13. Families of Languages

Finite automata over free monoids accept rational languages. Automata over other monoids can be used as acceptors for other classes of languages.

Definition 13.1. *The family of languages* $\mathfrak{F}(M, X)$ *determined by a monoid* M *and a nonempty subset* $X \subset M$ *is the collection of images* $\rho(X)$ *as* ρ *ranges over all rational relations from* M *to finitely generated free monoids.*

Example 13.2. $\mathfrak{F}(\{1\}, \{1\})$ *is the collection of rational languages. Indeed a rational relation* $\rho : \{1\} \to \Sigma^*$ *is just a set* $\{1\} \times R \subset \{1\} \times \Sigma^*$ *with* R *is rational, and consequently* $\rho(1) = R$.

$L \subset \Sigma^*$ is in $\mathfrak{F}(M, X)$ if and only if there is an automaton Γ over $M \times \Sigma^*$ such that L is the collection of all $w \in \Sigma^*$ with (m, w) the label of a successful path in Γ for some $m \in X$. We will say in this case that Γ accepts w. We have two notions of acceptance here. Γ accepts both a rational relation $\rho \subset M \times \Sigma^*$ and the language $L = \rho(X) \subset \Sigma^*$.

It can be shown that the context–free languages are $\mathfrak{F}(M_{cf}, 1)$ where M_{cf} is the countably generated monoid with zero

$$M_{cf} = \langle P_i, Q_i \mid P_i Q_i = 1, P_i Q_j = 0 \text{ if } i \neq j \rangle.$$

From this point of view a pushdown automaton accepting a language over Σ is a finite automaton over $M_{cf} \times \Sigma^*$. See Figure 1 for an example.

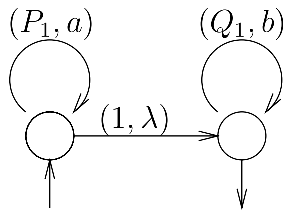

Figure 1. A pushdown automaton accepting $\{a^n b^n\}$

We see that the languages in $\mathfrak{F}(M, X)$ are characterized concretely in terms of automata at least if M and X are themselves sufficiently tangible. A more abstract characterization is given in the next theorem.

Theorem 13.3. *Let* $\mathfrak{F} = \mathfrak{F}(M, X)$ *be a family of languages.* $\mathfrak{F}$ *contains all rational languages and is closed under rational transduction and union. Conversely every collection of languages which contains a nonempty language and is closed under rational transduction and union is a family of languages.*

Proof. Pick $x \in X$. Each rational language $L \subset \Sigma^*$ is the image $\rho(X)$ under the rational transduction $\rho = \{x\} \times L$. Thus $L \in \mathfrak{F}$.

To show that $\mathfrak{F}$ is closed under union we must show that if L and L' are languages over Σ in $\mathfrak{F}$, then so is $L \cup L'$. It follows immediately from Definition 13.1 that enlarging Σ does not affect whether or not $L \in \mathfrak{F}$. Thus if L' were defined over Σ', we could replace Σ by $\Sigma + \Sigma'$.

$L = \rho(X)$ and $L' = \rho'(X)$ for rational relations $\rho : M \to \Sigma^*$ and $\rho' : M \to \Sigma^*$. It follows that $L \cup L' = (\rho \cup \rho')(X)$; and since the union of rational relations is rational, we are done. Likewise if $\tau : \Sigma^* \to \Delta^*$ is a rational transduction and $L \in \mathfrak{F}$, then $\tau(L) = \tau \circ \rho(L) \in \mathfrak{F}$ because $\tau \circ \rho$ is rational.

For the last part of the theorem let $\mathfrak{C}$ be a collection of languages containing a nonempty language and closed under transduction and union. Observe that if we alter a language by substituting new letters for its alphabet, then each version of the language is the image of the other under a homomorphism. Since these homomorphisms are rational transductions, we may pick a sub–collection $\mathfrak{C}_0$ of languages defined over disjoint alphabets such that every language in $\mathfrak{C}$ is the image under homomorphism of some language in $\mathfrak{C}_0$. Now let M be the free monoid over the union of the alphabets of languages in $\mathfrak{C}_0$, and define X to be the corresponding union of languages. Since $\mathfrak{C}$ contains a nonempty language, $X \neq \phi$. Take $\mathfrak{F} = \mathfrak{F}(M, X)$.

