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.

1 Introduction

Deterministic finite automata (*e.g.*, [10, 20, 24]) and sequential transducers [7, 22, 3, 8] are applied, for example, in lexical analysis [1, 2], digital image manipulation [9], and speech processing [16]. In the latter application area also very large sequential transducers, *i.e.*, transducers having several million states, over various monoids are encountered [16], so without minimization algorithms [21, 23, 15] the applicability of sequential transducers would be severely hampered.

In [16, 17] efficient algorithms for the minimization of sequential transducers are presented in case the weight is taken out of the monoid $(\Delta^*, \cdot, \varepsilon)$ of words over Δ with the operation of concatenation or out of the monoid $(\mathbb{R}_+, +, 0)$ of non-negative reals with the usual addition. A Myhill-Nerode theorem also allowing minimization is well-known for sequential transducers over groups [6, 4] and in [13, 5] the authors prove Myhill-Nerode theorems for bottom-up finite tree automata and deterministic bottom-up weighted finite tree automata over arbitrary commutative groups, respectively. We present a generalization of the classical Myhill-Nerode [18, 19] congruence relation to the setting of sequential transducers over unique GCD-monoids [11, 12], in which every two non-zero

* Financially supported by the German Research Council (DFG, GRK 334/3)

elements admit a unique greatest common divisor. Roughly speaking, a sequential transducer $M = (Q, q_0, F, \Sigma, \delta, \mathcal{A}, a_0, \mu)$ comprises of

(i) a non-empty and finite set Q of states,
(ii) an initial state $q_0 \in Q$ in which the computation is started,
(iii) a set $F \subseteq Q$ of final states marking the end of successful computations,
(iv) a finite set Σ, also called input alphabet, of symbols over which the input words are formed,
(v) a mapping $\delta\colon Q \times \Sigma \longrightarrow Q$ yielding the next state provided the current state and input symbol,
(vi) a monoid $\mathcal{A} = (A, \odot, \mathbf{1})$ with an absorbing element $\mathbf{0}$,
(vii) a non-zero element $a_0 \in A \setminus \{\mathbf{0}\}$ standing for the weight of the empty word, and
(viii) a mapping $\mu\colon Q \times \Sigma \longrightarrow A$ which assigns a weight to each state transition.

At any given time the sequential transducer M is in a certain state of Q and has accumulated a weight of A. Initially, its internal state is q_0 and the weight is set to a_0. Then M is presented the input word w one symbol at a time, changes its internal state according to δ, and updates the accumulated weight by multiplying it with the weight obtained from μ. After the word w has been completely processed, M is either in a final state, which means that M accepts the word w and outputs the accumulated weight, or M rejects the word w and outputs $\mathbf{0}$. Hence M computes a mapping from Σ^* to A, which is then called *sequential*.

More formally, the mappings $\widehat{\delta}\colon \Sigma^* \longrightarrow Q$ and $\widehat{\mu}\colon \Sigma^* \longrightarrow A$ are recursively defined for every $w \in \Sigma^*$ and $\sigma \in \Sigma$ by

(i) $\widehat{\delta}(\varepsilon) = q_0$ and $\widehat{\mu}(\varepsilon) = a_0$, and
(ii) $\widehat{\delta}(w{\cdot}\sigma) = \delta(\widehat{\delta}(w), \sigma)$ and $\widehat{\mu}(w{\cdot}\sigma) = \widehat{\mu}(w) \odot \mu(\widehat{\delta}(w), \sigma)$.

Then the mapping $S_M\colon \Sigma^* \longrightarrow A$ computed by M (or equivalently the power series recognized by M) is defined as

$$S_M(w) = \begin{cases} \widehat{\mu}(w) & \text{, if } \widehat{\delta}(w) \in F \\ \mathbf{0} & \text{, otherwise} \end{cases} .$$

We will prove that a given power series S, *i.e.*, a mapping $S\colon \Sigma^* \longrightarrow A$ into a monoid $(A, \odot, \mathbf{1})$ with absorbing element $\mathbf{0}$, is sequential, if and only if (i) our Myhill-Nerode equivalence relation has finite index, and in addition, (ii) $S(w) = \gcd_{u \in \Sigma^*,\, S(w \cdot u) \neq \mathbf{0}} S(w{\cdot}u)$ whenever $S(w) \neq \mathbf{0}$. Moreover, in case S is sequential, the equivalence relation will also permit the construction of a minimal (with respect to the number of states) sequential transducer recognizing S.

The paper is structured as follows. Section 2 reviews the mathematical foundations required in the sequel. In particular, it formally introduces the key notions of unique GCD-monoids and sequential transducers. In Section 3 we present our generalization of the Myhill-Nerode theorem along with the minimization of sequential transducers using a construction similar to [17]. Finally, Section 4 contains conclusions. We present the constructions in the main part of the paper, while most proof details can be found in the appendix.

