Myhill - Nerode Theorem

(John Myhill and Anil Nerode, 1958)

Not in the book. Only stated as exercise 1.51 and 1.52.

This provides a necessary and sufficient condition for a language to be regular.

Def 1: Let x, y be strings over Σ and L be a language over Σ. We say that x, y are distinguishable by L if $\exists z \in \Sigma^*$ such that $xz \in L$ and $yz \notin L$ or vice versa. ($xz \notin L$ and $yz \in L$)

If x, y are not distinguishable by L, we say that they are indistinguishable by L, and denote it by $x \equiv_L y$.

Exercise 1: Show that $\equiv_L$ is an equivalence relation.

(1) Reflexive : $x \equiv_L x$

(2) Symmetric : $x \equiv_L y \Rightarrow y \equiv_L x$

(3) Transitive : $x \equiv_L y$, and $y \equiv_L z \Rightarrow x \equiv_L z$

This implies that the relation $\equiv_L$ partitions Σ^* into equivalence classes.

Example: mod 5 equivalence relation partitions all integers into 5 equivalence classes – based on the remainder when divided by 5.

Similarly '$\equiv_L$' relation partitions Σ^* into equi. classes.

Def 2: Let L be a language and X be a set of strings. X is pairwise distinguishable by L if every two distinct strings $x, y \in X$ are distinguishable by L.

Def 3: The index of L is the size of the largest set X of strings such that X is pairwise distinguishable by L.

In other words, index of L = No. of equivalence classes of Σ^* as determined by $\equiv_L$.

Myhill-Nerode Theorem: A language L is regular iff (if and only if) it has a finite index.

Moreover, the index of L is equal to the size (no. of states) of a smallest DFA that recognizes L.

Lemma 1: If L is recognized by a DFA with k states, then $\text{index}(L) \le k$.

Lemma 2: If $\text{index}(L) = k < \infty$, then there exists a DFA with k states that recognizes L.

Proof of Myhill Nerode theorem assuming lemmas

($\Longrightarrow$) Suppose L is regular. Then there is a DFA that recognizes L. Consider a smallest DFA that recognizes L. Let this DFA be M, and let M have k states. By lemma 1, we have $\text{index}(L) \le k$.

Index(L) $\le$ Size of the smallest DFA

that recognizes L.

($\Leftarrow$) Suppose L has finite index, say k. By lemma 2, there exists a DFA with k states that recognizes L. So L is regular.

$$\boxed{\left.\begin{array}{l}\text{Size of the smallest DFA} \\ \text{that recognizes } L\end{array}\right\} \leq \text{Index}(L)}$$

Combining, we get

$$\text{Index}(L) = \text{Size of the smallest DFA that recognizes } L.$$

$\delta(q, a) = r$ $\qquad q \xrightarrow{a} r \xrightarrow{b} s$ $\qquad \delta^*(q, ab) = s$

<u>Notation</u>: $\delta^*(q, x)$ for $x \in \Sigma^*$ denotes the state reached by the DFA starting from q and reading the string x.

<u>Lemma 1</u>: If L is recognized by a DFA with k states, then $\text{index}(L) \leq k$.

Proof: We will show that any two strings that end in the same state are indistinguishable.

Suppose L is recognized by a DFA M with k states.
Suppose, for the sake of contradiction that $\text{index}(L) > k$.

This means there exists X such that X is pairwise distinguishable by L, and $|X| > k$.

Let q_0 be the starting state of M. By pigeonhole principle, there exists two strings $x, y \in X$, $x \neq y$ such that $\underbrace{\delta^*(q_0, x)}_{r} = \delta^*(q_0, y)$ (x and y end in the same state)

For any $z \in \Sigma^*$, we now have

$$\delta^*(q_0, xz) = \delta^*(\delta^*(q_0, x), z)$$
$$= \delta^*(r, z)$$
$$= \delta^*(\delta^*(q_0, y), z)$$
$$= \delta^*(q_0, yz)$$

So $xz \in L \iff yz \in L$. So x, y are indistinguishable by L. . So $x \equiv_L y$. This is a contradiction.
So $\text{index}(L) \leq k$.

lemma 2: If index $(L) = k < \infty$, then there exists a DFA with k states that recognizes L.

