The Pumping Lemma

Mark Greenstreet, CpSc 421, Term 1, 2006/07

Lecture Outline

Regular Expressions

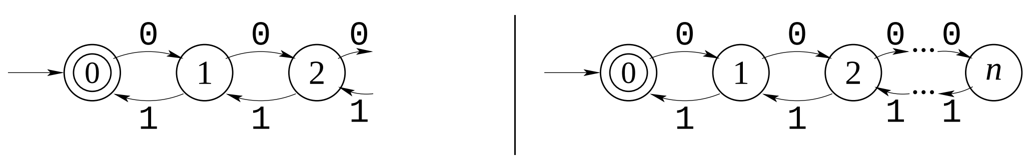
- An Example
- The Pumping Lemma

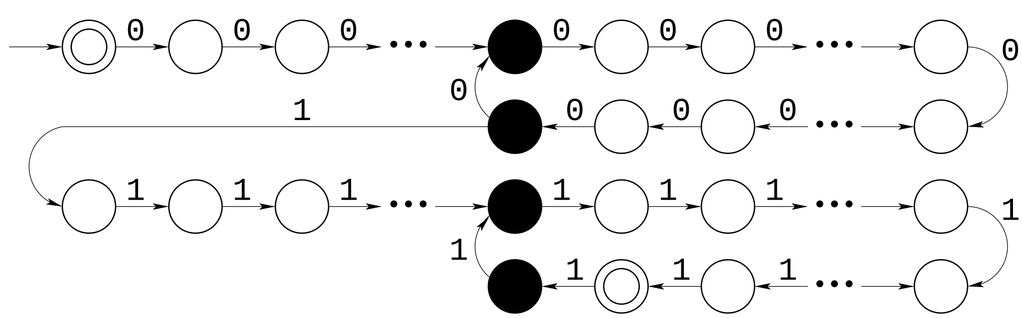
An Example

- Let $A = \{w \in \{0,1\}^* \mid \exists n \in \mathbb{N}.\ w = 0^n 1^n\}$.
- We abbreviate this as $A = 0^n 1^n$.
- Is A regular?

A DFA for $0^n 1^n$?

How about?





0^n1^n is not Regular

Proof (by contradiction):

- Let $A = 0^n1^n$.
- Suppose A were regular.
  Then there would be a DFA, $M = (Q, \{0,1\}, \delta, q_0, F)$ that recognizes A.
- Let $p = |Q|$. Note that $0^p1^p \in A$.
- M visits $p+1$ states (including the start state) when reading 0^p.
  Therefore, it visits at least one state twice.
- Let i_1, i_2 and j be integers such that:

$$\begin{array}{lcccccc} & 0 & \leq & i_1 & < & i_2 & \leq & p \\ \text{and} & \delta(q_0, 0^{i_1}) & = & \delta(q_0, 0^{i_2}) & = & q_j. \end{array}$$

  In English, the M completes a loop when reading the $i_2 - i_1$ `0`'s that follow 0^{i+1}.

- Let $i_{loop} = i_2 - i_1$, and note that $i_{loop} > 0$

0^n1^n is not Regular (end of proof)

- From the previous slide:

$$\begin{array}{rcll} M & = & (Q, \{0,1\}, \delta, q_0, F), & \text{a DFA} \\ p & = & |Q|, & \text{number of states of } M \\ \multicolumn{3}{c}{0 \leq i_1 < i_2 < p, j : \delta(q_0, 0^{i_1}) = \delta(q_0, i_2)\ ,} & \text{the “loop”} \\ i_{loop} & = & i_2 - i_1, & \text{the length of the loop} \end{array}$$

- We can make the machine take an extra lap around the loop, and it will still go to the same final state. In math,

$$\delta(q_0, 0^p0^{i_{loop}}1^p) \quad = \quad \delta(q_0, 0^p1^p) \quad \in \quad F$$

- This means that M accepts $0^p0^{i_{loop}}1^p$. But $0^p0^{i_{loop}}1^p \notin A$.
  Thus, M does not recognize A.
- We can apply this argument to any DFA.
  Therefore, there is no DFA that recognizes A.
  This proves that A is not regular.

  □

The Pumping Lemma

Let A be a regular language.

- There exists some integer p such that for any string w in language A with $|w| \geq p$ ...
- ... we can find strings x, y, and z such that $w = xyz$ and
  - $\forall i \geq 0.\ xy^iz \in A$,
  - $|y| > 0$, and
  - $|xy| \leq p$.
- The intuition behind the pumping lemma is that:
  - y is a string that takes a DFA that recognizes A through a cycle of states (i.e. a loop).
  - If $|w|$ is greater than the number of states in a DFA that recognizes A, then the DFA must visit the some state more than once when reading w. This provides the cycle.

