5.3 The Myhill-Nerode theorem and minimization of DFAs; decision properties of regular languages

(Here I follow Section 3.4 from the first edition[1] of HMU. Because this book will be inaccessible to most, this script is more detailed here than elsewhere.)

Recall: a binary relation R on a set X is an *equivalence relation* iff (1) R is reflexive, that is xRx for all $x \in X$; (2) R is symmetric, that is, $xRy \Rightarrow yRx$ for all $x, y \in X$; (3) R is transitive, that is, xRy and $yRz \Rightarrow xRz$ for all $x, y, z \in X$. An equivalence relation R on X partitions R into subsets X_i (where $i \in I$) called the equivalence classes of R, defined by (1) for all $i \in I$, for all $x, y \in X_i$, it holds that xRy, (2) for all $i,j \in I$, $i \neq j$, for all $x \in X_i$, $y \in X_j$, it holds that not xRy. The number $| I |$ of equivalence classes is the *index* of R. If $C \subseteq X$ is an equivalence class containing x, we call x a *representative* of C and write $C = [x]_R$.

Definition 5.2. An equivalence relation R on Σ^* is *right invariant* if for all $u, v, w \in \Sigma^*$, it holds that $uRv \Rightarrow uwRvw$.

Definition 5.3. Let A be a DFA with starting state q_0. Define a binary relation R_A on Σ^* by uR_Av iff $\delta(q_0, u) = \delta(q_0, v)$. This is an equivalence relation because "=" is one and R_A is defined via "=" of states.

Obviously R_A is right-invariant, has finite index, and $L(A)$ is the union of some of the equivalence classes of R_A (namely those for which $\delta(q_0, v)$ leads into an accepting state). This insight is considerably refined in the following theorem.

Theorem 5.4 (The Myhill-Nerode theorem; algebraic characterization of regular languages). Let $L \subseteq \Sigma^*$ be a language. The following three statements are equivalent:

1. $L \subseteq \Sigma^*$ is regular, that is, $L = L(A)$ for some DFA A.
2. L is the union of some of the equivalence classes of a right-invariant equivalence relation R on Σ^* of finite index.
3. Let an equivalence relation R_L on Σ^* be defined by:
   $$uR_Lv \quad \text{iff} \quad \text{for all } w \in \Sigma^*: uw \in L \Leftrightarrow vw \in L.$$
   Then R_L is of finite index.

The Myhill-Nerode theorem gives yet another characterization of the regular languages. Its proof involves the construction of minimal automata, which is an important technique in its own right.

Proof: 1. $\Rightarrow$ 2: We saw that already.
2. $\Rightarrow$ 3: We show that any equivalence relation R satisfying 2. is a refinement of R_L, that is, every equivalence class of R is entirely contained in some equivalence class of R_L. Thus the index of R_L cannot be greater than the index of R and therefore is finite. Assume uRv. Since R

[1] John E. Hopcroft, Jeffrey D. Ullman: *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley 1979

is right invariant, for every w in Σ^* we have $uwRvw$, and therefore uR_Lv, so R is a refinement of uR_Lv.
3. $\Rightarrow$ 1: We first show that R_L is right invariant. Suppose that uR_Lv. For any $y \in \Sigma^*$ we must show that uyR_Lvy, that is, for any $z \in \Sigma^*$, $uyz \in L \Leftrightarrow vyz \in L$. But this follows from the definition of R_L in 3., if we put $w = yz$.
We now construct a DFA A_{MN} accepting L. The basic trick is to use as states the equivalence classes of R_L (of which there is only a finite number!). So $A_{MN} = (Q, \Sigma, q_0, \delta, F) = (\{[w]_{R_L} \mid w \in \Sigma^*\}, \Sigma, q_0, \delta_{MN}, F_{MN})$. We put

i) $q_0 = [\varepsilon]_{R_L}$,
ii) for all $w \in \Sigma^*$, $a \in \Sigma$: $\delta_{MN}([w]_{R_L}, a) = [wa]_{R_L}$,
iii) $F_{MN} = \{[w]_{R_L} \mid w \in L\}$.

We first have to show that this is well-defined, especially that ii) makes sense, that is, the definition of δ_{MN} is independent of the choice of representative w of the equivalence class $[w]_{R_L}$. So we have to show that if wR_Lv, then $[wa]_{R_L} = \delta_{MN}([w]_{R_L}, a) = \delta_{MN}([v]_{R_L}, a) = [va]_{R_L}$. But if wR_Lv, then waR_Lva because R_L is right-invariant, therefore $[wa]_{R_L} = [va]_{R_L}$. So δ_{MN} is well-defined.

Finally we have to show that $L = L(A_{MN})$. This follows from $\hat{\delta}_{MN}([\varepsilon]_{R_L}, w) = [w]_{R_L}$ (an easy induction over w) and iii). □

Here is an example that illustrates the Myhill-Nerode theorem. Let L be the language 0*10*, that is, all words over {0,1} that contain exactly one 1. L is accepted, for instance, by the DFA A shown in Figure 5.1.

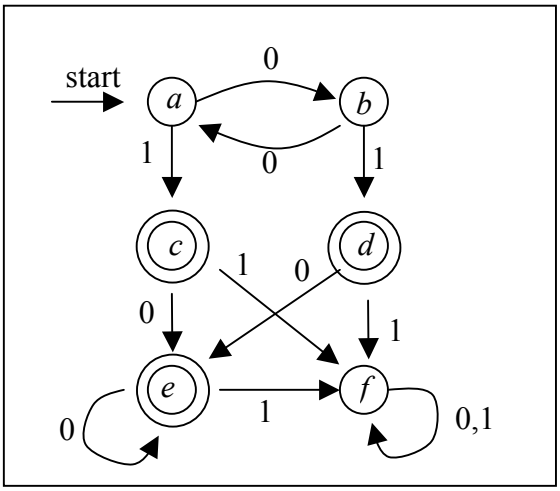

Figure 5.1. DFA A accepting $L = 0^*10^*$.

This DFA A yields a right-invariant equivalence relation R_A as follows. R_A has six equivalent classes corresponding to the states of A. These equivalence classes each contain the words that lead from the starting state into that class:

$C_a = (00)^*$ $C_b = (00)^*0$ $C_c = (00)^*1$

$C_d = (00)^*01$ $C_e = 0^*100^*$ $C_f = 0^*10^*1(0+1)^*$

L is the union of C_c, C_d and C_e.

The relation R_L for L has wR_Lv if and only if either

a) w and v have no 1's [= equivalence class D_1],
b) w and v each have one 1 [= equivalence class D_2],
c) w and v each have more than one 1 [= equivalence class D_3].

The classes C_i refine the classes D_j by $D_1 = C_a \cup C_b$, $D_2 = C_c \cup C_d \cup C_e$, and $D_3 = C_f$.

We may denote the three equivalence classes of R_L by $D_1 = [\varepsilon]_{R_L} = 0^*$, $d_2 = [1]_{R_L} = 0^*10^*$, $D_3 = [11]_{R_L} = 0^*10^*1(0+1)^*$. The DFA A_{MN} we get according to the construction in the proof of the Myhill-Nerode theorem is the following:

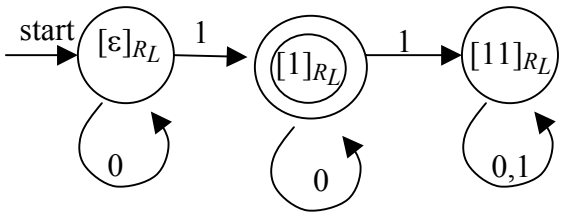

Figure 5.2: The DFA A_{MN}.