Computability and Complexity COSC 4200

The Pumping Lemma and Nonregular Languages

Suppose that A is accepted by a DFA $M=(Q,\Sigma,\delta,q_0,F)$. Let p=|Q| be the number of states in M. Let n=p+1, and let $w=0^n1^n$. Then M accepts w.

Suppose that A is accepted by a DFA $M=(Q,\Sigma,\delta,q_0,F)$. Let p=|Q| be the number of states in M. Let n=p+1, and let $w=0^n1^n$. Then M accepts w.

Let's look at the sequence of states that M visits on w. Let q_i be the state M is in after 0^i .

$$\xrightarrow{q_0} \xrightarrow{0} (q_1) \xrightarrow{0} (q_2) \xrightarrow{0} (q_{n-1}) \xrightarrow{0} (q_n) \xrightarrow{1^n} \xrightarrow{f}$$

Suppose that A is accepted by a DFA $M=(Q,\Sigma,\delta,q_0,F)$. Let p=|Q| be the number of states in M. Let n=p+1, and let $w=0^n1^n$. Then M accepts w.

Let's look at the sequence of states that M visits on w. Let q_i be the state M is in after 0^i .

$$\xrightarrow{q_0} \xrightarrow{0} \xrightarrow{q_1} \xrightarrow{0} \xrightarrow{q_2} \xrightarrow{0} \xrightarrow{q_{n-1}} \xrightarrow{0} \xrightarrow{q_n} \xrightarrow{0} \xrightarrow{q_n} \xrightarrow{1} \xrightarrow{r} \xrightarrow{r} \xrightarrow{r}$$

Since n > p, M must visit some state twice while reading the 0's.

Let
$$A = \{0^n 1^n \mid n \ge 0\}$$
. Is A regular?

Suppose that A is accepted by a DFA $M=(Q,\Sigma,\delta,q_0,F)$. Let p=|Q| be the number of states in M. Let n=p+1, and let $w=0^n1^n$. Then M accepts w.

Let's look at the sequence of states that M visits on w. Let q_i be the state M is in after 0^i .

$$\xrightarrow{q_0} \xrightarrow{0} \xrightarrow{q_1} \xrightarrow{0} \xrightarrow{q_2} \xrightarrow{0} \xrightarrow{q_{n-1}} \xrightarrow{0} \xrightarrow{q_n} \xrightarrow{0} \xrightarrow{q_n} \xrightarrow{1} \xrightarrow{r} \xrightarrow{r} \xrightarrow{r}$$

Since n > p, M must visit some state twice while reading the 0's. \longrightarrow For some i, j with $1 \le i < j \le n$, we must have $q_i = q_j$.

$$\xrightarrow{q_0} \xrightarrow{0} \xrightarrow{q_1} \xrightarrow{0} \xrightarrow{q_2} \xrightarrow{0} \xrightarrow{q_{n-1}} \xrightarrow{0} \xrightarrow{q_n} \xrightarrow{1^n} \xrightarrow{f}$$

For some i, j with $1 \le i < j \le n$, we must have $q_i = q_j$.

$$\delta^*(q_0, 0^i) = q_i \qquad \qquad \delta^*(q_0, 0^j) = q_i$$

$$\delta^*(q_i, 0^{j-i}) = q_i \qquad \qquad \delta^*(q_i, 0^{n-j}1^n) \in F$$

$$0^{j-i}$$

$$q_i = q_j$$

$$0^{n-j}1^n$$

$$\longrightarrow \overbrace{q_0} \xrightarrow{0} \overbrace{q_1} \xrightarrow{0} \overbrace{q_2} \xrightarrow{0} \xrightarrow{0} \overbrace{q_{n-1}} \xrightarrow{0} \overbrace{q_n} \xrightarrow{1} \xrightarrow{n} \xrightarrow{r} \underbrace{f}$$

For some i, j with $1 \le i < j \le n$, we must have $q_i = q_j$.

