The Pumping Lemma

If you reach the end of a section early, please do not work on the next section - instead work on the questions at the end ("Side Questions" section).

1. Infinite regular languages

a. Given a DFA, what features of the machine would you look for to determine if it accepts infinitely many strings or only finitely many? That is, fill in the blank with something that refers only to the structure of the machine (e.g. features of the diagram like paths and cycles): A DFA M accepts infinitely many strings if and only if **there is a cycle in the DFA, a path from the start state to that cycle, and a path from that cycle to an accept state**.

b. Given a DFA which recognizes an infinite language L, create a procedure which outputs a description of some infinite subset of L. (For example, if the machine actually recognizes $1 \cup (00)^*$, some valid outputs for your procedure include $(00)^*$ and $(0000)^*$. Describe your procedure in high-level steps/pseudocode - for example, you could have the step "First, find some path in the diagram from one accept state to another.")

   Solution: Find a cycle in the graph which is reachable from the start state, and from which an accept state can be reached. (We know this exists from part a.) Pick a state q in the cycle. Let X be the labels on the transitions along a path from the start state to q, Y be the labels along the cycle from q back to itself, and Z be the labels along the path from q to an accept state. Output XY^*Z.

The Pumping Lemma

[Theorem 1.70 in Sipser] 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. $|y| > 0$, and
3. $|xy| \leq p$.

2. Pumping Lemma + Regular Languages

a. Find the smallest possible pumping length for $11111 \cup (01)^*$.

   Solutions: The smallest pumping length is 6. We can see that any string who's size is at least 6 is of the form $010101(01)*$. Thus we can pick $x = 01, y = 01$, and z is the rest of the string. From here it is clear that $xy^iz \in L$.

We cannot pick a smaller pumping length, because we cannot divide the string 11111 into xyz such that the conditions of the pumping lemma are satisfied.

b. Does the pumping lemma also apply to *finite* regular languages? For example, is there a valid pumping length for the language $\{001, 11010\}$?

Solution:
Yes. We can pick the pumping length p to be anything larger than the length of the longest string in our language. (Since our language is finite and every string has a length, we know that there must exist such a maximum.) From here, there will be no strings in our language of length at least p, so the pumping lemma will be vacuously true.

c. Is there a way to use the pumping lemma as a tool to prove a language is regular? Is there a way to use it to prove a language is *not* regular?

Solution:
There is no way to use the pumping lemma as a tool to prove that a language is regular. The pumping lemma tells us that if a language is regular, then it is 'pumpable'. If you recall from previous logic classes, the converse is not always true. In this case we **cannot** say that if a language is 'pumpable' then it is regular. (In fact there are indeed examples of a 'pumpable' language that is not regular.)

We can (and do, below) use the pumping lemma to prove that a language is not regular. The pumping lemma tells us that if a language is regular, then it is 'pumpable'. If you recall from previous logic classes, the contrapositive is always true. So the pumping lemma tells us that if a language is not 'pumpable' then it is not regular. See examples below.

3. Pumping Lemma + Nonregular Languages

a. Let $L = \{ww^{\mathcal{R}} \mid w \in \{0,1\}^*\}$. *(Recall $w^{\mathcal{R}}$ is the string w reversed.)* Prove that L is nonregular using the pumping lemma by filling in this template:

Proof: Fix p an arbitrary positive integer. We will show that p is not a pumping length for L by picking a string $s \in L$ with length greater than or equal to p, and showing s cannot be pumped. Let s be $\boxed{1^p001^p}$. *(Some values for s will be much easier to work with than others, so list a few first, try to pick an easy one, and come back and change it if it doesnt' work.)*

Suppose $s = xyz$ with $|xy| \leq p$ and $|y| > 0$. Then the following must be true about x and/or y:

x and y are made up of entirely 1s and y is made up of at least one 1. Using variables, we can say $x = 1^m, y = 1^n, z = 1^{p-m-n}001^p$, where $n \geq 1$ and $m + n \leq p$.

But then when $i = \boxed{2}$, $xy^iz = \boxed{1^{p+n}001^p}$, which is not in L because our choice of w would need to have exactly one 0 since there are two 0s in xy^2z; the only way to split the string with one 0 in each part is between the two zeros (so $1^{p+n}0$ and 01^p), but then $1^{p+n}0 \neq 1^p0$, so there is no possible choice of w that would allow us to divde xy^2z into $ww^{\mathcal{R}}$

So s cannot be pumped, which means p is not a pumping length for L. And since p was arbitrary, we've shown that there is *no* pumping length for L. So by the (contrapositive of the) pumping lemma, L is not regular.

b. Prove $L = \{0^j1^k \mid j \geq k \geq 0\}$ is nonregular using the pumping lemma.
Solution:
Fix p an arbitrary positive integer. We will show that p is not a pumping length for L by picking a string $s \in L$ with length greater than or equal to p, and showing s cannot be pumped. Let s be 0^p1^p.
Since s is in L and has length greater than or equal to p, the Pumping Lemma guarantees that s can be divided into parts x, y, z where $s = xyz$, $|y| > 0$, $|xy| \leq p$, and for each $i \geq 0$, $xy^iz \in L$. Since the first p letters of s are all 0 and $|xy| \leq p$, we know that x and y are made up of all 0s. Say $x = 0^m, y = 0^n, z = 0^{p-m-n}1^p$, where $m + n \leq p$ and $n \geq 1$. If we let $i = 0$, we get $xy^0z = 0^{p-n}1^p$. Since $n > 0$, $p - n < p$, so we can conclude that $xy^0z \notin L$.
So s cannot be pumped, which means p is not a pumping length for L. And since p was arbitrary, we've shown that there is *no* pumping length for L. So by the (contrapositive of the) pumping lemma, L is not regular.

c. Prove $L = \overline{\{0^j1^k \mid j \geq k \geq 0\}}$ is nonregular. (Hint: the pumping lemma is not the best tool for this.)
Solution:
Let's assume towards a contradiction that L is in fact regular. Since regular languages are closed under complementation (as we saw in lecture 1), we get that $\overline{\overline{\{0^j1^k \mid j \geq k \geq 0\}}} = \{0^j1^k \mid j \geq k \geq 0\}$ is also regular. However, this contradicts what we showed in part b: that $\{0^j1^k \mid j \geq k \geq 0\}$ is nonregular. So our original assumption must have been false and so L is nonregular.

d. Prove $L = \{0^n1^m0^n \mid m, n \geq 0\}$ is nonregular.
Solution:
Fix p an arbitrary positive integer. We will show that p is not a pumping length for L by picking a string $s \in L$ with length greater than or equal to p, and showing s cannot be pumped. Let s be 0^p10^p.
Since s is in L and has length greater than or equal to p, the Pumping Lemma guarantees that s can be divided into parts x, y, z where $s = xyz$, $|y| > 0$, $|xy| \leq p$, and for each $i \geq 0$, $xy^iz \in L$. Since the first p letters of s are all 0 and $|xy| \leq p$, we know that x and y are made up of all 0s. Say $x = 0^m, y = 0^n, z = 0^{p-m-n}1^p$, where $m + n \leq p$ and $n \geq 1$. If we let $i = 2$, we get $xy^2z = 0^{p+n}10^p$. Since $n > 0$, $p + n \neq p$, so we can conclude that $xy^2z \notin L$.
So s cannot be pumped, which means p is not a pumping length for L. And since p was arbitrary, we've shown that there is *no* pumping length for L. So by the (contrapositive of the) pumping lemma, L is not regular.