How to Use The Pumping Theorem

Ben Wiedermann (ben@cs.utexas.edu)

October 4, 2004

1 The Theorem

Student: Hey! What is this pumping theorem I keep hearing about?

Teacher:
If L is regular
$\exists N \geq 1$**, such that**
$\forall$ **strings** $w \in L$**, where** $|w| \geq N$
$\exists x, y, z$**, such that** $w = xyz$
and $|xy| \leq N$
and $y \neq \epsilon$
and $\forall q \geq 0, xy^q z \in L$

Student: Ummm. . . Okay. That's what it said in the class notes. But what does it mean?

Teacher: Well, let's go through it line by line.

Student: Yeah, but. . . I don't see how that's going to help.

Teacher: Let's do it anyway, because I think it might help later.

Student: Okay.

Teacher: Okay. The first line is
If L is regular
This means that our theorem applies only if the language is regular. It doesn't apply if L is not regular.

Student: But I thought we use the pumping theorem to show that a language *isn't* regular.

Teacher: You're jumping ahead a bit. What you said is right, but we use a proof by contradiction. In other words, we assume L *is* regular, then we show that it doesn't satisfy the pumping theorem. This gives us a contradiction, so our initial assumption that L is regular must be wrong.

Student: That makes sense, I guess. But before we get into all that, I want to make sure I understand the rest of the theorem.

Teacher: Right. Let's look at the next line:
$\exists N \geq 1$**, such that**
This means that there's some "magic number" N for the language L, such that all the things in the rest of the theorem are true.

Student: So does that mean N is different for every regular language?

Teacher: That's right. Or rather, it means that it *could* be different. In other words, we don't know. We can't say anything about N other than that it must exist.

Student: Didn't Descartes say something like that?

Teacher: Ummm... Let's move on. The next line says:
$\forall$ **strings** $w \in L$**, where** $|w| \geq N$
This means that the rest of the theorem is only going to apply to certain strings. Namely, the ones that are at least as long as that magic number N.

Student: I'm with ya.

Teacher: The next line says:
$\exists x, y, z$**, such that** $w = xyz$

Student: So this means that x, y, and z exist, but we don't necessarily know what they are?

Teacher: Right just like with N. By the way, this is the line of the proof that we later use to show when a language can't be regular. If there's no possible way to find an x, y, and z for some string w, then L can't be regular, right?

Student: Yeah... I guess I see how that's true, but it seems like there's so many possible values for x, y, and z that it would be hard to show that there's absolutely no way to split up some w like that.

Teacher: Well, fortunately, we have the last three lines of the theorem. Let's look at the first one:
and $|xy| \leq N$

Student: Okay. Next.

Teacher:
and $y \neq \epsilon$

Student: So that means $1 \leq |y| \leq N$.

Teacher: Good! That fact will help us in our proofs later on.

Student: Okay. And now the last line of the theorem
and $\forall q \geq 0, xy^q z \in L$
Yikes!

Teacher: Ah. This is the "pumping" part of the theorem. It says "no matter how many copies of y I remove or add to the string, the resulting string is still in the original language L".

Student: Oh. So when $q = 0$, that's "pumping out", because we've removed one copy of y from the string. And when $q = 1$, well that's just the original string w. But when $q > 1$, then we're pumping copies of y into the string.

Teacher: Right. In fact, we're pumping $q - 1$ copies. See, because $q = 2$ means we have two copies of y, which is one more than we had in the original string.

Student: So, when you say “pump in once”, you’re just saying, “let $q = 2$”?

Teacher: Exactly.

Student: So let me see if I understand it all. If L is regular, then there’s some magic number N. And if I take any string w that is at least as long N, then I can break it up into three parts x, y, and z. Now, the length of xy is less than or equal to N, and also the length of y is greater than or equal to 1 and less than or equal to N.

Teacher: Good...

Student: Okay, and finally, I can remove y, or I can add any number of copies of y, and the resulting string should still be in L.

Teacher: Excellent

Student: Okay, I got it. Now how do I use it?

Teacher: Good question.

2 Using the Pumping Theorem

Teacher: Remember how I said we could use a proof by contradiction to show that a language is not regular?

Student: Yep.

Teacher: Do you think you could show how to do that?

Student: Nope.

Teacher: Oh, come on. I think you can.