$$\delta^*(q_0, 0^i) = q_i \qquad \delta^*(q_0, 0^j) = q_i$$

$$\delta^*(q_i, 0^{j-i}) = q_i \qquad \delta^*(q_i, 0^{n-j}1^n) \in F$$

$$0^{j-i}$$

$$q_i = q_j$$

$$0^{n-j}1^n$$

We must also have

$$0^{i} (0^{j-i}) (0^{j-i}) 0^{n-j} 1^{n} = 0^{n+(j-i)} 1^{n}$$

accepted! Therefore M does not accept A.

This technique is formalized as the "Pumping Lemma."

Pumping Lemma

If A is a regular language, then there is a number p (the pumping constant) such that for every $w \in A$ with $|w| \ge p$, w can be divided into three pieces w = xyz satisfying the following conditions:

- For all $i \ge 0$, $xy^iz \in A$.
- |y| > 0.
- $|xy| \leq p$.

This technique is formalized as the "Pumping Lemma."

Pumping Lemma

If A is a regular language, then there is a number p (the pumping constant) such that for every $w \in A$ with $|w| \ge p$, w can be divided into three pieces w = xyz satisfying the following conditions:

- For all $i \ge 0$, $xy^iz \in A$.
- |y| > 0.
- $|xy| \leq p$.

Comments.

- Condition **1** says that xz, xyz, xyyz, xyyyz, ... are all in A.
- Condition ② says that $y \neq \epsilon$. It is possible for x and z to be ϵ .
- Condition 3 says the combined length of x and y is at most p.

Proof. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA that recognizes A. Let p = |Q| be the number of states in M.

Proof. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA that recognizes A. Let p = |Q| be the number of states in M.

Let $w \in A$ have length $|w| = n \ge p$. For any $1 \le i \le j \le n$, let w[i..j] be the substring of the i^{th} through j^{th} symbols of w. For each $1 \le i \le n$, let $q_i = \delta^*(q_0, w[1..i])$. Then $q_n \in F$ since $w \in A$.

Proof. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA that recognizes A. Let p = |Q| be the number of states in M.

Let $w \in A$ have length $|w| = n \ge p$. For any $1 \le i \le j \le n$, let w[i..j] be the substring of the i^{th} through j^{th} symbols of w. For each $1 \le i \le n$, let $q_i = \delta^*(q_0, w[1..i])$. Then $q_n \in F$ since $w \in A$.

Since n+1>p and there are only p states in Q, some state must be repeated in the sequence

$$q_0, q_1, \ldots, q_n$$

of n+1 states. Let $0 \le i < j \le p$ such that $q_i = q_j$.

We now break up w as w = xyz where

$$x = w[1..i]$$

 $y = w[i+1..j]$
 $z = w[j+1..n]$.

Since i < j, |y| > 0, and since $j \le p$, $|xy| \le p$. Therefore conditions (2) and (3) are satisfied.

We now break up w as w = xyz where

$$x = w[1..i]$$

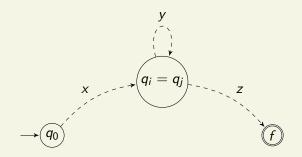
 $y = w[i+1..j]$
 $z = w[j+1..n]$.

Since i < j, |y| > 0, and since $j \le p$, $|xy| \le p$. Therefore conditions (2) and (3) are satisfied.

Then

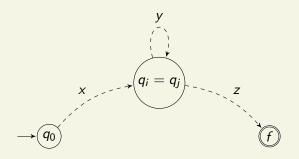
$$\delta^*(q_0, x) = q_i,
\delta^*(q_i, y) = q_j = q_i,
\delta^*(q_i, z) = q_n.$$

We also have $\delta^*(q_i, y^k) = q_i$ for all k (by induction).



