Nonregular Languages

- How to show that a given language is nonregular.
- In some sense, we need to prove that No DFA is possible to recognize the language.
- How we do this?

Some properties can help us

- L is regular => L obeys "Pumping Lemma"
- DFA must have finite number of states.
  - For the given L, we need infinite number of states in the DFA.
  - Myhill-Nerode Theorem (Gives a necessary and sufficient condition for regular languages).
- There are other ways …

- Pumping Lemma is useful to show that L is nonregular.
- It cannot be used to show that L is regular.
- Why?

the sufficient condition

IF

$\Longrightarrow$

(If you assume this, you'll get what you want.)

the necessary condition

ONLY IF

$\Longleftarrow$

(You can't get what you want without assuming this.)

- A => B (A being true is a sufficient condition for B to be true)
  - If A is true, we know B is true.
  - If A is false, what about B?
- A <= B (A being true is a necessary condition for B to be true)
  - If A is false, we know B is false.
  - If A is true, what about B?

- A => B (A being true is a sufficient condition for B to be true)
  - If A is true, we know B is true.
  - If A is false, what about B?
  - If B is false, then what about A?

- L is regular => L obeys “Pumping Lemma”
- If L fails to obey “Pumping Lemma” then L is nonregular.

- Moral of the story: Never use Pumping Lemma to prove that L is regular.

- Nonregular examples

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

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

- But, the following is regular

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

- Nonregular examples

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

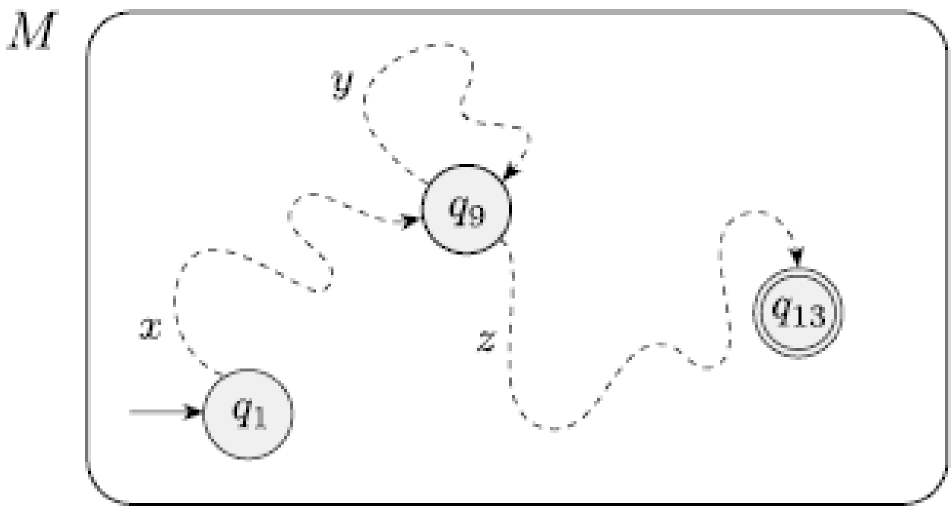
$$C = \{w \mid w \text{ has an equal number of 0s and 1s}\}$$

- But, the following is regular

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

  - http://www.cs.gordon.edu/courses/cps220/Notes/nonregular_languages
    - Follow the above URL for an answer to show D is regular.

Pumping Lemma



- In any DFA, if $w = xyz$ is "long enough", then such a loop must occur. Why?

Pigeonhole Principle

The following figure shows the string s and the sequence of states that M goes through when processing s. State q_9 is the one that repeats.

$$
\begin{array}{cccccccccccccccccc}
s = & & s_1 & & s_2 & & s_3 & & s_4 & & s_5 & & s_6 & & \ldots & & s_n & \\
& \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow \\
& q_1 & & q_3 & & q_{20} & & q_9 & & q_{17} & & q_9 & & q_6 & & q_{35} & & q_{13}
\end{array}
$$

FIGURE 1.71
Example showing state q_9 repeating when M reads s

Pumping Lemma

THEOREM 1.70

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 can 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. $|xy| \leq p$.

THEOREM 1.70

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 can 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. $|xy| \leq p$.

When s is divided into xyz, either x or z may be ε, but condition 2 says that $y \neq \varepsilon$.

Observe that without condition 2 the theorem would be trivially true.

Negation of Pumping Lemma

- There is **a** string *w* in *L* which of length atleast of the pumping length *(p)*, where **every** division of *w* into *xyz* fails to satisfy at-least **one** of the following –
  - $|y| \neq 0$
  - $|xy| \leq p$
  - $x\, y^i\, z$ is in L for all i in {0,1,2,…}.

