Lecture 10

Homomorphisms

A *homomorphism* is a map $h : \Sigma^* \to \Gamma^*$ such that for all $x, y \in \Sigma^*$,

$$h(xy) = h(x)h(y), \tag{10.1}$$

$$h(\epsilon) = \epsilon. \tag{10.2}$$

Actually, (10.2) is a consequence of (10.1):

$$\begin{aligned} |h(\epsilon)| &= |h(\epsilon\epsilon)| \\ &= |h(\epsilon)h(\epsilon)| \\ &= |h(\epsilon)| + |h(\epsilon)|; \end{aligned}$$

subtracting $|h(\epsilon)|$ from both sides, we have $|h(\epsilon)| = 0$, therefore $h(\epsilon) = \epsilon$.

It follows from these properties that any homomorphism defined on Σ^* is uniquely determined by its values on Σ. For example, if $h(a) = ccc$ and $h(b) = dd$, then

$$h(abaab) = h(a)h(b)h(a)h(a)h(b) = cccddcccccddd.$$

Moreover, any map $h : \Sigma \to \Gamma^*$ extends uniquely by induction to a homomorphism defined on all of Σ^*. Therefore, in order to specify a homomorphism completely, we need only say what values it takes on elements of Σ.

If $A \subseteq \Sigma^*$, define

$$h(A) \stackrel{\text{def}}{=} \{h(x) \mid x \in A\} \subseteq \Gamma^*,$$

and if $B \subseteq \Gamma^*$, define

$$h^{-1}(B) \stackrel{\text{def}}{=} \{x \mid h(x) \in B\} \subseteq \Sigma^*.$$

The set $h(A)$ is called the *image* of A under h, and the set $h^{-1}(B)$ is called the *preimage* of B under h.

We will show two useful closure properties of the regular sets: any homomorphic image or homomorphic preimage of a regular set is regular.

Theorem 10.1 *Let $h : \Sigma^* \to \Gamma^*$ be a homomorphism. If $B \subseteq \Gamma^*$ is regular, then so is its preimage $h^{-1}(B)$ under h.*

Proof. Let $M = (Q, \Gamma, \delta, s, F)$ be a DFA such that $L(M) = B$. Create a new DFA $M' = (Q, \Sigma, \delta', s, F)$ for $h^{-1}(B)$ as follows. The set of states, start state, and final states of M' are the same as in M. The input alphabet is Σ instead of Γ. The transition function δ' is defined by

$$\delta'(q, a) \stackrel{\text{def}}{=} \widehat{\delta}(q, h(a)).$$

Note that we have to use $\widehat{\delta}$ on the right-hand side, since $h(a)$ need not be a single letter.

Now it follows by induction on $|x|$ that for all $x \in \Sigma^*$,

$$\widehat{\delta'}(q, x) = \widehat{\delta}(q, h(x)). \tag{10.3}$$

For the basis $x = \epsilon$, using (10.1),

$$\widehat{\delta'}(q, \epsilon) = q = \widehat{\delta}(q, \epsilon) = \widehat{\delta}(q, h(\epsilon)).$$

For the induction step, assume that $\widehat{\delta'}(q, x) = \widehat{\delta}(q, h(x))$. Then

$$\begin{aligned}
\widehat{\delta'}(q, xa) &= \delta'(\widehat{\delta'}(q, x), a) && \text{definition of } \widehat{\delta'} \\
&= \delta'(\widehat{\delta}(q, h(x)), a) && \text{induction hypothesis} \\
&= \widehat{\delta}(\widehat{\delta}(q, h(x)), h(a)) && \text{definition of } \delta' \\
&= \widehat{\delta}(q, h(x)h(a)) && \text{Homework 1, Exercise 3} \\
&= \widehat{\delta}(q, h(xa)) && \text{property (10.2) of homomorphisms.}
\end{aligned}$$

Now we can use (10.3) to prove that $L(M') = h^{-1}(L(M))$. For any $x \in \Sigma^*$,

$$\begin{aligned}
x \in L(M') &\iff \widehat{\delta'}(s, x) \in F && \text{definition of acceptance} \\
&\iff \widehat{\delta}(s, h(x)) \in F && \text{by (10.3)} \\
&\iff h(x) \in L(M) && \text{definition of acceptance} \\
&\iff x \in h^{-1}(L(M)) && \text{definition of } h^{-1}(L(M)).
\end{aligned}$$

□

Theorem 10.2 *Let $h : \Sigma^* \to \Gamma^*$ be a homomorphism. If $A \subseteq \Sigma^*$ is regular, then so is its image $h(A)$ under h.*