Therefore, for any $k \ge 0$,

$$\delta^*(q_0, xy^k) = \delta^*(\underbrace{\delta^*(q_0, x)}_{q_i}, y^k) = q_i,$$



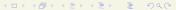
Therefore, for any $k \ge 0$,

$$\delta^*(q_0, xy^k) = \delta^*(\underbrace{\delta^*(q_0, x)}_{q_i}, y^k) = q_i,$$

so

$$\delta^*(q_0, xy^k z) = \delta^*(\underbrace{\delta^*(q_0, xy^k)}_{q_i}, z) = q_n \in F$$

and $xy^kz \in A$.



Proof. Suppose (for sake of argument) that A is regular, and let p be the pumping constant for A. Let $w=0^p1^p$. Then w can be broken into w=xyz where $|xy| \le p$ and |y| > 0 such that $xy^iz \in A$ for all i.

Proof. Suppose (for sake of argument) that A is regular, and let p be the pumping constant for A. Let $w=0^p1^p$. Then w can be broken into w=xyz where $|xy|\leq p$ and |y|>0 such that $xy^iz\in A$ for all i.

Then we know that $x = 0^k$, $y = 0^l$, and $z = 0^{p-k-l}1^p$ for some $k \ge 0$, l > 0.

Let i = 2. Then

$$xy^{2}z = 0^{k}0^{2l}0^{p-k-l}1^{p}$$
$$= 0^{p+l}1^{p}$$

should be in A according to the Pumping Lemma.

Proof. Suppose (for sake of argument) that A is regular, and let p be the pumping constant for A. Let $w=0^p1^p$. Then w can be broken into w=xyz where $|xy|\leq p$ and |y|>0 such that $xy^iz\in A$ for all i.

Then we know that $x = 0^k$, $y = 0^l$, and $z = 0^{p-k-l}1^p$ for some $k \ge 0$, l > 0.

Let i = 2. Then

$$xy^2z = 0^k 0^{2l} 0^{p-k-l} 1^p$$

= $0^{p+l} 1^p$

should be in A according to the Pumping Lemma. But since l > 0, $p + l \neq p$, so $xy^2z \notin A$. This is a contradiction, so A must not be regular.

 Assume that A is regular, and let p be the pumping constant for A.

- Assume that A is regular, and let p be the pumping constant for A.
- **2** Choose a string $w \in A$ with $|w| \ge p$.

- Assume that A is regular, and let p be the pumping constant for A.
- **2** Choose a string $w \in A$ with $|w| \ge p$.
- **3** Let xyz = w be any way of breaking w into three pieces such that $y \neq \epsilon$ and $|xy| \leq p$.

- Assume that A is regular, and let p be the pumping constant for A.
- **2** Choose a string $w \in A$ with $|w| \ge p$.
- **3** Let xyz = w be any way of breaking w into three pieces such that $y \neq \epsilon$ and $|xy| \leq p$.
- Find some $i \ge 0$ such that $xy^iz \notin A$.

- Assume that A is regular, and let p be the pumping constant for A.
- ② Choose a string $w \in A$ with $|w| \ge p$.
- **1** Let xyz = w be any way of breaking w into three pieces such that $y \neq \epsilon$ and $|xy| \leq p$.
- Find some $i \ge 0$ such that $xy^iz \notin A$.
- **5** Conclude that *A* is not regular.

- Assume that A is regular, and let p be the pumping constant for A.
- **2** Choose a string $w \in A$ with $|w| \ge p$.
- **3** Let xyz = w be any way of breaking w into three pieces such that $y \neq \epsilon$ and $|xy| \leq p$.
- Find some $i \ge 0$ such that $xy^iz \notin A$.
- **5** Conclude that *A* is not regular.

Note:

- We get to choose w and i in steps 2 and 4.
- We do not get to choose p and x, y, z in steps 1 and 3.
 We must argue for any possibility of p and x, y, z.

