[Lecture Tue March 31, 2015] "Proof Scripts"

Myhill-Nerode Theorem: Suppose we can establish that (1) [for a given language L] (1)

(Part I)

THERE EXISTS an infinite set $S \subseteq \Sigma^*$ such that

FOR ALL pairs $x, y \in S$ $(x \neq y)$

THERE EXISTS $z \in \Sigma^*$ such that $L(xz) \neq L(yz)$.

$xz \in L \wedge yz \notin L$ or $xz \notin L \wedge yz \in L$.

THEN L is not regular.

Let

The general corresponding "Proof Script":

Take $S =$ _____. S is infinite (because ....)

Let any $x, y \in S$, $x \neq y$, be given. (Relate to your choice of S)

Then we can write x, y helpfully as $x =$ ____ and $y =$ ____ where _____

Take $z =$ _____. Then $L(xz) \neq L(yz)$ because _____

∴ S is an infinite pairwise distinguishing set for L, so by MNT L is not regular. ☒

Same script for $L = \{a^n b^n : n \geq 0\}$ Easy to modify for $L' = \{a^N x : x \neq b^N\}$.

Example: $L = \{x \in \{a,b\}^* : \#a(x) = \#b(x)\}$. $S = \{a^n : n \geq 0\}$.

Take $S = a^*$. Clearly S is infinite.

Let any $x, y \in S$, $x \neq y$, be given. Then there are natural numbers $m, n \geq 0$, $m \neq n$, such that $x = a^m$ and $y = a^n$.

Take $z = b^m$. Then $xz = a^m b^m \in L$ ($\notin L'$), but $yz = a^n b^m \notin L$ ($\in L'$) since $m \neq n$. Hence $L(xz) \neq L(yz)$ ($L'(xz) \neq L'(yz)$).

Thus S is PO for L, and since S is infinite, L is not regular by MNT. ☒

$L = \{ x \in \{a,b\}^* : \#a(x) \geq \#b(x) \}$. $x = a^n b^n$ (2)

Example 2. Take $S = a^*$ (Generally: $S = \{$critical first-half conditions for $L\}$)

Clearly S is infinite. Let any $x, y \in S$, $x \neq y$, be given. Then there are natural numbers $m, n \geq 0$ such that $x = a^m$ and $y = a^n$, where without loss of generality (wlog) $m < n$.

Take $z = b^m$. Then $xz = a^m b^n \notin L$ since $m < n$.

But $yz = a^n b^n \in L$. So $L(xz) \neq L(yz)$, etc. L is not regular.

Example 3:

Sometimes we can be "proactive" in the choice of S.

$L = \{ ww : w \in \{a,b\}^* \}$.

$L = \{ x : x$ can be broken as $x = ww$ for some $w \in \Sigma^* \}$

NON-PROOF: Take $S = a^*$. Let $x, y \in S$, $x \neq y$ be given. Then $x = a^m$ & $y = a^n$ where $m \neq n$ (wlog $m < n$). Take $z = a^m$. Then $xz = a^m a^m \in L$ but $yz = a^n a^m$... Not a proof since $a^n a^m$ could be in L too.

Take $S = a^* b$. Clearly S is infinite. Let any $x, y \in S$, $x \neq y$, be given. Then there are $m, n \geq 0$ ($m \neq n$) s.t. $x = a^m b$ and $y = a^n b$.

Take $z = a^m b$. Then $xz = a^m b a^m b \in L$, but $yz = a^n b a^m b \notin L$

because the central b must end the first part of any possible $yz = ww$ breakdown, but $m \neq n$ makes the a^i mismatch, $\therefore S$ is PD. $\therefore$ L is not regular. ⊗

This can work with $S = a^*$, $z = b a^m b$ too, but the above puts the key idea more forward.

Example 4²

$\Sigma = \{a\}$ only! $L = \{a^n : n \text{ is a perfect square} : 0, 1, 4, 9, 16, \ldots\}$

Take $S = a^*$. Clearly S is infinite.