Student: Okay. Well, let’s say I have a language that I think is not regular. First, I assume it’s regular. Then, I show that it violates the pumping theorem. I know how to assume things. It’s the showing things that I always have a problem with.

Teacher: Well, we know that if the language *were* regular, then there would be some magic number N such that the rest of the theorem is true.

Student: Oh! So I can just show that there can’t possibly be an N!

Teacher: Well, you could do that, but I think it would be pretty hard. You’d have to consider all possible values for N and then show that the pumping theorem doesn’t hold. So let’s just assume that some N does exist. Can we still find a way to show the pumping theorem doesn’t hold for L?

Student. Hmmm. . .. If there *were* an N, then every string w that was at least as long as N would have to satisfy the rest of the theorem.

Teacher: So, what if you want show that that’s not true?

Student: I would have to show that there was some w that didn’t satisfy the rest of the theorem.

Teacher: Right. And what would that w look like?

Student: Oh, I remember now. w looks like xyz, where $|xy| < N$ and $y \neq \epsilon$ and. . .

Teacher: . . . and if L *were* regular, then, no matter how we pick y, we could find a q such that $w' = xy^q z \in L$.

Student: Oh, okay. So I just need to show that, no matter how we pick y, we could find a q such that $w' = xy^q z \notin L$.

Teacher: Good job. It looks like you really understand it. Have a nice day and let me know if you have any more questions.

Student: Wait, wait, wait, wait, wait. Can we try some examples?

Teacher: I was hoping you'd say that! Let's try to show that. . .

3 $L = a^n b^n$ is not regular

Student: Okay, so I know that there's a magic number N. Now I just need to find a string w and show that I can pump y to make w go out of L.

Teacher: Good. Let's try the string $w = a^N b^N$.

Student: Hey! *I* wanted to come up with w.

Teacher: Well, picking a good w can be hard. Let's just assume we already have a w, and show that we can pump y and make w go out of L.

Student: Okay.

Teacher: The first thing I do when I have a w is to list all the possible choices for y.

Student: Oh, okay. $1 \leq |y| \leq N$, so y must be a^i, when $1 \leq i \leq N$.

Teacher: Good. Once I know what y looks like, I then try to write down what w looks like.

Student: I thought we already know what w looks like.

Teacher: Well, we *do*. But I meant that I try to write down w in terms of x, y, and z.

Student: Okay. If $y = a^i$ and $|xy| \leq N$, then $x = a^j$, where $i + j \leq N$. So $w = (a^j)(a^i)(a^{N-(i+j)} b^N)$.

Teacher: Very good. Now all we have to do is show that, for every possible value of y, we can find a q such that $w' = a^j (a^i)^q a^{N-(i+j)} b^N \notin L$.

Student: Okay. Well, let's take $q = 0$. Then $w' = a^j a^{N-j} b^N$, which I can rewrite as $a^N b^N$. Wait. That string is still in L. What happened?

Teacher: You got your is and qs mixed up. Remember, the q applies only to the y part of the string, but the i has implications on both the x and y part of the string. You plugged in a 0 for q, but you also plugged in a 0 for the first i.

Student: Oh, I see. So, I should have said $w' = a^j a^{N-(i+j)} b^N$. I can rewrite that as $w = a^{N-i} b^N$. That looks much better. I proved that L is not regular!

Teacher: Well, not quite. What if $i = 0$?

Student: Hey! That's cheating. You don't get to pick what i is.

Teacher: That's true. But you do have to show that, for every possible value of i, $a^{N-i} b^N \notin L$. So what if $i = 0$?

Student: Hmmm... Oh, wait! i *can't* be 0. We said so before. i has to be greater than or equal 1 and less than or equal to N.

Teacher: Good. So, given the range for i, is it possible that $a^{N-i} b^N \in L$?

Student: No. See, the most number of as the string could have is $N - 1$, and the least number of as it could have is 0. So there's no way.

Teacher: Good. *Now* you've proved that L is not regular.

Student: Okay. But every example I've ever seen is $a^n b^n$. I *know* it's not regular. Can we try another one?

Teacher: Sure. Let's try to show:

4 $L = \{a^i b^j c^k$, where $i + j = k\}$ Is Not Regular

Student: Okay. I know we're going to talk about how to pick good ws later, so can you give me a w?

Teacher: Try $w = a^N b^N c^{2N}$.

Student: Okay, thanks. Let's see. First I need to say what all the possible values for y are. Well, it's obvious that $y = a^i$, where $1 \leq i \leq N$.