Proof. For this proof, we will use regular expressions. Let α be a regular expression over Σ such that $L(\alpha) = A$. Let α' be the regular expression obtained by replacing each letter $a \in \Sigma$ appearing in α with the string $h(a) \in \Gamma^*$. For example, if $h(a) = ccc$ and $h(b) = dd$, then

$$((a+b)^*ab)' = (ccc + dd)^*cccdd.$$

Formally, α' is defined by induction:

$$\begin{aligned} a' &= h(a), \quad a \in \Sigma, \\ \emptyset' &= \emptyset, \\ (\beta + \gamma)' &= \beta' + \gamma', \\ (\beta\gamma)' &= \beta'\gamma', \\ \beta^{*\prime} &= \beta'^*. \end{aligned}$$

We claim that for any regular expression β over Σ,

$$L(\beta') = h(L(\beta)); \tag{10.4}$$

in particular, $L(\alpha') = h(A)$. This can be proved by induction on the structure of β. To do this, we will need two facts about homomorphisms: for any pair of subsets $C, D \subseteq \Sigma^*$ and any family of subsets $C_i \subseteq \Sigma^*$, $i \in I$,

$$h(CD) = h(C)h(D), \tag{10.5}$$

$$h\Big(\bigcup_{i\in I} C_i\Big) = \bigcup_{i\in I} h(C_i). \tag{10.6}$$

To prove (10.5),

$$\begin{aligned} h(CD) &= \{h(w) \mid w \in CD\} \\ &= \{h(yz) \mid y \in C,\ z \in D\} \\ &= \{h(y)h(z) \mid y \in C,\ z \in D\} \\ &= \{uv \mid u \in h(C),\ v \in h(D)\} \\ &= h(C)h(D). \end{aligned}$$

To prove (10.6),

$$\begin{aligned} h\Big(\bigcup_i C_i\Big) &= \{h(w) \mid w \in \bigcup_i C_i\} \\ &= \{h(w) \mid \exists i\ w \in C_i\} \\ &= \bigcup_i \{h(w) \mid w \in C_i\} \\ &= \bigcup_i h(C_i). \end{aligned}$$

Now we prove (10.4) by induction. There are two base cases:

$$L(a') = L(h(a)) = \{h(a)\} = h(\{a\}) = h(L(a))$$

and

$$L(\emptyset') = L(\emptyset) = \varnothing = h(\varnothing) = h(L(\emptyset)).$$

The case of ϵ is covered by the other cases, since $\epsilon = \varnothing^*$.

There are three induction cases, one for each of the operators $+$, $\cdot$, and *. For $+$,

$$\begin{aligned}
L((\beta+\gamma)') &= L(\beta'+\gamma') && \text{definition of } ' \\
&= L(\beta') \cup L(\gamma') && \text{definition of } + \\
&= h(L(\beta)) \cup h(L(\gamma)) && \text{induction hypothesis} \\
&= h(L(\beta) \cup L(\gamma)) && \text{property (10.6)} \\
&= h(L(\beta+\gamma)) && \text{definition of } +.
\end{aligned}$$

The proof for $\cdot$ is similar, using property (10.5) instead of (10.6). Finally, for *,

$$\begin{aligned}
&L(\beta^{*\prime}) \\
&= L(\beta'^*) && \text{definition of } ' \\
&= L(\beta')^* && \text{definition of regular expression operator } ^* \\
&= h(L(\beta))^* && \text{induction hypothesis} \\
&= \bigcup_{n\geq 0} h(L(\beta))^n && \text{definition of set operator } ^* \\
&= \bigcup_{n\geq 0} h(L(\beta)^n) && \text{property (10.5)} \\
&= h\Big(\bigcup_{n\geq 0} L(\beta)^n\Big) && \text{property (10.6)} \\
&= h(L(\beta)^*) && \text{definition of set operator } ^* \\
&= h(L(\beta^*)) && \text{definition of regular expression operator } ^*. \qquad \square
\end{aligned}$$

Warning: It is *not* true that A is regular whenever $h(A)$ is. This is not what Theorem 10.1 says. We will show later that the set $\{a^n b^n \mid n \geq 0\}$ is not regular, but the image of this set under the homomorphism $h(a) = h(b) = a$ is the regular set $\{a^n \mid n \text{ is even}\}$. The preimage $h^{-1}(\{a^n \mid n \text{ is even}\})$ is not $\{a^n b^n \mid n \geq 0\}$, but $\{x \in \{a,b\}^* \mid |x| \text{ is even}\}$, which is regular.

