# Lecture 15

# Myhill–Nerode Relations

Two deterministic finite automata

$$\begin{aligned} M &= (Q_M,\, \Sigma,\, \delta_M,\, s_M,\, F_M), \\ N &= (Q_N,\, \Sigma,\, \delta_N,\, s_N,\, F_N) \end{aligned}$$

are said to be *isomorphic* (Greek for "same form") if there is a one-to-one and onto mapping $f : Q_M \to Q_N$ such that

- $f(s_M) = s_N$,
- $f(\delta_M(p, a)) = \delta_N(f(p), a)$ for all $p \in Q_M$, $a \in \Sigma$, and
- $p \in F_M$ iff $f(p) \in F_N$.

That is, they are essentially the same automaton up to renaming of states. It is easily argued that isomorphic automata accept the same set.

In this lecture and the next we will show that if $M$ and $N$ are any two automata with no inaccessible states accepting the same set, then the quotient automata $M/\approx$ and $N/\approx$ obtained by the collapsing algorithm of Lecture 14 are isomorphic. Thus the DFA obtained by the collapsing algorithm is the minimal DFA for the set it accepts, and this automaton is unique up to isomorphism.

We will do this by exploiting a profound and beautiful correspondence between finite automata with input alphabet $\Sigma$ and certain equivalence relations on $\Sigma^*$. We will show that the unique minimal DFA for a regular set $R$ can be defined in a natural way *directly from* $R$, and that any minimal automaton for $R$ is isomorphic to this automaton.

## Myhill–Nerode Relations

Let $R \subseteq \Sigma^*$ be a regular set, and let $M = (Q,\ \Sigma,\ \delta,\ s,\ F)$ be a DFA for $R$ with no inaccessible states. The automaton $M$ induces an equivalence relation $\equiv_M$ on $\Sigma^*$ defined by

$$x \equiv_M y \quad \stackrel{\text{def}}{\iff} \quad \widehat{\delta}(s,x) = \widehat{\delta}(s,y).$$

(Don't confuse this relation with the collapsing relation $\approx$ of Lecture 13—that relation was defined on $Q$, whereas $\equiv_M$ is defined on $\Sigma^*$.)

One can easily show that the relation $\equiv_M$ is an equivalence relation; that is, that it is reflexive, symmetric, and transitive. In addition, $\equiv_M$ satisfies a few other useful properties:

(i) It is a *right congruence*: for any $x, y \in \Sigma^*$ and $a \in \Sigma$,

$$x \equiv_M y \quad \Rightarrow \quad xa \equiv_M ya.$$

To see this, assume that $x \equiv_M y$. Then

$$\begin{aligned} \widehat{\delta}(s,xa) &= \delta(\widehat{\delta}(s,x),a) \\ &= \delta(\widehat{\delta}(s,y),a) \quad \text{by assumption} \\ &= \widehat{\delta}(s,ya). \end{aligned}$$

(ii) It *refines* $R$: for any $x, y \in \Sigma^*$,

$$x \equiv_M y \quad \Rightarrow \quad (x \in R \iff y \in R).$$

This is because $\widehat{\delta}(s,x) = \widehat{\delta}(s,y)$, and this is either an accept or a reject state, so either both $x$ and $y$ are accepted or both are rejected. Another way to say this is that every $\equiv_M$-class has either all its elements in $R$ or none of its elements in $R$; in other words, $R$ is a union of $\equiv_M$-classes.

(iii) It is of *finite index*; that is, it has only finitely many equivalence classes. This is because there is exactly one equivalence class

$$\{x \in \Sigma^* \mid \widehat{\delta}(s,x) = q\}$$

corresponding to each state $q$ of $M$.

Let us call an equivalence relation $\equiv$ on $\Sigma^*$ a *Myhill–Nerode relation for R* if it satisfies properties (i), (ii), and (iii); that is, if it is a right congruence of finite index refining $R$.

The interesting thing about this definition is that it characterizes exactly the relations on $\Sigma^*$ that are $\equiv_M$ for some automaton $M$. In other words, we can reconstruct $M$ from $\equiv_M$ using only the fact that $\equiv_M$ is Myhill–Nerode. To see this, we will show how to construct an automaton $M_\equiv$ for $R$ from any given Myhill–Nerode relation $\equiv$ for $R$. We will show later that the two constructions

$$\begin{aligned} M &\mapsto \equiv_M, \\ \equiv &\mapsto M_\equiv \end{aligned}$$

are inverses up to isomorphism of automata.