Contrapositive Form of the Pumping Lemma

Original Form:

```
If A is regular, then there is some constant p such that for every w \in A with |w| \ge p there exist x, y, z with w = xyz, |xy| \le p, |y| > 0 such that for every i \ge 0, xy^iz \in A,
```

Contrapositive Form of the Pumping Lemma

Original Form:

```
If A is regular, then there is some constant p such that for every w \in A with |w| \ge p there exist x, y, z with w = xyz, |xy| \le p, |y| > 0 such that for every i \ge 0, xy^iz \in A,
```

Contrapositive Form:

```
If for every constant p there is some w \in A with |w| \ge p such that for every x, y, z with w = xyz, |xy| \le p, |y| > 0 there is some i \ge 0 such that xy^iz \notin A,
```

then A is not regular.



Proof. Suppose that A is regular, and let p be the pumping constant for A.

Let $w = 0^{p^2}$. Then $|w| \ge p$ and $w \in A$. Let w = xyz where $|xy| \le p$ and |y| > 0.

Proof. Suppose that A is regular, and let p be the pumping constant for A.

Let $w = 0^{p^2}$. Then $|w| \ge p$ and $w \in A$. Let w = xyz where $|xy| \le p$ and |y| > 0.

Then we know that $x = 0^k$, $y = 0^l$, and $z = 0^{p^2 - k - l}$ for some $k \ge 0$, l > 0.

Proof. Suppose that A is regular, and let p be the pumping constant for A.

Let $w=0^{p^2}$. Then $|w|\geq p$ and $w\in A$. Let w=xyz where $|xy|\leq p$ and |y|>0.

Then we know that $x = 0^k$, $y = 0^l$, and $z = 0^{p^2 - k - l}$ for some $k \ge 0$, l > 0.

Let i = 2. We claim that

$$xy^2z = 0^k 0^{2l} 0^{p^2 - k - l}$$

= $0^{p^2 + l} \notin A$.

We must show that $p^2 + I$ is not a perfect square.

We will show $p^2 + I$ is not a perfect square by showing it lies strictly between two consecutive perfect squares. We have

$$(p+1)^2 = p^2 + 2p + 1$$

We will show $p^2 + I$ is not a perfect square by showing it lies strictly between two consecutive perfect squares. We have

$$(p+1)^2 = p^2 + 2p + 1$$

> $p^2 + p$

We will show $p^2 + I$ is not a perfect square by showing it lies strictly between two consecutive perfect squares. We have

$$(p+1)^2 = p^2 + 2p + 1$$

$$> p^2 + p$$

$$\geq p^2 + l$$

We will show $p^2 + I$ is not a perfect square by showing it lies strictly between two consecutive perfect squares. We have

$$(p+1)^2 = p^2 + 2p + 1$$

> $p^2 + p$
 $\geq p^2 + l$
> p^2 ,

We will show $p^2 + I$ is not a perfect square by showing it lies strictly between two consecutive perfect squares. We have

$$(p+1)^2 = p^2 + 2p + 1$$

> $p^2 + p$
 $\geq p^2 + l$
> p^2 ,

so

$$p^2 < p^2 + l < (p+1)^2$$
.

and $p^2 + I$ is not a perfect square.

We will show $p^2 + I$ is not a perfect square by showing it lies strictly between two consecutive perfect squares. We have

$$(p+1)^2 = p^2 + 2p + 1$$

> $p^2 + p$
 $\geq p^2 + l$
> p^2 ,

so

$$p^2 < p^2 + l < (p+1)^2$$
.

and $p^2 + I$ is not a perfect square. Therefore A is not regular by the pumping lemma.

Example. $E = \{0^i 1^j \mid i \ge j\}$ is not regular.

Proof. Suppose that E is regular, and let p be the pumping constant for E. Let $w=0^p1^p$. Let w=xyz where $|xy|\leq p$ and |y|>0.

Example. $E = \{0^i 1^j \mid i \ge j\}$ is not regular.