Teacher: Good. So now what does w look like, in terms of x, y, and z?

Student: $w = (a^j)(a^i)(a^{N-(i+j)} b^N c^{2N})$, where $i + j \leq N$.

Teacher: Good. Now what?

Student: Now I need to show that, for every possible value of y, there's some way to pump it such that the resulting string is no longer in L.

Teacher: Good.

Student: So, if I let $q = 2$ ...

Teacher:... which means you're going to pump in one extra copy of y ...

Student: ... then I get $w' = a^j a^i a^i a^{N-(i+j)} b^N c^{2N}$. Now, I can rewrite that as $a^{N+i} b^N c^{2N}$.

Teacher: Right. Now what?

Student: Now I need to show that for all possible values of i, $w' \notin L$. Well, that would mean that for all possible values of i, $N + i + N \neq 2N$. That *has* to be true, because the smallest value for i is 1. So, I've proved that L is not regular!

Teacher: Nicely done.

Student: I fell pretty good about that.

Teacher: Good. Let's take a break for a second and talk about—

Student: —the Longhorns football team?

Teacher: No, not that. This:

5 Picking a good w

Student: That sounds like an impossibility, when it comes to Presidents.

Teacher: Whoa! I'm not gonna talk about politics. I think it's against the Teacher's Code or something. Let's just stick to regular languages.

Student: Is that all you ever talk about?

Teacher: Yes. Well, I'm also really interested in glaciology.

Student: Ummm. . . Let's go back to languages.

Teacher: Okay. The key to picking a good w is that you want the pumping to affect only one portion of the string. The best way to do this is to make an N-sized region at the beginning of w. And preferably one that contains only one kind of character from Σ.

Student: Why is that a good idea?

Teacher: For two reasons. First, the pumping only happens to the y portion of the string. So, if we have an N-sized region at the beginning of the string, the pumping will affect only that region, and leave the other regions untouched.

Student: Oh. That how you came up with $a^N b^N c^{2N}$ for that last one.

Teacher: Right. That way when we pumped in, we only affected one part of the addition, leaving the other parts the same, and pretty much guaranteeing that w' would be out of the language.

Student: So, why do we want the region to contain only one kind of character from Σ?

Teacher: That way, we have fewer cases to consider for y. For example, in our first problem, y had to be a^i, because a was the only character that appeared in the first N characters of w. See, we could have chosen $w = a^{N/2} b^{N/2} c^N \ldots$

Student: Oh, but then y could be some combination of as and bs, so there'd be more cases to consider for what xy could be.

Teacher: Right, and then we'd have to consider what the implications on the addition operands are. For example, if y lies entirely in the a region, then only the first $N/2$ operand is changed. But what if y spans the as and bs?

Student: What a mess.

Teacher: Exactly. So we saved ourselves the hassle by picking the w we did. Our job is hard enough as it is, so we have to do everything we can to make it easy on us.

Student: Sounds good to me. So I should always pick a w that has at least N of the same characters at the beginning.

Teacher: Well, that's not always possible. For example, what if I ask you about $L = ab^n c^n$?

Student: That's not regular.

Teacher: Sure. But you have to tell me why.

Student: Okay, okay. Well, I see that I can't get a w with all the same character at the beginning. So I'll get as close to that as I can. How about $w = ab^{N-1}c^N$?

Teacher: That seems like it will work. Why did you pick that $N - 1$?

Student: Because I've already got that a at the beginning, and I want to try to limit $|xy|$ to be exactly N. That way, when I pump, I don't affect the region after N characters.

Teacher: That's a good strategy. Hey, let me ask you a question. Could you ever use a^{N-2} in one of your strings?

Student: Sure. Why not?

Teacher: I'll tell you why. Because the pumping theorem says that $N \geq 1$. So, if we had a^{N-2}, then it's possible for that to mean a^{-1}. And that's just silly.

Student: I guess that *is* silly. I'll have to remember that bit about $N - 2$.

Teacher: Or N minus anything greater than 1.

Student: Right.

Teacher: Let's get back to our problem. How do you show that your w doesn't satisfy the pumping theorem?

Student: First I think about all the possible values of y. Let's see. y can be ab^i.

Teacher: Good. And what are the possible values for i?

Student: Because $|xy| \leq N$, i can be greater than or equal to 0 or less than or equal to $N - 1$

Teacher: Excellent! So what does w look like in this case?