Let $R \subseteq \Sigma^*$, and let $\equiv$ be an arbitrary Myhill–Nerode relation for $R$. Right now we're not assuming that $R$ is regular, only that the relation $\equiv$ satisfies (i), (ii), and (iii). The $\equiv$-class of the string $x$ is

$$[x] \stackrel{\text{def}}{=} \{y \mid y \equiv x\}.$$

Although there are infinitely many strings, there are only finitely many $\equiv$-classes, by property (iii).

Now define the DFA $M_\equiv = (Q,\ \Sigma,\ \delta,\ s,\ F)$, where

$$\begin{aligned} Q &\stackrel{\text{def}}{=} \{[x] \mid x \in \Sigma^*\}, \\ s &\stackrel{\text{def}}{=} [\epsilon], \\ F &\stackrel{\text{def}}{=} \{[x] \mid x \in R\}, \\ \delta([x], a) &\stackrel{\text{def}}{=} [xa]. \end{aligned}$$

It follows from property (i) of Myhill–Nerode relations that $\delta$ is well defined. In other words, we have defined the action of $\delta$ on an equivalence class $[x]$ in terms of an element $x$ chosen from that class, and it is conceivable that we could have gotten something different had we chosen another $y \in [x]$ such that $[xa] \neq [ya]$. The property of right congruence says exactly that this cannot happen.

Finally, observe that

$$x \in R \iff [x] \in F. \tag{15.1}$$

The implication ($\Rightarrow$) is from the definition of $F$, and ($\Leftarrow$) follows from the definition of $F$ and property (ii) of Myhill–Nerode relations.

Now we are ready to prove that $L(M_\equiv) = R$.

**Lemma 15.1** $\widehat{\delta}([x], y) = [xy]$.

*Proof.* Induction on $|y|$.

Basis

$$\widehat{\delta}([x],\epsilon) \quad = \quad [x] = [x\epsilon].$$

Induction step

$$\begin{aligned} \widehat{\delta}([x],ya) &= \delta(\widehat{\delta}([x],y),a) && \text{definition of } \widehat{\delta} \\ &= \delta([xy],a) && \text{induction hypothesis} \\ &= [xya] && \text{definition of } \delta. \end{aligned}$$

□

**Theorem 15.2** $L(M_\equiv) = R$.

*Proof.*

$$\begin{aligned} x \in L(M_\equiv) &\iff \widehat{\delta}([\epsilon],x) \in F && \text{definition of acceptance} \\ &\iff [x] \in F && \text{Lemma 15.1} \\ &\iff x \in R && \text{property (15.1).} \end{aligned}$$

□

## $M \mapsto \equiv_M$ and $\equiv \mapsto M_\equiv$ Are Inverses

We have described two natural constructions, one taking a given automaton $M$ for $R$ with no inaccessible states to a corresponding Myhill–Nerode relation $\equiv_M$ for $R$, and one taking a given Myhill–Nerode relation $\equiv$ for $R$ to a DFA $M_\equiv$ for $R$. We now wish to show that these two operations are inverses up to isomorphism.

**Lemma 15.3** *(i) If $\equiv$ is a Myhill–Nerode relation for $R$, and if we apply the construction $\equiv \mapsto M_\equiv$ and then apply the construction $M \mapsto \equiv_M$ to the result, the resulting relation $\equiv_{M_\equiv}$ is identical to $\equiv$.*

*(ii) If $M$ is a DFA for $R$ with no inaccessible states, and if we apply the construction $M \mapsto \equiv_M$ and then apply the construction $\equiv \mapsto M_\equiv$ to the result, the resulting DFA $M_{\equiv_M}$ is isomorphic to $M$.*

*Proof.* (i) Let $M_\equiv = (Q,\ \Sigma,\ \delta,\ s,\ F)$ be the automaton constructed from $\equiv$ as described above. Then for any $x, y \in \Sigma^*$,

$$\begin{aligned} x \equiv_{M_\equiv} y &\iff \widehat{\delta}(s,x) = \widehat{\delta}(s,y) && \text{definition of } \equiv_{M_\equiv} \\ &\iff \widehat{\delta}([\epsilon],x) = \widehat{\delta}([\epsilon],y) && \text{definition of } s \\ &\iff [x] = [y] && \text{Lemma 15.1} \\ &\iff x \equiv y. \end{aligned}$$

(ii) Let $M = (Q, \Sigma, \delta, s, F)$ and let $M_{\equiv_M} = (Q', \Sigma, \delta', s', F')$. Recall from the construction that

$$\begin{aligned}
[x] &= \{y \mid y \equiv_M x\} = \{y \mid \widehat{\delta}(s,y) = \widehat{\delta}(s,x)\}, \\
Q' &= \{[x] \mid x \in \Sigma^*\}, \\
s' &= [\epsilon], \\
F' &= \{[x] \mid x \in R\}, \\
\delta'([x],a) &= [xa].
\end{aligned}$$