Proof. Suppose that E is regular, and let p be the pumping constant for E. Let $w=0^p1^p$. Let w=xyz where $|xy|\leq p$ and |y|>0.

Then we know that $x = 0^k$, $y = 0^l$, and $z = 0^{p-k-l}1^p$ for some $k \ge 0$, l > 0.

Example. $E = \{0^i 1^j \mid i \ge j\}$ is not regular.

Proof. Suppose that E is regular, and let p be the pumping constant for E. Let $w=0^p1^p$. Let w=xyz where $|xy|\leq p$ and |y|>0.

Then we know that $x=0^k$, $y=0^l$, and $z=0^{p-k-l}1^p$ for some $k \ge 0$, l > 0.

Let i = 0. Then

$$xy^0z = 0^k 0^{p-k-l} 1^p$$

= $0^{p-l} 1^p \notin E$

because l > 0. Therefore E is not regular by the pumping lemma.

Proof. Suppose B is regular, and let p be the pumping constant for B.

Let $w = 0^p 10^p 1$.

Proof. Suppose B is regular, and let p be the pumping constant for B.

Let $w=0^p10^p1$. Then $w\in B$ and $|w|\geq p$. Let xyz=w such that $|xy|\leq p$ and |y|>0.

Proof. Suppose B is regular, and let p be the pumping constant for B.

Let $w = 0^p 10^p 1$. Then $w \in B$ and $|w| \ge p$. Let xyz = w such that $|xy| \le p$ and |y| > 0.

We must have have $x = 0^k$, $y = 0^l$, and $z = 0^{p-k-l}10^p1$ for some $k \ge 0$, l > 0.

Proof. Suppose B is regular, and let p be the pumping constant for B.

Let $w = 0^p 10^p 1$. Then $w \in B$ and $|w| \ge p$. Let xyz = w such that $|xy| \le p$ and |y| > 0.

We must have have $x = 0^k$, $y = 0^l$, and $z = 0^{p-k-l}10^p1$ for some $k \ge 0$, l > 0.

Then

$$xy^2z = 0^{p+l}10^p1 \notin B,$$

so B is not regular by the pumping lemma.



We can also use closure properties to prove nonregularity.

Example. $A = \{w \in \{0,1\}^* \mid \#_0(w) = \#_1(w)\}$ is not regular.

We can also use closure properties to prove nonregularity.

Example. $A = \{w \in \{0,1\}^* \mid \#_0(w) = \#_1(w)\}$ is not regular. **Proof.** Suppose A is regular. Then since the regular languages are closed under intersection,

$$A \cap 0^*1^* = \{0^n1^n \mid n \ge 0\}$$

is also regular.

We can also use closure properties to prove nonregularity.

Example. $A = \{w \in \{0,1\}^* \mid \#_0(w) = \#_1(w)\}$ is not regular. **Proof.** Suppose A is regular. Then since the regular languages are closed under intersection,

$$A \cap 0^*1^* = \{0^n1^n \mid n \ge 0\}$$

is also regular. This is a contradiction since we already proved $\{0^n1^n\mid n\geq 0\}$ is not regular.



Proof. Suppose ADD is regular, and let p be the pumping constant for ADD.

Proof. Suppose ADD is regular, and let p be the pumping constant for ADD.

Let w be the string " $1^p + 1 = 10^p$ ".

Proof. Suppose ADD is regular, and let p be the pumping constant for ADD.

Let w be the string " $1^p + 1 = 10^p$ ". Then $w \in ADD$ and $|w| \ge p$. Let xyz = w such that $|xy| \le p$ and |y| > 0.

We must have $x = 1^k$, $y = 1^l$, and z is the string " $1^{p-k-l} + 1 = 10^p$ ".

Proof. Suppose ADD is regular, and let p be the pumping constant for ADD.

Let w be the string " $1^p + 1 = 10^p$ ". Then $w \in ADD$ and $|w| \ge p$. Let xyz = w such that $|xy| \le p$ and |y| > 0.