Negation of Pumping Lemma (we simplify)

- There is **a** string *w* in *L* which of length atleast of the pumping length *(p)*, where **every** division of *w* into *xyz* that obeys
  - $|y| \neq 0$
  - $|xy| \leq p$

  Fails to satisfy the following for at-least one *i.*
  - $x\ y^i\ z$ is in *L* for all i in *{0,1,2,…}.*

Show that, $\boldsymbol{L = \{0^{n^2} | n \geq 1\}}$ is nonregular.

Show that, $L = \{0^{n^2} | n \geq 1\}$ is nonregular.

<u>Proof</u> [by Pumping Lemma]:

Let p be the pumping length.

Let us choose $w = 0^{p^2}$ and we know, $|w| \geq p$.

Show that, $L = \{0^{n^2} | n \geq 1\}$ is nonregular.

<u>Proof</u> [by Pumping Lemma]:

Let p be the pumping length.

Let us choose $w = 0^{p^2}$ and we know, $|w| \geq p$.

Let us choose $x = 0^m,\ y = 0^n,\ z = 0^r$ and $m + n + r = p^2$.

Show that, $L = \{0^{n^2} | n \geq 1\}$ is nonregular.

<u>Proof</u> [by Pumping Lemma]:

Let p be the pumping length.

Let us choose $w = 0^{p^2}$ and we know, $|w| \geq p$.

Let us choose $x = 0^m,\ y = 0^n,\ z = 0^r$ and $m + n + r = p^2$.

Along with this, we have $1 \leq n \leq p$.

Show that, $L = \{0^{n^2} | n \geq 1\}$ is nonregular.

<u>Proof</u> [by Pumping Lemma]:

Let p be the pumping length.

Let us choose $w = 0^{p^2}$ and we know, $|w| \geq p$.

Let us choose $x = 0^m,\ y = 0^n,\ z = 0^r$ and $m + n + r = p^2$.

Along with this, we have $1 \leq n \leq p$.

Now, consider the string xy^2z. Note, $|xy^2z| = p^2 + n$.

We have, $p^2 < p^2 + n < (p+1)^2$.

Show that, $L = \{0^{n^2} | n \geq 1\}$ is nonregular.

<u>Proof</u> [by Pumping Lemma]:

Let p be the pumping length.

Let us choose $w = 0^{p^2}$ and we know, $|w| \geq p$.

Let us choose $x = 0^m,\ y = 0^n,\ z = 0^r$ and $m + n + r = p^2$.

Along with this, we have $1 \leq n \leq p$.

Now, consider the string xy^2z. Note, $|xy^2z| = p^2 + n$.

We have, $p^2 < p^2 + n < (p+1)^2$.

So, $|xy^2z|$ is not a perfect square, hence xy^2z is not in L .

Thus, Pumping Lemma failed for L.

EXAMPLE

Let B be the language $\{0^n1^n | n \geq 0\}$. We use the pumping lemma to prove that B is not regular.

Let p be the pumping length.

Let $s = 0^{\lceil p/2 \rceil}1^{\lceil p/2 \rceil}$

Because s is a member of B and s has length more than 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 is in B with conditions $y \neq \epsilon$ and $|xy| \leq p$.

We consider three cases to show that this result is impossible.

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

A good choice for the string

EXAMPLE

Let B be the language $\{0^n1^n | n \geq 0\}$. We use the pumping lemma to prove that B is not regular.

Let p be the pumping length.

Choose s to be the string 0^p1^p.

Because s is a member of B and s has length more than 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 is in B.

With conditions $y \neq \epsilon$ and $|xy| \leq p$,

y can be of only 0s,

and the string xy^2z will clearly have more 0s than 1s,

hence is not in the language.

Following are all nonregular.

1. Strings having equal number of 0s and 1s.

2. Dyck language.

   $\Sigma = \{(,)\}$. Dyck language is the set of all balanced strings like $\{(), ()(), (()()),...\}$

3. Palindromes (over any alphabet, other than unary alphabet).

4. Copy language, i.e., $L = \{ww | w \in \Sigma^*\}$.

5. $L = \{0^n 1 0^n | n \geq 0\}$.

6. $L = \{ww^R | w \in \Sigma^*\}$.

- Can you prove for each of these that PL fails.

What is wrong?

In order to show that the set of palindromes over $\Sigma = \{0,1\}$ regular,

I have chosen $s = 0^{\lceil p/2 \rceil} 1 0^{\lceil p/2 \rceil}$.

Now, I split $s = xyz$, with $y = 1$.

I can pump y as many times as I want and the resulting string is in the language.

So, the language is regular.