Proof: Suppose index $(L) = k < \infty$. We will construct a DFA M with k states that recognizes L. Let $X = \{x_1, x_2, \dots x_k\} \subseteq \Sigma^*$ be a set of strings pairwise distinguishable by L.

$M = (Q, \Sigma, \delta, q_0, F)$. $Q = \{q_1, q_2, \dots q_k\}$

Each state $q_i \in Q$ corresponds to $x_i \in X$.

For each $a \in \Sigma$, $\delta(q_i, a)$ is defined as follows.

We have $x_i a \equiv_L x_j$ for some $x_j \in X$. Else, we can add $\{x_i a\} \cup X$ to get a bigger pairwise distinguishable set. Now set $\delta(q_i, a) = q_j$.

Similarly, $\varepsilon \equiv_L x_m$ for some $x_m \in X$. Set $q_0 = q_m$.

Finally, define $F = \{q_i \mid x_i \in L\}$. Now we need to show that M recognizes L.

$$q_i \xrightarrow{w} q_j$$

$x_i w \equiv_L x_j$

We first state a claim.

$\boxed{\text{Claim: } \delta^*(q_i, w) = q_j \iff x_i w \equiv_L x_j \text{ for all } i, j \text{ and } w \in \Sigma^*.}$

We will prove the lemma using the claim.

Suppose $x \in L$. Then $x \equiv_L x_i$ for some $x_i \in X \cap L$.

$$x = \varepsilon x \equiv_L x_i \iff \delta^*(q_0, x) = q_i \in F.$$

Therefore x is accepted by M.

Suppose $x \notin L$. Then $x \equiv_L x_j$ for some $x_j \in X$, and $x_j \notin L$. Similar to the above, we get that

$\delta^*(q_0, x) = q_j$, where $q_j \notin F$. So M does not accept x.

Thus M recognizes the language L.

$$q_i \xrightarrow{\ w\ } q_j$$

$$x_i w \equiv_L x_j$$

<u>Claim</u>: $\delta^*(q_i, w) = q_j \iff x_i w \equiv_L x_j$ for all i, j and $w \in \Sigma^*$.

<u>Proof</u>: By induction on $|w|$.

When $|w| = 0$, $w = \varepsilon$.

$\delta^*(q_i, \varepsilon) = q_i$ and $x_i \varepsilon = x_i \equiv_L x_i$

When $|w| = 1$, $w = a \in \Sigma$.

Let $\delta^*(q_i, a) = \delta(q_i, a) = q_j$

By definition of δ, we have $x_i a \equiv_L x_j$.

So claim is true.

Induction step: let $|w| = l \geq 1$. let $w = va$ where
$|w| = |v| + 1$ and $a \in \Sigma$.

$$\delta^*(q_i, w) = \delta(\delta^*(q_i, v), a) = \delta(q_{j_1}, a) = q_{j_2}$$

where $q_{j_1} = \delta^*(q_i, v)$.

By induction, we have $x_{j_1} \equiv_L x_i v$, and

by definition of δ, $x_{j_2} \equiv_L x_{j_1} a$

$$x_{j_2} \equiv_L x_{j_1} a \equiv_L x_i va = x_i w.$$

So claim holds for w, $|w| \geq 1$, as well.

Example: $A = \{0^n 1^n \mid n \geq 0\} = \{\varepsilon, 01, 0011, 000111, \ldots\}$

Consider $x_i = 0^i$ for $i = 0, 1, 2, 3, \ldots$

The set $X = \{x_i \mid i \geq 0\}$ is pairwise distinguishable by A.

$X = \{0^i \mid i \geq 0\}$
$= \{\varepsilon, 0, 00, 000, \ldots\}$

Given x_i, x_j such that $i \neq j$.

Consider the string 1^i. We have $x_i 1^i \in A$, but $x_j 1^i \notin A$, when $i \neq j$. So 1^i distinguishes x_i and x_j. So X is an infinite set, pairwise distinguishable by A.

Hence A is not regular.

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

$X = \{0, 00, 000, 0000\}$

01 is in A and $001 \notin A$. So 0 and 00 are distinguishable by A.

$0011 \in A$ but $00011 \notin A$. So 00 and 000 are distinguishable by A.

We can verify that above X is pairwise distinguishable by A.