DISCLAIMER: These notes are not necessarily an accurate representation of what I said during the class. They are mostly what I intend to say, and have not been carefully edited.

Today’s topics:
1. Set Theoretic Operations on Regular Languages
2. More Examples of Pumping Lemma

1 Set Theoretic Operations on Regular Languages

Regular languages are also closed under a variety of operations. Many of which can be created by taking and interpreting combinations of states.

They include:

1. Complement of a language $L$: we turn all accepting states to rejecting, and all rejecting states to accepting (aka. replace $F$ by $Q \setminus F$).
2. Intersection of two languages $L_1 \cap L_2$
3. Union of two languages: can obtain from complement of union.

However, note that these statements say nothing about the non-regular languages. In particular, it’s possible for the square of a non-regular language (e.g. strings with same numbers of 0s and 1) to be regular.

2 Pumping Lemma

Then the question is how to show some language is not regular. Here we can turn to the pumping lemma.

Theorem 2.1 (Pumping Lemma for Regular Languages). If $A$ is a regular language, then there is a number $p$ (the pumping length) where, if $s$ is a string in $A$ and $|s| \geq p$, 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. $|y| > 0$, and
3. $|xy| \leq p$.

We can use this lemma to show that the language

$$B = \{0^n1^n \mid n \geq 0\}$$

is not regular.

Assume for the sake of contradiction, that $B$ is regular. Then, let $p$ be the pumping length given by the lemma. Pick a string $s = 0^p1^p$. Since $s \in B$ and $|s| \geq p$, the pumping lemma implies that $s$ can be divided into three pieces $s = xyz$, such that $xy^iz \in B$ for all $i \geq 0$. Let us consider the ways in which $s$ cannot be divided into $x$, $y$ and $z$:

1. Suppose $y$ has only 0s. Then $xy^2z$ has more 0s than 1s (note that it cannot be length 0 from condition 2 of the lemma), so it cannot be in $B$.

2. The case where $y$ has only 1s follows similarly.

3. Suppose $y$ has both 0s and 1s. Then, in $xy^2z$, some 1s appear before 0s and so it is not in $B$.

Hence, there is no division of $s$ which satisfies the properties of the lemma. This must mean that our initial assumption that $B$ is regular was false.

**Things to note**

- You **cannot** choose $p$ - you have to assume this is given by the lemma.
- You **can** choose the string $s$.
- You **cannot** choose one particular division of $s$. You have to prove that no possible division can satisfy the three conditions.

For example, the language

$$\{w \in \{0,1\}^* \mid w \text{ has an even number of zeros}\}$$

is regular: let $p = 10$, we can consider any $w$ of length greater than 10.

### 3 Proof of Pumping Lemma

We applied the lemma as a black box, but let us see why it is true.

**Proof of Theorem 2.1.** Since $A$ is regular, we can construct a DFA $M$ recognizing it. Let this be $M = (Q, \Sigma, \delta, q_1, F)$. Let us assign the pumping length to be $p = |Q| + 1$. Now, let $s$ be any string accepted by $M$ of length at least $p$. (What if there are no such strings? Then the lemma is true by default!)
Consider the set of states visited by $M$ on input $s$. Since $s$ has at least $|Q| + 1$ characters, at least one state must be visited twice. Let $q_i$ be the first such repeated state. The sequences of states visited look like:

$$q_1, \ldots, q_i, \ldots, q_i, \ldots, q_F$$

Let $x$ denote the first part of $s$ which the machine reads before reaching $q_i$ the first time. Let $y$ denote the part of $s$ reads between $q_i$ and returning to $q_i$. Let $z$ be the rest of the string. Now, it is easy to see that replacing $y$ with $y^i$ for any $i \geq 0$ will also lead the machine to reach $q_F$ (condition 1). $y$ cannot be empty since it is between two separate occurrences of a state (condition 2). Lastly, since $q_i$ is the first repetition of a state, the number of characters read until then is at most $|Q| + 1 = p$ (condition 3).

Note: The converse of the pumping lemma is not true! That is, a language satisfying the lemma may still be non-regular.

### 4 Examples of Using Pumping Lemma

**Example 1: pumping up**

Let $C$ be the language $\{1^n^2 \mid n \geq 0\}$. $C$ is a unary language, which only accepts strings whose lengths are perfect squares. Let’s use the pumping lemma to prove that $C$ is not regular.

Assume for the sake of contradiction, that $C$ is regular. Then, let $p$ be the pumping length given by the lemma. Pick a string $s = 1^p^2$. Since $s \in C$ and $|s| \geq p$, the pumping lemma implies that $s$ can be divided into three pieces $s = xyz$, such that $xy^i z \in D$ for all $i \geq 0$. So, we have that $|xy^i z|$ is always a perfect square.

From condition 3, we know that $|xy| \leq p \implies |y| \leq p$. Then, $|xy^2 z| = |xyz| + |y| \leq p^2 + p < (p + 1)^2$. Hence, the only way $|xy^2 z|$ can be a perfect square is if $|y| = 0$, which contradicts condition 2 of the lemma.

**Example 2: pumping down**

Let $D$ be the language $\{0^i1^j \mid i > j\}$. Let’s use the pumping lemma to prove that $D$ is not regular.

Assume for the sake of contradiction, that $D$ is regular. Then, let $p$ be the pumping length given by the lemma. Pick a string $s = 0^{p+1}1^p$. Since $s \in C$ and $|s| \geq p$, the pumping lemma implies that $s$ can be divided into three pieces $s = xyz$, such that $xy^i z \in D$ for all $i \geq 0$.

From condition 3, we know that $|xy| \leq p \implies y$ has only 0s. Then, $xy^0 z = xz$ has equal or lesser 0s than 1s, which contradicts condition 1 of the lemma.