We will show that $M_{\equiv_M}$ and $M$ are isomorphic under the map

$$\begin{aligned}
f : Q' &\rightarrow Q, \\
f([x]) &= \widehat{\delta}(s,x).
\end{aligned}$$

By the definition of $\equiv_M$, $[x] = [y]$ iff $\widehat{\delta}(s,x) = \widehat{\delta}(s,y)$, so the map $f$ is well defined on $\equiv_M$-classes and is one-to-one. Since $M$ has no inaccessible states, $f$ is onto.

To show that $f$ is an isomorphism of automata, we need to show that $f$ preserves all automata-theoretic structure: the start state, transition function, and final states. That is, we need to show

- $f(s') = s$,
- $f(\delta'([x],a)) = \delta(f([x]),a)$,
- $[x] \in F' \iff f([x]) \in F$.

These are argued as follows:

$$\begin{aligned}
f(s') &= f([\epsilon]) && \text{definition of } s' \\
&= \widehat{\delta}(s,\epsilon) && \text{definition of } f \\
&= s && \text{definition of } \widehat{\delta};
\end{aligned}$$

$$\begin{aligned}
f(\delta'([x],a)) &= f([xa]) && \text{definition of } \delta' \\
&= \widehat{\delta}(s,xa) && \text{definition of } f \\
&= \delta(\widehat{\delta}(s,x),a) && \text{definition of } \widehat{\delta} \\
&= \delta(f([x]),a) && \text{definition of } f;
\end{aligned}$$

$$\begin{aligned}
[x] \in F' &\iff x \in R && \text{definition of } F \text{ and property (ii)} \\
&\iff \widehat{\delta}(s,x) \in F && \text{since } L(M) = R \\
&\iff f([x]) \in F && \text{definition of } f.
\end{aligned}$$

□

We have shown:

**Theorem 15.4** *Let $\Sigma$ be a finite alphabet. Up to isomorphism of automata, there is a one-to-one correspondence between deterministic finite automata over $\Sigma$ with no inaccessible states accepting $R$ and Myhill–Nerode relations for $R$ on $\Sigma^*$.*

# Lecture 16

# The Myhill–Nerode Theorem

Let $R \subseteq \Sigma^*$ be a regular set. Recall from Lecture 15 that a *Myhill–Nerode relation for R* is an equivalence relation $\equiv$ on $\Sigma^*$ satisfying the following three properties:

(i) $\equiv$ is a *right congruence*: for any $x, y \in \Sigma^*$ and $a \in \Sigma$,

$$x \equiv y \quad \Rightarrow \quad xa \equiv ya;$$

(ii) $\equiv$ *refines* $R$: for any $x, y \in \Sigma^*$,

$$x \equiv y \quad \Rightarrow \quad (x \in R \Longleftrightarrow y \in R);$$

(iii) $\equiv$ is of *finite index*; that is, $\equiv$ has only finitely many equivalence classes.

We showed that there was a natural one-to-one correspondence (up to isomorphism of automata) between

- deterministic finite automata for $R$ with input alphabet $\Sigma$ and with no inaccessible states, and
- Myhill–Nerode relations for $R$ on $\Sigma^*$.

This is interesting, because it says we can deal with regular sets and finite automata in terms of a few simple, purely algebraic properties.

In this lecture we will show that there exists a *coarsest* Myhill–Nerode relation $\equiv_R$ for any given regular set $R$; that is, one that every other Myhill–Nerode relation for $R$ refines. The notions of *coarsest* and *refinement* will be defined below. The relation $\equiv_R$ corresponds to the unique minimal DFA for $R$.

Recall from Lecture 15 the two constructions

- $M \mapsto \equiv_M$, which takes an arbitrary DFA $M = (Q, \Sigma, \delta, s, F)$ with no inaccessible states accepting $R$ and produces a Myhill–Nerode relation $\equiv_M$ for $R$:

$$x \equiv_M y \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad \widehat{\delta}(s,x) = \widehat{\delta}(s,y);$$

- $\equiv \mapsto M_\equiv$, which takes an arbitrary Myhill–Nerode relation $\equiv$ on $\Sigma^*$ for $R$ and produces a DFA $M_\equiv = (Q, \Sigma, \delta, s, F)$ accepting $R$:

$$\begin{aligned}
[x] &\stackrel{\text{def}}{=} \{y \mid y \equiv x\}, \\
Q &\stackrel{\text{def}}{=} \{[x] \mid x \in \Sigma^*\}, \\
s &\stackrel{\text{def}}{=} [\epsilon], \\
\delta([x], a) &\stackrel{\text{def}}{=} [xa], \\
F &\stackrel{\text{def}}{=} \{[x] \mid x \in R\}.
\end{aligned}$$

We showed that these two constructions are inverses up to isomorphism.