We must have $x = 1^k$, $y = 1^l$, and z is the string " $1^{p-k-l} + 1 = 10^p$ ".

Then

$$xy^2z = "1^{p+l} + 1 = 10^p" \notin ADD$$

because
$$1^{p+l} + 1 = 10^{p+l} \neq 10^p$$
 since $l > 0$.



The Pumping Lemma Does Not Always Apply

The Pumping Lemma applies in most cases, but there are nonregular languages that satisfy the conclusion of the pumping lemma. Here is one example.

Let
$$F = \{a^i b^j c^k \mid i, j, k \ge 0 \text{ and if } i = 1 \text{ then } j = k\}.$$

The Pumping Lemma Does Not Always Apply

The Pumping Lemma applies in most cases, but there are nonregular languages that satisfy the conclusion of the pumping lemma. Here is one example.

Let
$$F = \{a^i b^j c^k \mid i, j, k \ge 0 \text{ and if } i = 1 \text{ then } j = k\}.$$

Suppose F is regular. Then

$$F \cap ab^*c^* = \{ab^nc^n \mid n \ge 0\}$$

is also regular by closure under intersection. This language is not regular by a proof similar to the one we wrote for $\{0^n1^n\mid n\geq 0\}$. This is a contradiction, so F is not regular.

$$F = \{a^m b^n c^l \mid m, n, l \ge 0 \text{ and if } m = 1 \text{ then } n = l\}$$

$$F = \{a^m b^n c^l \mid m, n, l \ge 0 \text{ and if } m = 1 \text{ then } n = l\}$$

Define p=1. Let $w=a^mb^nc^l\in F$ have $|w|\geq p$. We consider three cases.

$$F = \{a^m b^n c^l \mid m, n, l \ge 0 \text{ and if } m = 1 \text{ then } n = l\}$$

Define p = 1. Let $w = a^m b^n c^l \in F$ have $|w| \ge p$. We consider three cases.

• If m=1, then $w=ab^nc^n$. Let $x=\epsilon$, y=a, and $z=b^nc^n$. Then for all $i\geq 0$,

$$xy^iz=a^ib^nc^n\in F.$$

$$F = \{a^m b^n c^l \mid m, n, l \ge 0 \text{ and if } m = 1 \text{ then } n = l\}$$

Define p = 1. Let $w = a^m b^n c^l \in F$ have $|w| \ge p$. We consider three cases.

• If m=1, then $w=ab^nc^n$. Let $x=\epsilon$, y=a, and $z=b^nc^n$. Then for all $i\geq 0$,

$$xy^iz=a^ib^nc^n\in F.$$

• If $m \ge 2$, then let $x = \epsilon$, y = aa, and $z = a^{m-2}b^nc^l$. For all $i \ge 0$, $xy^iz = a^{m+2i}b^nc^l \in F.$



$$F = \{a^m b^n c^l \mid m, n, l \ge 0 \text{ and if } m = 1 \text{ then } n = l\}$$

Define p = 1. Let $w = a^m b^n c^l \in F$ have $|w| \ge p$. We consider three cases.

• If m=1, then $w=ab^nc^n$. Let $x=\epsilon$, y=a, and $z=b^nc^n$. Then for all $i\geq 0$,

$$xy^iz=a^ib^nc^n\in F.$$

- If $m \ge 2$, then let $x = \epsilon$, y = aa, and $z = a^{m-2}b^nc^l$. For all $i \ge 0$, $xy^iz = a^{m+2i}b^nc^l \in F.$
- If m = 0, then $w = b^n c^l$. There are two subcases:
 - If i > 0, then let $x = \epsilon$, y = b, and $z = b^{n-1}c^{i}$.
 - If i = 0, then let $x = \epsilon$, y = c, and $z = c^{l-1}$.

In either subcase, $xy^iz \in F$ for all i > 0.

