Myhill-Nerode Theorem for Sequential Transducers over Unique GCD-Monoids

Andreas Maletti*

Faculty of Computer Science, Dresden University of Technology
D–01062 Dresden, Germany. email: maletti@tcs.inf.tu-dresden.de

Abstract. We generalize the classical MYHILL-NERODE theorem for finite automata to the setting of sequential transducers over unique GCD-monoids, which are cancellative monoids in which every two non-zero elements admit a unique greatest common (left) divisor. We prove that a given formal power series is sequential, if and only if it is directed and our MYHILL-NERODE equivalence relation has finite index. As in the classical case, our MYHILL-NERODE equivalence relation also admits the construction of a minimal (with respect to the number of states) sequential transducer recognizing the given formal power series.

Deterministic finite automata and sequential transducers are applied, for example, in lexical analysis, digital image manipulation, and speech processing [2]. In the latter application area also very large sequential transducers, *i.e.*, transducers having several million states, over various monoids are encountered [2], so without minimization algorithms [4] the applicability of sequential transducers would be severely hampered.

In [2, 3] efficient algorithms for the minimization of sequential transducers are presented in case the weight is taken out of the monoid $(\Delta^*, \cdot, \varepsilon)$ or out of the monoid $(\mathbb{R}_+, +, 0)$. A MYHILL-NERODE theorem also allowing minimization is well-known for sequential transducers over groups [1].

We use $(A, \odot, \mathbf{1}, \mathbf{0})$ to denote a monoid with the absorbing element $\mathbf{0}$. A *unique GCD-monoid* is a cancellation monoid $(A, \odot, \mathbf{1}, \mathbf{0})$ in which (i) $a|\mathbf{1}$ implies $a = \mathbf{1}$, (ii) a greatest common divisor (gcd) exists for every two non-zero elements, and (iii) a least common multiple (lcm) exists for every two non-zero elements having a common multiple. Unique GCD-monoids exist in abundance (e.g., $(\mathbb{N} \cup \{\infty\}, +, 0, \infty)$ and $(\mathbb{N}, \cdot, 1, 0)$ as well as the monoids mentioned in the previous paragraph).

A *sequential transducer* (ST) is a tuple $M = (Q, q_0, F, \Sigma, \delta, \mathcal{A}, a_0, \mu)$ where (i) Q is a finite set, (ii) $q_0 \in Q$, (iii) $F \subseteq Q$, (iv) Σ is an alphabet, (v) $\delta\colon Q \times \Sigma \longrightarrow Q$, (vi) $\mathcal{A} = (A, \odot, \mathbf{1}, \mathbf{0})$ is a monoid, (vii) $a_0 \in A \setminus \{\mathbf{0}\}$,

* Financially supported by the German Research Foundation (DFG, GK 334/3)

and (viii) $\mu\colon Q \times \Sigma \longrightarrow A$. For every $q \in Q$ the mappings $\widehat{\delta}_q\colon \Sigma^* \longrightarrow Q$ and $\widehat{\mu}_q\colon \Sigma^* \longrightarrow A$ are recursively defined by (i) $\widehat{\delta}_q(\varepsilon) = q$ and $\widehat{\mu}_q(\varepsilon) = \mathbf{1}$, and for every $w \in \Sigma^*$ and $\sigma \in \Sigma$ by (ii) $\widehat{\delta}_q(w{\cdot}\sigma) = \delta(\widehat{\delta}_q(w), \sigma)$ and $\widehat{\mu}_q(w{\cdot}\sigma) = \widehat{\mu}_q(w) \odot \mu(\widehat{\delta}_q(w), \sigma)$. Finally, the power series $S_M \in A\langle\!\langle \Sigma^* \rangle\!\rangle$ *recognized by* M is then defined to be $(S_M, w) = a_0 \odot \widehat{\mu}_{q_0}(w)$, if $\widehat{\delta}_{q_0}(w) \in F$, otherwise $\mathbf{0}$. We call a power series $S \in A\langle\!\langle \Sigma^* \rangle\!\rangle$ *sequential* (with respect to $\mathcal{A}$), if there exists a sequential transducer M such that $S = S_M$.

In the following, let $\mathcal{A} = (A, \odot, \mathbf{1}, \mathbf{0})$ be a unique GCD-monoid, $M = (Q, q_0, F, \Sigma, \delta, \mathcal{A}, a_0, \mu)$ be a ST, and $S \in A\langle\!\langle \Sigma^* \rangle\!\rangle$. Moreover, we use $g(w) = \gcd_{u \in \Sigma^*,\, w\cdot u \in \mathrm{supp}(S)} (S, w{\cdot}u)$ for every $w \in \Sigma^*$. If $(S, w) = g(w)$ for all $w \in \mathrm{supp}(S)$, then S is called *directed.*

Definition 1. *The ST M is* normalized, *if there exists $\perp \in Q \setminus (F \cup \{q_0\})$ such that $\delta(\perp, \sigma) = \perp$ for every $\sigma \in \Sigma$ and $\mu(q, \sigma) = \mathbf{0} \iff \delta(q, \sigma) = \perp$ for every $q \in Q$.*

Definition 2. *We define the* Myhill-Nerode *relation $\equiv_S \subseteq \Sigma^* \times \Sigma^*$ by $w_1 \equiv_S w_2$, iff there exist $a_1, a_2 \in A \setminus \{\mathbf{0}\}$ such that for every $w \in \Sigma^*$*

$$w_1{\cdot}w \in \mathrm{supp}(S) \iff w_2{\cdot}w \in \mathrm{supp}(S) \;\; \textit{and} \;\; a_1^{-1} g(w_1{\cdot}w) = a_2^{-1} g(w_2{\cdot}w).$$

Proposition 3. *If S is directed and $\equiv_S$ has finite index, then there exists a sequential transducer M with* $\mathrm{index}(\equiv_S)$ *states such that $S_M = S$.*

Proof. In the proof we write $[w]$ and $[\Sigma^*]$ instead of $[w]_{\equiv_S}$ and $[\Sigma^*]_{\equiv_S}$. Let $M = (Q, q_0, F, \Sigma, \delta, \mathcal{A}, a_0, \mu)$ where for every $w \in \Sigma^*$ and $\sigma \in \Sigma$

(i) $Q = [\Sigma^*]$, $q_0 = [\varepsilon]$, $F = \{\, [w] \mid w \in \mathrm{supp}(S) \,\}$,
(ii) $\delta([w], \sigma) = [w{\cdot}\sigma]$, $a_0 = g(\varepsilon)$, and $\mu([w], \sigma) = g(w)^{-1} \odot g(w{\cdot}\sigma)$.

Moreover, the constructed ST is minimal with respect to the number of states amongst all normalized deterministic ST computing S.

Theorem 4. *The following are equivalent.*

(i) S is directed and $\equiv_S$ has finite index.
(ii) S is sequential.

References

1. Jack W. Carlyle and Azaria Paz. Realizations by stochastic finite automaton. *Journal of Computer and System Sciences*, 5(1):26–40, 1971.
2. Mehryar Mohri. Finite-state transducers in language and speech processing. *Computational Linguistics*, 23(2):269–311, 1997.
3. Mehryar Mohri. Minimization algorithms for sequential transducers. *Theoretical Computer Science*, 234(1–2):177–201, 2000.
4. Marcel P. Schützenberger and Christophe Reutenauer. Minimization of rational word functions. *SIAM Journal of Computing*, 20(4):669–685, 1991.