- There are two mistakes.

Mistakes.

- The argument is not for a particular division of the chosen s in to xyz.

  But, for every division, satisfying the conditions $y \neq \epsilon$ and $|xy| \leq p$.

Mistakes.

- If you want to show that Pumping Lemma is true, then you have to show for all strings s, such that $|s| \geq p$, s can be divided in to xyz, satisfying the three conditions. Just showing it for one s is not enough.

- Note, on the otherhand, to show that Pumping Lemma is false, you can choose just one string s whose length is at-least p, but, now, for every division of s in to xyz, at-least one of the three conditions is not satisfied.

- But, the serious mistake is, you have not learnt the moral.
- Never use PL to show that a language is regular.

Set of primes – a nonregular language

- Let p be the pumping length.
- Consider $s = 0^n$ where n is prime and let $n \geq p$, represents a prime number.

Set of primes – a nonregular language

- Let p be the pumping length.
- Consider $s = 0^n$ where n is prime and let $n \geq p$, represents a prime number.
- Now, let us divide $s = 0^x 0^y 0^{(n-x-y)}$ and $y > 0, x + y \leq p$.

Set of primes – a nonregular language

- Let p be the pumping length.
- Consider $s = 0^n$ where n is prime and let $n \geq p$, represents a prime number.
- Now, let us divide $s = 0^x 0^y 0^{(n-x-y)}$ and $y > 0, x + y \leq p$.
- Consider $i = n + 1$.
- We show $0^x \, (0^y)^i \, 0^{(n-x-y)}$ does not represent a prime number

- $0^x \, (0^y)^i \, 0^{(n-x-y)} \;\; = 0^{x+y(n+1)+n-x-y}$

  $= 0^{n(y+1)}$

Show that $L = \{a^i b^j | i \neq j\}$ is non-regular.

Direct proof using pumping lemma is somewhat an involved one. All the trouble is in choosing an appropriate $s \in L$ for which the lemma is going to fail. After some investigation the following s is found which will ease out the proof.

Show that $L = \{a^i b^j | i \neq j\}$ is non-regular.

Direct proof using pumping lemma is somewhat an involved one. All the trouble is in choosing an appropriate $s \in L$ for which the lemma is going to fail. After some investigation the following s is found which will ease out the proof.

Proof:

Let the pumping length be p.

Choose $s = a^p b^{p!+p}$. Here $p!$ is factorial of p.

Let $s = xyz$ where $x = a^{p-n}, y = a^n,\ z = b^{p!+p}$ such that $1 \leq |n| \leq p$.

This division of s into xyz satisfies the constraints, viz., (i) $|y| \neq 0$, and (ii) $|xy| \leq p$.

We show that for some $i,\ xy^i z \notin L$.

Show that $L = \{a^i b^j | i \neq j\}$ is non-regular.

Direct proof using pumping lemma is somewhat an involved one. All the trouble is in choosing an appropriate $s \in L$ for which the lemma is going to fail. After some investigation the following s is found which will ease out the proof.

Proof:

Let the pumping length be p.

Choose $s = a^p b^{p!+p}$. Here $p!$ is factorial of p.

Let $s = xyz$ where $x = a^{p-n}, y = a^n,\ z = b^{p!+p}$ such that $1 \leq |n| \leq p$.

This division of s into xyz satisfies the constraints, viz., (i) $|y| \neq 0,$ and (ii) $|xy| \leq p$.

We show that for some $i,\ xy^i z \notin L$.

Choose $i = \frac{p!}{n} + 1$. Note this i is a non-negative integer. Then $xy^i z =$

$a^{p-n}(a^n)^{\frac{p!}{n}+1} b^{p!+p} = a^{p-n+p!+n} b^{p!+p} = a^{p!+p} b^{p!+p} \notin L.$

An easy way to show $\{a^i b^j | i \neq j\}$ is nonregular

We know $a^* b^*$ is regular (why?)

We know $\{a^n b^n | n \geq 0\}$ is nonregular. Since PL fails for this.

Now assume $\{a^i b^j | i \neq j\}$ is regular.

Now this leads to a contradiction.

Can you prove these

1. $\{a^m b^n | m < n\}$ is nonregular.

2. $\{a^m b^n | m \leq n\}$ is nonregular.

3. $\{a^m b^n | m > n\}$ is nonregular.

4. $\{a^m b^n | m \geq n\}$ is nonregular.

- Prove or disprove: “every finite language is regular”.
- Prove or disprove: “every infinite language is nonregular”.

- Prove or disprove: “every finite language is regular”.
- True. We can build a NFA.
- Prove or disprove: “every infinite language is nonregular”.
- False. Counter example is: a*b*

- Prove or disprove : “nonregular languages are closed under union”.