Pumping Lemma: Some Definitions

- Given a regular language A,
- let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA that recognizes A.
- Let $p = |Q|$.
- Let w be any string in A with $|w| \geq p$.
- Let

$$\begin{aligned}
\textit{prefix}(w, n) &= \text{the first } n \text{ symbols of } w \\
&= w_0 \cdot w_2 \cdots w_{n-1} \\
&\quad \text{if } n \geq |w|, \text{ then } \textit{prefix}(w, n) = w \\
&\quad \text{if } n \leq 0, \text{ then } \textit{prefix}(w, n) = \epsilon \\
\textit{suffix}(w, n) &= \text{the string } s \text{ such that } w = \textit{prefix}(w, n) \cdot s \\
&= w_n \cdot w_{n+1} \cdots w_{|w|-1} \\
&\quad \text{if } n \geq |w|, \text{ then } \textit{suffix}(w, n) = \epsilon \\
&\quad \text{if } n \leq 0, \text{ then } \textit{suffix}(w, n) = w \\
\textit{substring}(w, n_1, n_2) &= \textit{prefix}(\textit{suffix}(w, n_2), n_2 - n_1) \\
&= w_{n_1} \cdot w_{n_1+1} \cdots w_{n_2-1}
\end{aligned}$$

Pumping Lemma: The Proof

- Note that M must visit some state twice by the time it has read $\mathit{prefix}(w, p)$. This is because M only has p states, and it has visited $p + 1$ states (including the start state) by the time it reads $\mathit{prefix}(w, p)$.
- Let $0 \leq i_1 < i_2 \leq p$ be integers such that

$$\delta(q_0, \mathit{prefix}(w, i_1)) \quad = \quad \delta(q_0, \mathit{prefix}(w, i_2))$$

- Now, let

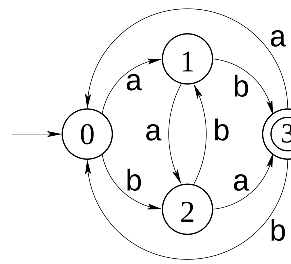
$$\begin{aligned} x &= \mathit{prefix}(w, i_1) \\ y &= \mathit{substring}(w, i_1, i_2) \\ z &= \mathit{suffix}(w, i_2) \end{aligned}$$

- We have
  - $w = xyz$: by the definitions of x, y, and z.
  - $xy^iz \in A$: see the next slide.
  - $|y| > 0$: $i_1 < i_2$.
  - $|xy| \leq p$: $i_2 \leq p$.
- Thus, the claims of the pumping lemma are satisfied.

Proof that $xy^iz \in A$

- Let $q_j = \delta(q_0, x)$.
- $q_j = \delta(q_0, xy) = \delta(\delta(q_0, x), y) = \delta(q_j, y)$.
  In short, $\delta(q_j, y) = q_j$.
- $\delta(q_0, xy^i) = q_j$, by induction on i:
  - Base case, $i = 0$: $\delta(q_0, xy^0) \; = \; \delta(q_0, x) \; = \; q_j$.
  - Induction step, assume for i, prove for $i + 1$:
    $\delta(q_0, xy^{i+1}) \; = \; \delta(q_0, xy^iy) \; = \; \delta(\delta(q_0, xy^i), y) \; = \; \delta(q_j, y) \; = \; q_j$
- $\delta(q_0, xy^iz) = \delta(q_j, z) = \delta(\delta(q_0, xy), z) \in A$.
- $\square$

Pumping Lemma: Example



- Let $M = (Q, \Sigma, \delta, q_0, F)$ be the DFA shown above, with $Q = \{0, 1, 2, 3\}$, $\Sigma = \{\mathtt{a}, \mathtt{b}\}$, $q_0 = 0$, and $F = \{3\}$.
- Let $A = L(M)$. Let $p = |Q| = 4$.
- Let $w = \mathtt{aabaa}$. Note that $w \in A$.
- We can show that the claims of the pumping lemma are satisfied by choosing $x = \mathtt{a}$, $y = \mathtt{ab}$ and $z = \mathtt{aa}$.
  - $\forall i.\ xy^iz \in A$.
  - $|y| = 2 > 0$.
  - $|xy| = 3 < 4 = p$.

Using the Pumping Lemma

- Typically, we use the pumping lemma to show that a language is not regular.
- To do so, we use the contrapositive of the pumping lemma:
  - If it is not possible to choose an integer p such that for any string $w \in A$ there are strings x, y, z such that
    - $w = xyz$,
    - $\forall i.\ xy^iz \in A$,
    - $|y| > 0$, and
    - $|xy| \leq p$.
  - then A is not a regular language.
  - Note that p is chosen first, and then w can be chosen according to the choice of p.
  - Typically, we find a w (depending on the choice of p) such that there is no way to break w into x, y, and z such that $\forall i.\ xy^iz \in A$.
  - Often, the counterexample uses $i = 2$ or $i = 0$.

The Pumping Game

- We can see this as a game between Alice and Bob. Alice wants to show that language A is not regular, and Bob wants to thwart her.
- Bob has to make the first move by stating a value for p.
- Based on the value for p, Alice puts forward a string $w \in A$.
- Bob now gives strings x, y, and z such that $w = xyz$, $|y| > 0$, and $|xy| \leq p$.
- If Alice can find a value for i such that $xy^i z \notin A$, then Alice wins. ☺ Otherwise, Bob wins. ☹

One More Example

- Let $A = 1^p$ where p is a prime number.

A Few Remarks

- WARNING: There are non-regular languages that satisfy the pumping lemma. For example,

$$\begin{aligned} \Sigma &= \{\mathsf{a}, \mathsf{b}, \mathsf{c}\} \\ A &= (\mathsf{a}\mathsf{a}^*\mathsf{c})^n(\mathsf{b}\mathsf{b}^*\mathsf{c})^n \cup \Sigma^*\mathsf{c}\mathsf{c}\Sigma^* \end{aligned}$$

  The language A is not regular, but it satisfies the conditions of the pumping lemma.

- Satisfying the conditions of the pumping lemma is a necessary but not sufficient condition for showing that a language is regular.
- If A is finite (i.e. $|A|$ is finite), then A trivially satisfies the pumping lemma. Let

$$p = 1 + \max_{w \in A} |w|$$

  There are no strings in A with length at least p, and the conditions of the pumping lemma are (vacuously) satisfied.