Let any $x, y \in S$, $x \neq y$ be given. Then $x = a^m$, $y = a^n$ where wlog we have $m < n$. Take $z = a^p$. We control p!

Then $xz = a^m a^p = a^{m+p} \in L$ since $m+p$ is a perfect square

and $yz = a^n a^p = a^{n+p} \notin L$ since $n+p$ is not a perfect square.

Whatever the m and n that were given are, their difference $K = n - m$ is also "given" — it is nailed down at that step.

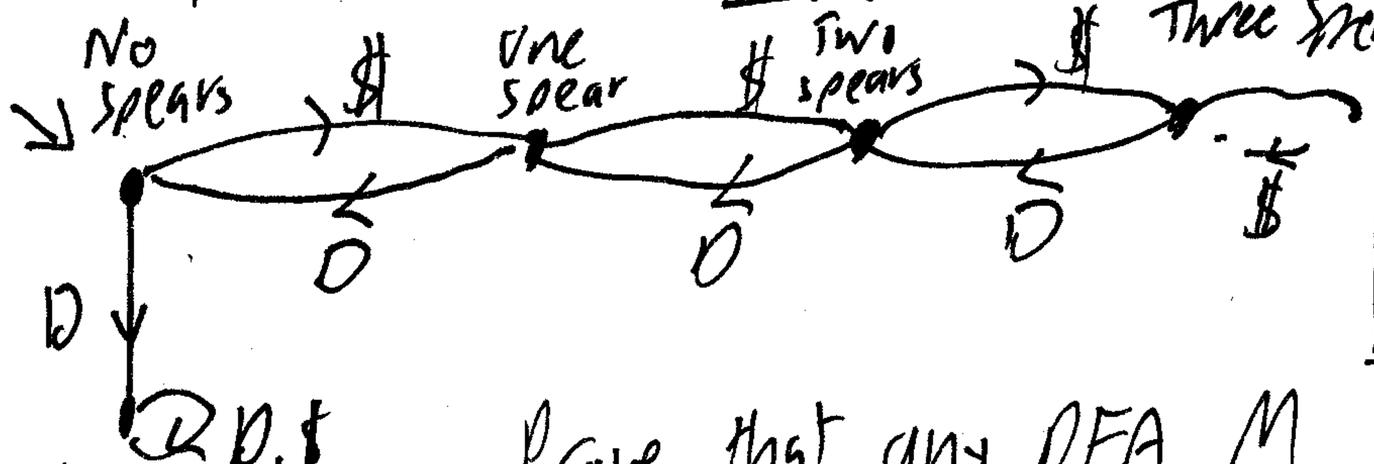
There is — we can find — a square q s.t. $(q+1)^2 > q + K$

and also such that $q > m$. Having found q, we take $p = q - m$.

Then $xz = a^{m+p} = a^q \in L$ since q is a square.

while $yz = a^{n+p} = a^{q+(n-m)} \notin L$ since $q + (n-m) = q + K < (q+1)^2$.

so $n+p$ is not a square. $\therefore S$ is PD for L, $\therefore L$ is not regular.



Example 5: If no limit on saving spears, this becomes a proof that the language becomes non-regular.

For any $K \geq 1$, let A_K stand for the language of my "Dungeon" game where you can save up to K spears. $\Sigma = \{\$, D\}$

- All pairs (D, y) $y \in \$^*$ are distinguished by $z = \varepsilon$.

Prove that any DFA M s.t. $L(M) = A_K$ needs (at least) $K+2$ states.

$S = \{D, \$^0 (\varepsilon), \$, \$^2, \ldots, \$^K\}$ $|S| = K+2$. So we need only show that S is PD for L.

Other pairs $(x, y) \in S$, $x \neq y$ have the form $x = \$^m$, $y = \$^n$ where wlog $m < n$.

Take $z = D^n$. Then $xz = \$^m D^n \notin A_K$ since $m < n$, but $yz = \$^n D^n \in A_K$. $\therefore S$ is PD for A_K.