- Prove or disprove : “nonregular languages are closed under union”.
- False.
- Counter example:

$\{a^n b^n | n \geq 0\} \cup \{a^i b^j | i \neq j\}$ is equal to $a^* b^*$, which is regular.

- Prove or disprove : “nonregular languages are closed under intersection”.

- Prove or disprove : "nonregular languages are closed under intersection".
- False.
- Counter example:

$\{a^n b^n | n \geq 0\} \cap \{a^i b^j | i \neq j\}$ is empty language, which is regular.

- Prove or disprove : “nonregular languages are closed under complementation”.

- Prove or disprove : “nonregular languages are closed under complementation”.
- True.
- Proof: [by contradiction] using the fact that regular languages are closed under complementation.

Example of nonregular language that satisfies the pumping lemma

Let $\Sigma = \{\$, a, b\}$.

Consider the language $L = \{\$a^n b^n | n \geq 1\} \cup \{ \$^k w | k \neq 1, w \in \{a, b\}^* \}$.

Let p be the pumping length.

For every string s such that $|s| \geq p$, we show that $s = xyz$ satisfying,

1. $|y| \neq 0$,

2. $|xy| \leq p$, and

3. For all i, $xy^i z \in L$.

Example of nonregular language that satisfies the pumping lemma

Let $\Sigma = \{\$, a, b\}$.

Consider the language $L = \{\$a^n b^n | n \geq 1\} \cup \{ \$^k w | k \neq 1, w \in \{a, b\}^* \}$.

Let p be the pumping length.

For every string s such that $|s| \geq p$, we show that $s = xyz$ satisfying,

1. $|y| \neq 0$,
2. $|xy| \leq p$, and
3. For all i, $xy^i z \in L$.

There are two cases. The string s might be of the form $\$a^n b^n$ or of the form $\$^k w$ where $k \neq 1$.

For all cases, except when $k = 0$, consider $y = \$$. This satisfies all three conditions.

For the case when $s \in \{ \$^k w | k = 0, w \in \Sigma^* \}$. Then s can be any string from $\{a, b\}^*$.

And, y can be any nonempty substring of s. This satisfies all three conditions.

Example of nonregular language that satisfies the pumping lemma

- $L = \{\$a^n b^n | n \geq 1\} \cup \{ \$^k w | k \neq 1, w \in \{a, b\}^* \}$.

Now, how is that we show L is nonregular.

This is through proof by contradiction.

Assume that L is regular.

We know $\$a^* b^*$ is regular. (why?)

Since regular languages are closed under intersection. Intersection of L and $\$a^* b^*$ must be regular.

But, Intersection of L and $\$a^* b^*$ is $\{\$a^n b^n | n \geq 0\}$, and this is nonregular (why?).

Hence, the contradiction.

Yet another language that is nonregular, but for which PL is satisfied.

- This language is very similar to the previous one.
- $L = \{a^i b^j c^k | (i = 1) \Rightarrow (j = k)\}$.

An Important Other way of showing that a language is nonregular

- By using Myhill-Nerode Theorem
  - DFA or NFA for a regular language must have finite number of states.
  - If you show that infinite number of states are needed, then it is equivalent to showing that the language in nonregular.
  - Apart from this, Myhill-Nerode theorem has one important application, viz., minimization of a DFA.

Myhill-Nerode Theorem is much more than the Pumping Lemma

- Myhill-Nerode theorem can be used to show that a language is regular also. Of course it can be used to show that a language is nonregular.
  - This gives a necessary and sufficient condition for a language being regular.
- Note, the pumping lemma, on the otherhand can be used only to show that a language is nonregular.
  - Pumping lemma should not be used to show that a language is regular.

1.29 Use the pumping lemma to show that the following languages are not regular.

[A]**a.** $A_1 = \{0^n1^n2^n \mid n \geq 0\}$

b. $A_2 = \{www \mid w \in \{a, b\}^*\}$

[A]**c.** $A_3 = \{a^{2^n} \mid n \geq 0\}$ (Here, a^{2^n} means a string of 2^n a's.)

- Reading Assignment – From Sipser's book

1.30 Describe the error in the following "proof" that 0^*1^* is not a regular language. (An error must exist because 0^*1^* *is* regular.) The proof is by contradiction. Assume that 0^*1^* is regular. Let p be the pumping length for 0^*1^* given by the pumping lemma. Choose s to be the string 0^p1^p. You know that s is a member of 0^*1^*, but Example 1.73 shows that s cannot be pumped. Thus you have a contradiction. So 0^*1^* is not regular.