**Definition 16.1** A relation $\equiv_1$ is said to *refine* another relation $\equiv_2$ if $\equiv_1 \subseteq \equiv_2$, considered as sets of ordered pairs. In other words, $\equiv_1$ *refines* $\equiv_2$ if for all $x$ and $y$, $x \equiv_1 y$ implies $x \equiv_2 y$. For equivalence relations $\equiv_1$ and $\equiv_2$, this is the same as saying that for every $x$, the $\equiv_1$-class of $x$ is included in the $\equiv_2$-class of $x$. □

For example, the equivalence relation $x \equiv y \bmod 6$ on the integers refines the equivalence relation $x \equiv y \bmod 3$. For another example, clause (ii) of the definition of Myhill–Nerode relations says that a Myhill–Nerode relation $\equiv$ for $R$ refines the equivalence relation with equivalence classes $R$ and $\Sigma^* - R$.

The relation of *refinement* between equivalence relations is a partial order: it is reflexive (every relation refines itself), transitive (if $\equiv_1$ refines $\equiv_2$ and $\equiv_2$ refines $\equiv_3$, then $\equiv_1$ refines $\equiv_3$), and antisymmetric (if $\equiv_1$ refines $\equiv_2$ and $\equiv_2$ refines $\equiv_1$, then $\equiv_1$ and $\equiv_2$ are the same relation).

If $\equiv_1$ refines $\equiv_2$, then $\equiv_1$ is the *finer* and $\equiv_2$ is the *coarser* of the two relations. There is always a finest and a coarsest equivalence relation on any set $U$, namely the *identity relation* $\{(x,x) \mid x \in U\}$ and the *universal relation* $\{(x,y) \mid x,y \in U\}$, respectively.

Now let $R \subseteq \Sigma^*$, regular or not. We define an equivalence relation $\equiv_R$ on $\Sigma^*$ in terms of $R$ as follows:

$$x \equiv_R y \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad \forall z \in \Sigma^* \ (xz \in R \Longleftrightarrow yz \in R). \tag{16.1}$$

In other words, two strings are equivalent under $\equiv_R$ if, whenever you append the same string to both of them, the resulting two strings are either both in $R$ or both not in $R$. It is not hard to show that this is an equivalence relation for any $R$.

We show that for any set $R$, regular or not, the relation $\equiv_R$ satisfies the first two properties (i) and (ii) of Myhill–Nerode relations and is the coarsest such relation on $\Sigma^*$. In case $R$ is regular, this relation is also of finite index, therefore a Myhill–Nerode relation for $R$. In fact, it is the coarsest possible Myhill–Nerode relation for $R$ and corresponds to the unique minimal finite automaton for $R$.

**Lemma 16.2** *Let $R \subseteq \Sigma^*$, regular or not. The relation $\equiv_R$ defined by (16.1) is a right congruence refining $R$ and is the coarsest such relation on $\Sigma^*$.*

*Proof.* To show that $\equiv_R$ is a right congruence, take $z = aw$ in the definition of $\equiv_R$:

$$\begin{aligned} x \equiv_R y \quad &\Rightarrow \quad \forall a \in \Sigma \ \forall w \in \Sigma^* (xaw \in R \Longleftrightarrow yaw \in R) \\ &\Rightarrow \quad \forall a \in \Sigma \ (xa \equiv_R ya). \end{aligned}$$

To show that $\equiv_R$ refines $R$, take $z = \epsilon$ in the definition of $\equiv_R$:

$$x \equiv_R y \quad \Rightarrow \quad (x \in R \Longleftrightarrow y \in R).$$

Moreover, $\equiv_R$ is the coarsest such relation, because any other equivalence relation $\equiv$ satisfying (i) and (ii) refines $\equiv_R$:

$$\begin{aligned} & x \equiv y \\ & \quad \Rightarrow \quad \forall z \ (xz \equiv yz) && \text{by induction on } |z|, \text{ using property (i)} \\ & \quad \Rightarrow \quad \forall z \ (xz \in R \Longleftrightarrow yz \in R) && \text{property (ii)} \\ & \quad \Rightarrow \quad x \equiv_R y && \text{definition of } \equiv_R. \end{aligned}$$