Example 6 (regular but with "exponential explosion"):

For any $K \geq 1$, consider $L_K = (0+1)^* 1 (0+1)^{K-1}$

$= \{x \in \{0,1\}^* :$ the Kth bit from the right is a $1\}$.

$K=3$: even DFA M s.t. $L(M) = L_3$ needs 8 states.

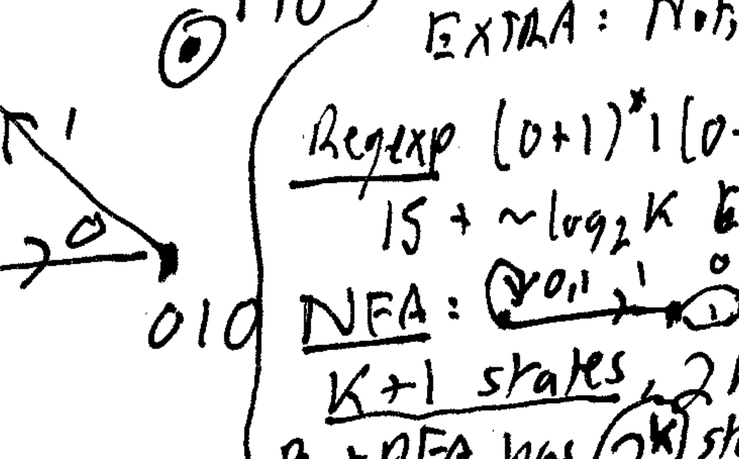
Theorem: For all $K \geq 1$, every DFA M s.t. $L(M) = L_K$ needs at least 2^K states — (and can be done with that number.)

Proof: Take $S = \{0,1\}^K$. Clearly $|S| = 2^K$. Show S is PD. for L_K.

Let any $x, y \in S$, $x \neq y$, be given.

Then there exists i, $1 \leq i \leq K$, such that

$x_i = 0$ and $y_i = 1$ or $x_i = 1$ and $y_i = 0$.



Take $z = 0^{i-1}$. This moves column i to be Kth from end in both xz and yz, so $L(xz) \neq L(yz)$.

$\therefore S$ is a PD set of size 2^K for A_K, so the design "Remember last K bits read and accept iff first of those is a 1" cannot be improved upon.

EXTRA: Notice the size:

Regexp $(0+1)^*1(0+1)^{\wedge}(K-1)$:

$15 + \sim \log_2 K$ characters.

NFA:

$K+1$ states, $2K+1$ arcs. $O(K)$.

But DFA has 2^K states, humongous!

Extra To answer the question in class about what happens if you mirror-image the relation used in the Myhill-Nerode Theorem, writing $X \sim_L^R Y$ iff $(\forall z \in \Sigma^*)[zX \in L \Leftrightarrow zY \in L]$.

Well, let's mirror-image the language too: $L^R = \{x^R : x \in L\}$ where x^R means x reversed. Now:

$zX \in L \Leftrightarrow (zX)^R \in L^R \Leftrightarrow X^R \cdot z^R \in L^R$ } Thus $X \sim_L^R Y \Leftrightarrow (\forall z^R \in \Sigma^*)[X^R z^R \in L^R \Leftrightarrow Y^R z^R \in L^R] \Leftrightarrow X^R \sim_{L^R} Y^R$.

$\Updownarrow$

$zY \in L \Leftrightarrow (zY)^R \in L^R \Leftrightarrow Y^R \cdot z^R \in L^R$ }

Thus the reversed relation for L is equivalent to the original relation $\sim_{L^R}$ for L^R using reversed strings.

Now, L is regular $\Leftrightarrow$ L^R is regular: if you "well-ground" a DFA s.t. $L(M) = L$ by adding ε-arcs from all old final states to one new final state, make that the start state, reverse all arc arrows, and make the old start state the new final state, you get an NFA N such that $L(N) = L^R$. Though the index of $\sim_{L^R}$ vs. $\sim_L$ remains finite, it can go up considerably, as L_K versus L_K^R above shows.