Every language in $\mathfrak{C}_0$ is the image of X under the partial identity homomorphism whose domain is the free submonoid of M generated by the alphabet corresponding to that language. Hence every language in $\mathfrak{C}$ is the image of X under a partial homomorphism whose domain is a finitely generated free submonoid of M. Consequently $\mathfrak{C} \subset \mathfrak{F}$.

It remains to show that $\mathfrak{F} \subset \mathfrak{C}$. Suppose $L = \rho(X)$ for some rational relation $\rho : M \to \Sigma^*$. We know that $\rho \subset M_0 \times \Sigma^*$ where M_0 is the free monoid over the union of a finite number of the alphabets used to define M. Consequently $L = \rho(X \cap M_0)$. By definition of X, $X \cap M_0$ is a finite union of languages in $\mathfrak{C}_0$. As $\mathfrak{C}$ is closed under union, $X \cap M_0 \in \mathfrak{C}$; and as $\mathfrak{C}$ is closed under transduction, $\rho(X \cap M_0) \in \mathfrak{C}$. □

Corollary 13.4. *Every family of languages is closed under intersection with rational sets.*

Proof. If L and R are languages over Σ, then $L \cap R = \rho_R(L)$ where $\rho_R = \{(w, w) \mid w \in R\}$. If R is rational, then ρ_R is a rational transduction. □

The next theorem shows that it makes sense to speak of the word problem of a group as being in a particular family.

Theorem 13.5. *Let $\mathfrak{F}$ be a family of languages. If the word problem for G with respect to one choice of generators is in $\mathfrak{F}$, then the word problem with respect to any choice of generators is also. Further the class of groups whose word problems lie in $\mathfrak{F}$ is closed under isomorphism, finitely generated subgroup and finite extension.*

Proof. Suppose $\sigma : \Sigma^* \to G$ is a choice of generators for G, and $\mu : \Delta^* \to H$ is one for the subgroup $H \subset G$. Let $W(G) = \sigma^{-1}(1)$ be the word problem of G with respect to the choice of generators σ, and let $W(H) = \mu^{-1}(1)$ be the word problem of H. There is a monoid homomorphism $f : \Delta^* \to \Sigma^*$ such that $\mu = \sigma \circ f$, and consequently $W(H) = f^{-1}(W(G))$. $W(G) \in \mathfrak{F}$ implies $W(H) \in \mathfrak{F}$. In particular when $H = G$ we see that whether or not $W(G) \in \mathfrak{F}$ depends only on G and not on the choice of generators.

Clearly the class of groups whose word problems lie in $\mathfrak{F}$ is closed under isomorphism. Thus it remains only to prove the last assertion of Theorem 13.5. At this point we are free to make any convenient choices of generators for G and H.

Make any choice of generators $\sigma : \Sigma^* \to G$ for G, and by taking H as the initial and terminal vertex turn the coset diagram for H in G into a finite automaton $\mathcal{A}$ accepting the language of all words in Σ^* representing elements of H. $\mathcal{A}$ admits a spanning tree Γ with root H and edges directed away from the root. The edges of $\mathcal{A}$ not in Γ determine a choice of generators $\mu : \Delta^* \to H$.

Alter the edge labels of $\mathcal{A}$ as follows. Replace each label a of an edge in Γ by (a, λ), and replace each label a of an edge in $\Gamma - \Gamma_0$ by (a, d) where $d \in \Delta$ is the element of Δ corresponding to that edge. Γ now accepts a rational transduction $\rho : \Sigma^* \to \Delta^*$ which rewrites each word in Σ^* representing an element of H as a word in Δ^* representing the same element. It follows immediately that $W(G) = \rho^{-1}(W(H)) \in \mathfrak{F}$. $\square$

Exercise 13.6. *Let G be a fixed finite group. Characterize the collection of groups with word problems in $\mathfrak{F}(G, 1)$.*