Student: $w = (\epsilon)(ab^i)(b^{N-1-i}c^N)$. I threw that ϵ in there, because we're saying that y starts at the beginning of the string.

Teacher: Good. Are there any other possibilities for y?

Student: Sure. It could be that $y = b^j$, where $1 \leq j \leq N-1$. In this case $w = (ab^k)(b^j)(b^{N-1-(k+j)}c^N)$, where $k + j \leq N - 1$.

Teacher: Solid. Now we have to show that for every possible value for y, we can find a q such that $w' = x(y)^q z \notin L$. We know there's two possibilities for the form of y. Let's take the first one.

Student: Okay, in this case, I need to make sure that, for every possible value of i, there is a q such that $w' = (ab^i)^q b^{N-1-i}c^N$.

Teacher: That's good. A lot of people mix up their qs and is at this point. You were right to remember that we still need to make sure that every i has a q that pumps the string out of the language. Most of the time, we can find one q that will work for every i. Sometimes, though, we need to find one q that satisfies some of the is and another q that satisfies the rest of them.

Student: I'll keep that in mind. But I think I can find one q that works for all the is in this case.

Teacher: Go for it.

Student: Okay, so I let $q = 2$, then I have $w' = ab^i ab^i b^{N-1-i}c^N$. I can stop right there, because this has too many as in it.

Teacher: Right. So now we need to do the other possibility for y.

Student: Okay. I let $q = 0$, so we get $w' = ab^k b^{N-1-(k+j)}c^N = ab^{N-1-j}c^N$. Hmmm... I got a little confused here. What do I do next?

Teacher: You're almost there. You just need to show that $N - 1 - j \neq N$ for all possible values of j.

Student: Oh, right! Well, that's easy. Because of the range of values we have for j, $0 \leq N - 1 - j \leq N - 2$. So, there we go.

Teacher: Good job. How do you feel?

Student: Good. I really appreciate it. I think I'm going to go have a big cup of coffee or something now.

Teacher: Hold on a second. I have one more important thing to ask you.

Student: Uh-oh. I feel one of those big headings coming up.

Teacher: That's right:

6 Is $L = \{ab^i c^j, i \neq j\}$ regular?

Student: No.

Teacher: Hey, that's a good guess, but can you prove it to me?

Student: Well, I know how to do it when i and j are equal; that's what we just did. But this seems really hard. Is there some trick to it?

Teacher: There is. Remember how we defined all those operations under which regular languages are closed?

Student: Oh, yeah. Like, if L is regular, then $\neg L$ is regular. Oh! So we can just take the negation of that L up there and do just like we did before.

Teacher: Not quite. $\neg L$ does include all the strings of the form $ab^N c^N$. But it also includes all the strings where there's some cs before the bs, and it includes the empty string, and generally lots of other stuff that will make it very hard to use the pumping theorem.

Student: Oh. I didn't think about that. This seems like just as hard a problem as we had before.

Teacher: Well, what we want is to somehow "pull out" only the strings of the form $ab^N c^N$. Can you see a way to do that?

Student: Oh! $ab^N c^N = \neg L \cap ab^* c^*$

Teacher: Right. So our proof goes like this: Assume L is regular. Then $L' = \neg L \cap ab^* c^* = ab^N c^N$ must be regular. Now we show that L' is not regular, using your excellent analysis from before. Because L' is not regular, then our initial assumption that L was regular is wrong. L must not be regular.

Student: I get it. Huh. That $\cap$ operation was like a bit mask that let me select what part of the language I wanted to look at.

Teacher: That's not a bad way to think about it.

Student: Whew. I've got a lot to think about. I'm probably going to need a whole pot of coffee. I feel a lot better about the pumping theorem, though. Thanks.

Teacher: You're welcome.

7 Epilogue

Teacher: You know, this whole thing reminds of that play *The Miracle Worker*, where they're down by the water pump. . .

Student: Geez. My Descartes reference was less obscure than that! I'm outta here.

Teacher: Good luck on the test.

Student: I wouldn't need luck if you made the test really easy.

Teacher: I wouldn't be much of a "test" if it were easy.

Student: It would be a test of your compassion.

Teacher: It sounds like a test of my patience. . .

Student: Okay, okay. I know how to do these problems now anyway. The test should be a piece of cake.

Teacher: Mmmmm. . . cake.

Student: Bye.