Non-regular Languages §1.4

Which languages cannot be recognized by finite automata?

Example: $L = \{ 0^n1^n | n \in \mathbb{N} \}$

- ‘Playing around’ with FA convinces you that the ‘finiteness’ of FA is problematic for “all $n \in \mathbb{N}$”
- The problem occurs between the $0^n$ and the $1^n$
- Informal: the memory of a FA is limited by the the number of states $|Q|$
Proving non-regularity

- Prove a general statement -- NO DFA exists for a given problem.
- Cannot assume an automaton structure or a specific strategy
- Need an argument that holds for ALL DFA’s

Repeating DFA Paths

Consider an accepting DFA M with size $|Q|$
On a string of length $p$, $p+1$ states get visited
For $p \geq |Q|$, there must be a $j$ such that the computational path looks like: $q_1, \ldots, q_j, \ldots, q_j, \ldots, q_k$
Repeating DFA Paths

The action of the DFA in $q_j$ is always the same. If we repeat (or ignore) the $q_j, \ldots, q_j$ part, the new path will again be an accepting path.

Proof by contradiction:

• Assume that $L$ is regular
• Hence, there is a DFA $M$ that recognizes $L$
• For strings of length $\geq |Q|$ the DFA $M$ has to ‘repeat itself’
• Show that $M$ will accept strings outside $L$
• Conclude that the assumption was wrong

Note that we use the simple DFA, not the more elaborate (but equivalent) NFA or GNFA.
Pumping Lemma (Thm 1.37)

For every regular language $L$, there is a pumping length $p$, such that for any string $s \in L$ and $|s| \geq p$, we can write $s = xyz$ with

1) $x \cdot y^i \cdot z \in L$ for every $i \in \{0, 1, 2, \ldots\}$
2) $|y| \geq 1$
3) $|xy| \leq p$

Note that 1) implies that $xz \in L$
2) says that $y$ cannot be the empty string $\varepsilon$
Condition 3) is not always used

Formal Proof of Pumping Lemma

Let $M = (Q, \Sigma, \delta, q_1, F)$ with $Q = \{q_1, \ldots, q_p\}$
Let $s = s_1 \ldots s_n \in L(M)$ with $|s| = n \geq p$
Computational path of $M$ on $s$ is the sequence $r_1, \ldots, r_{n+1} \in Q^{n+1}$ with
$r_1 = q_1, r_{n+1} \in F$ and $r_{t+1} = \delta(r_t, s_t)$ for $1 \leq t \leq n$
Because $n+1 \geq p+1$, there are two states such that $r_j = r_k$ (with $j < k$ and $k \leq p+1$)
Let $x = s_1 \ldots s_{j-1}$, $y = s_j \ldots s_{k-1}$, and $z = s_k \ldots s_{n+1}$
$x$ takes $M$ from $q_i = r_1$ to $r_j$, $y$ takes $M$ from $r_j$ to $r_j$,
and $z$ takes $M$ from $r_j$ to $r_{n+1} \in F$
As a result: $x \cdot y^i \cdot z$ takes $M$ from $q_1$ to $r_{n+1} \in F$ ($i \geq 0$)
Formal Proof of Pumping Lemma

Let $M = (Q, \Sigma, \delta, q_1, F)$ with $Q = \{q_1, \ldots, q_p\}$
Let $s = s_1 \ldots s_n \in L(M)$ with $|s| = n \geq p$
Computational path of $M$ on $s$ is the sequence $r_1, \ldots, r_{n+1} \in Q^{n+1}$ with $r_1 = q_1$, $r_{n+1} \in F$ and $r_{t+1} = \delta(r_t, s_t)$ for $1 \leq t \leq n$
Because $n+1 \geq p+1$, there are two terms such that $r_j = r_k$ (with $j < k$ and $k \leq p+1$)
Let $x = s_1 \ldots s_{j-1}$, $y = s_j \ldots s_{k-1}$, and $z = s_k \ldots s_{n+1}$
x takes $M$ from $q_1$ to $r_j$, y takes $M$ from $r_j$ to $r_j$, and z takes $M$ from $r_j$ to $r_{n+1} \in F$
As a result, $x y z \in L(M)$ for every $i \in \{0, 1, 2, \ldots\}$

Pumping $0^n 1^n$ (Ex. 1.38)

Assume that $B = \{0^n 1^n \mid n \geq 0\}$ is regular
Let $p$ be the pumping length, and $s = 0^p 1^p \in B$
By P.L.: $s = xyz = 0^p 1^p$, with $xy^i z \in B$ for all $i \geq 0$
Three options for $y$:
1) $y = 0^k$, hence $xyz = 0^{p+k} 1^p \not\in B$
2) $y = 1^k$, hence $xyz = 0^p 1^{k+p} \not\in B$
3) $y = 0^k 1^l$, hence $xyz = 0^{p-k} 0^k 1^l 0^k 1^l 1^{p-l} \not\in B$
Contradiction! So the language $B$ is not regular.
Another example

\[ F = \{ \text{ww} \mid w \in \{0,1\}^* \} \text{ (Ex. 1.40)} \]

Let \( p \) be the pumping length, and take \( s = 0^p10^p1 \)
Let \( s = xyz = 0^p10^p1 \) with condition 3) \(|xy| \leq p\)
Only one option: \( y = 0^k \), with \( xyyz = 0^{p+k}10^p1 \not\in F \)

Without 3) this would have been a pain.

Another Strategy: Intersecting Regular Languages

Let \( C = \{ w \mid \# \text{ of 0s in } w \text{ equals } \# \text{ of 1s in } w \} \)
Problem: If \( xyz \in C \) with \( y \in C \), then \( xy^iz \in C \)
Idea: If \( C \) is regular and \( F \) is regular, then the intersection \( C \cap F \) has to be regular as well

Solution: Assume that \( C \) is regular
Take the regular \( F = \{ 0^n1^m \mid n,m \in \mathbb{N} \} \), then for the intersection: \( C \cap F = \{ 0^n1^n \mid n \in \mathbb{N} \} \)
But we know that \( C \cap F \) is not regular
Conclusion: \( C \) is not regular
Pumping Down $E = \{ 0^i1^j \mid i \geq j \}$

Problem: ‘pumping up’ $s=0^p1^p$ with $y=0^k$ gives $x\alpha\beta\gamma = 0^{p+k}1^p$, $x\alpha^3\beta \gamma = 0^{p+2k}1^p$, which are all in $E$ (hence do not give contradictions)
Solution: pump down to $xz = 0^{p-k}1^p$.
Overall for $s = xyz = 0^p1^p$ (with $|xy| \leq p$):

$y=0^k$, hence $xz = 0^{p-k}1^p \not\in E$

Contradiction! So $E$ is not regular

Pumping lemma usage - steps

• You are given a pumping number
• You choose a string
• You are told $x,y,z$ (satisfying some criteria)
• You choose $i$ in $xy^iz$, and show it violates criterion of set for that $i$. 
Alternatives for proving non-regularity

• Simpler technique (not in the text)
  – Based on the Myhill-Nerode Theorem
  – less general