Automata with ϵ-transitions

Here is an example of how to use homomorphisms to give a clean treatment of ϵ-transitions. Define an *NFA with ϵ-transitions* to be a structure

$$M = (Q, \Sigma, \epsilon, \Delta, S, F)$$

such that ϵ is a special symbol not in Σ and

$$M_\epsilon = (Q, \Sigma \cup \{\epsilon\}, \Delta, S, F)$$

is an ordinary NFA over the alphabet $\Sigma \cup \{\epsilon\}$. We define acceptance for automata with ϵ-transitions as follows: for any $x \in \Sigma^*$, M *accepts* x if there exists $y \in (\Sigma \cup \{\epsilon\})^*$ such that

- M_ϵ accepts y under the ordinary definition of acceptance for NFAs, and
- x is obtained from y by erasing all occurrences of the symbol ϵ; that is, $x = h(y)$, where

  $$h : (\Sigma \cup \{\epsilon\})^* \to \Sigma^*$$

  is the homomorphism defined by

  $$h(a) \stackrel{\text{def}}{=} a, \quad a \in \Sigma,$$
  $$h(\epsilon) \stackrel{\text{def}}{=} \epsilon.$$

In other words,

$$L(M) \stackrel{\text{def}}{=} h(L(M_\epsilon)).$$

This definition and the definition involving ϵ-closure described in Lecture 6 are equivalent (Miscellaneous Exercise 10). It is immediate from this definition and Theorem 10.2 that the set accepted by any finite automaton with ϵ-transitions is regular.

Hamming Distance

Here is another example of the use of homomorphisms. We can use them to give slick solutions to Exercise 3 of Homework 2 and Miscellaneous Exercise 8, the problems involving Hamming distance. Let $\Sigma = \{0, 1\}$ and consider the alphabet

$$\Sigma \times \Sigma = \left\{ \boxed{\begin{matrix}0\\0\end{matrix}}, \boxed{\begin{matrix}0\\1\end{matrix}}, \boxed{\begin{matrix}1\\0\end{matrix}}, \boxed{\begin{matrix}1\\1\end{matrix}} \right\}.$$

The elements of $\Sigma \times \Sigma$ are ordered pairs, but we write the components one on top of the other. Let $\mathbf{top} : \Sigma \times \Sigma \to \Sigma$ and $\mathbf{bottom} : \Sigma \times \Sigma \to \Sigma$ be the

two projections

$$\mathbf{top}\left(\begin{array}{|c|}\hline a \\ \hline b \\ \hline\end{array}\right) = a,$$

$$\mathbf{bottom}\left(\begin{array}{|c|}\hline a \\ \hline b \\ \hline\end{array}\right) = b.$$

These maps extend uniquely to homomorphisms $(\Sigma \times \Sigma)^* \to \Sigma^*$, which we also denote by **top** and **bottom**. For example,

$$\mathbf{top}\left(\begin{array}{|c|c|c|c|}\hline 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 1 & 1 \\ \hline\end{array}\right) = 0010,$$

$$\mathbf{bottom}\left(\begin{array}{|c|c|c|c|}\hline 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 1 & 1 \\ \hline\end{array}\right) = 0111.$$

Thus we can think of strings in $(\Sigma \times \Sigma)^*$ as consisting of two tracks, and the homomorphisms **top** and **bottom** give the contents of the top and bottom track, respectively.

For fixed k, let D_k be the set of all strings in $(\Sigma \times \Sigma)^*$ containing no more than k occurrences of

$$\begin{array}{|c|}\hline 0 \\ \hline 1 \\ \hline\end{array} \text{ or } \begin{array}{|c|}\hline 1 \\ \hline 0 \\ \hline\end{array}.$$

This is certainly a regular set. Note also that

$$D_k = \{x \in (\Sigma \times \Sigma)^* \mid H(\mathbf{top}(x), \mathbf{bottom}(x)) \leq k\},$$

where H is the Hamming distance function. Now take any regular set $A \subseteq \Sigma^*$, and consider the set

$$\mathbf{top}(\mathbf{bottom}^{-1}(A) \cap D_k). \tag{10.7}$$

Believe it or not, this set is exactly $N_k(A)$, the set of strings in Σ^* of Hamming distance at most k from some string in A. The set $\mathbf{bottom}^{-1}(A)$ is the set of strings whose bottom track is in A; the set $\mathbf{bottom}^{-1}(A) \cap D_k$ is the set of strings whose bottom track is in A and whose top track is of Hamming distance at most k from the bottom track; and the set (10.7) is the set of top tracks of all such strings.

Moreover, the set (10.7) is a regular set, because the regular sets are closed under intersection, homomorphic image, and homomorphic preimage.