□

At this point all the hard work is done. We can now state and prove the *Myhill–Nerode theorem*:

**Theorem 16.3** **(Myhill–Nerode theorem)** *Let $R \subseteq \Sigma^*$. The following statements are equivalent:*

*(a) $R$ is regular;*

*(b) there exists a Myhill–Nerode relation for $R$;*

(c) *the relation $\equiv_R$ is of finite index.*

*Proof.* (a) $\Rightarrow$ (b) Given a DFA $M$ for $R$, the construction $M \mapsto \equiv_M$ produces a Myhill–Nerode relation for $R$.

(b) $\Rightarrow$ (c) By Lemma 16.2, any Myhill–Nerode relation for $R$ is of finite index and refines $\equiv_R$; therefore $\equiv_R$ is of finite index.

(c) $\Rightarrow$ (a) If $\equiv_R$ is of finite index, then it is a Myhill–Nerode relation for $R$, and the construction $\equiv \mapsto M_\equiv$ produces a DFA for $R$. □

Since $\equiv_R$ is the unique coarsest Myhill–Nerode relation for a regular set $R$, it corresponds to the DFA for $R$ with the fewest states among all DFAs for $R$.

The collapsing algorithm of Lecture 14 actually gives this automaton. Suppose $M = (Q, \Sigma, \delta, s, F)$ is a DFA for $R$ that is already collapsed; that is, there are no inaccessible states, and the collapsing relation

$$p \approx q \quad \overset{\text{def}}{\iff} \quad \forall x \in \Sigma^* \; (\widehat{\delta}(p, x) \in F \iff \widehat{\delta}(q, x) \in F)$$

is the identity relation on $Q$. Then the Myhill–Nerode relation $\equiv_M$ corresponding to $M$ is exactly $\equiv_R$:

$$\begin{array}{lll} x \equiv_R y & & \\ \iff & \forall z \in \Sigma^* \; (xz \in R \iff yz \in R) & \text{definition of } \equiv_R \\ \iff & \forall z \in \Sigma^* \; (\widehat{\delta}(s, xz) \in F \iff \widehat{\delta}(s, yz) \in F) & \text{definition of acceptance} \\ \iff & \forall z \in \Sigma^* \; (\widehat{\delta}(\widehat{\delta}(s, x), z) \in F \iff \widehat{\delta}(\widehat{\delta}(s, y), z) \in F) & \\ & & \text{Homework 1, Exercise 3} \\ \iff & \widehat{\delta}(s, x) \approx \widehat{\delta}(s, y) & \text{definition of } \approx \\ \iff & \widehat{\delta}(s, x) = \widehat{\delta}(s, y) & \text{since } M \text{ is collapsed} \\ \iff & x \equiv_M y & \text{definition of } \equiv_M. \end{array}$$

## An Application

The Myhill–Nerode theorem can be used to determine whether a set $R$ is regular or nonregular by determining the number of $\equiv_R$-classes. For example, consider the set

$$A \quad = \quad \{a^n b^n \mid n \geq 0\}.$$

If $k \neq m$, then $a^k \not\equiv_A a^m$, since $a^k b^k \in A$ but $a^m b^k \notin A$. Therefore, there are infinitely many $\equiv_A$-classes, at least one for each $a^k$, $k \geq 0$. By the Myhill–Nerode theorem, $A$ is not regular.

In fact, one can show that the $\equiv_A$-classes are exactly

$$\begin{aligned} G_k &= \{a^k\}, \quad k \geq 0, \\ H_k &= \{a^{n+k}b^n \mid 1 \leq n\}, \quad k \geq 0, \\ E &= \Sigma^* - \bigcup_{k \geq 0} G_k \cup H_k = \Sigma^* - \{a^m b^n \mid 0 \leq n \leq m\}. \end{aligned}$$

For strings in $G_k$, all and only strings in $\{a^n b^{n+k} \mid n \geq 0\}$ can be appended to obtain a string in $A$; for strings in $H_k$, only the string $b^k$ can be appended to obtain a string in $A$; and no string can be appended to a string in $E$ to obtain a string in $A$.

We will see another application of the Myhill–Nerode theorem involving two-way finite automata in Lectures 17 and 18.

## Historical Notes

Minimization of DFAs was studied by Huffman [54], Moore [84], Nerode [88], and Hopcroft [53], among others. The Myhill–Nerode theorem is due independently to Myhill [85] and Nerode [88] in slightly different forms.