2 Preliminaries

Sets, Relations, and Words The set $\{0, 1, 2, \ldots\}$ of all non-negative integers is denoted by $\mathbb{N}$ and we let $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$. We write $\mathrm{card}(A)$ for the *cardinality* of a set A. Any subset $\rho \subseteq A \times A$ is called *relation on* A. Usually we prefer to write $a_1 \; \rho \; a_2$ instead of $(a_1, a_2) \in \rho$. Given a relation $\equiv$ on A, we say that $\equiv$ is an *equivalence relation*, if $\equiv$ is (i) *reflexive*, *i.e.*, for every $a \in A$ we have $a \equiv a$, (ii) *symmetric*, *i.e.*, $a_1 \equiv a_2$ if and only if $a_2 \equiv a_1$ for every $a_1, a_2 \in A$, and (iii) *transitive*, *i.e.*, for every $a_1, a_2, a_3 \in A$ the facts $a_1 \equiv a_2$ and $a_2 \equiv a_3$ imply $a_1 \equiv a_3$. The set $[a]_\equiv = \{ a' \in A \mid a \equiv a' \}$ is called the *equivalence class of* a (with respect to $\equiv$). Furthermore, we let $[A']_\equiv = \{ [a]_\equiv \mid a \in A' \}$ for every $A' \subseteq A$. The *index of* $\equiv$ is defined as $\mathrm{index}(\equiv) = \mathrm{card}([A]_\equiv)$.

A non-empty and finite set Σ is also called *alphabet*. In the following let Σ be an alphabet. Every finite sequence of elements of Σ is called a *word over* Σ and the set of all words over Σ is denoted by Σ^*. We use $w_1 \cdot w_2$ to denote the word obtained by concatenation of the two words $w_1, w_2 \in \Sigma^*$. In particular, we write ε for the *empty word*, *i.e.*, the sequence of length 0.

Monoids A *monoid* is defined to be an algebraic structure $\mathcal{A} = (A, \odot, \mathbf{1})$ with *carrier set* A, an *associative* operation $\odot \colon A^2 \longrightarrow A$, *i.e.*, we have $a_1 \odot (a_2 \odot a_3) = (a_1 \odot a_2) \odot a_3$ for every $a_1, a_2, a_3 \in A$, and an element $\mathbf{1} \in A$ such that $\mathbf{1} \odot a = a = a \odot \mathbf{1}$ for every $a \in A$. *Commutative* monoids additionally satisfy $a_1 \odot a_2 = a_2 \odot a_1$ for every $a_1, a_2 \in A$, and if there exists an element $\mathbf{0} \in A$ which acts as an *absorbing element*, *i.e.*, for every $a \in A$ we have $a \odot \mathbf{0} = \mathbf{0} = \mathbf{0} \odot a$, then this element is clearly unique and we use $(A, \odot, \mathbf{1}, \mathbf{0})$ to denote a monoid with the absorbing element $\mathbf{0}$. In case $\mathcal{A}$ has no absorbing element an absorbing element may be adjoined.

The monoid $(A, \odot, \mathbf{1}, \mathbf{0})$ is *zero-divisor free*, if $a_1 \odot a_2 = \mathbf{0}$ implies $a_1 = \mathbf{0}$ or $a_2 = \mathbf{0}$. In the sequel let $\mathcal{A} = (A, \odot, \mathbf{1}, \mathbf{0})$ be a monoid such that $\mathbf{0} \neq \mathbf{1}$.

Let $a, a_1, a_2 \in A$ such that $a \neq \mathbf{0}$. The monoid $\mathcal{A}$ is termed *(restricted) cancellation monoid*, if each of the two statements $a \odot a_1 = a \odot a_2$ and $a_1 \odot a = a_2 \odot a$ implies $a_1 = a_2$. We say that a_1 is a *(left) divisor* of a_2, in symbols $a_1 | a_2$, if there exists an $a \in A$ such that $a_1 \odot a = a_2$. Note that the element a is unique in a cancellation monoid, so that $a_1^{-1} \odot a_2$ denotes it provided that $a_1 | a_2$. Given two elements $a_1 \neq \mathbf{0} \neq a_2$, an element $a \in A$ is called *greatest common (left) divisor* (gcd) of a_1 and a_2, if (i) $a | a_1$, (ii) $a | a_2$, and (iii) for every $a' \in A$ satisfying $a' | a_1$ and $a' | a_2$ we have $a' | a$. Although greatest common divisors are neither guaranteed to exist nor unique, according to tradition we write $\gcd(a_1, a_2)$ to denote any gcd of a_1 and a_2. Dually to the notion of greatest common divisors the concept of least common multiples is defined. Precisely, a is a *least common (left) multiple* (lcm) of a_1 and a_2, if (i) $a_1 | a$, (ii) $a_2 | a$, and (iii) for every $a' \in A$ with $a_1 | a'$ and $a_2 | a'$ we have $a | a'$. 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 gcd exists for every two non-zero elements, and (iii) an lcm exists for every two non-zero elements having a common multiple. In particular this yields that every gcd is indeed unique. We extend the definition of a gcd to arbitrary many elements as follows. Let $k \in \mathbb{N}_+$ and $\{a_1, \ldots, a_k\} \subseteq A \setminus \{\mathbf{0}\}$.

$$\gcd_{i \in \{1,\ldots,k\}} a_i = \gcd(a_1, \gcd(a_2, \ldots, \gcd(a_{k-1}, a_k) \ldots)) \tag{1}$$

with $\gcd_{i \in \{1\}} a_i = a_1$. Given an infinite set I and a family $(a_i)_{i \in I}$, we define $\gcd_{i \in I} a_i = \gcd_{j \in J} a_j$, if there exists a finite set $J \subseteq I$ such that for every $i \in I$ there exists a $j \in J$ with $a_j | a_i$. Otherwise, we define $\gcd_{i \in I} a_i = \mathbf{1}$ and call this gcd *flawed*. For completeness we also define $\gcd_{i \in \emptyset} a_i = \mathbf{1}$.

Several important monoids are unique GCD-monoids such as

- the monoid $(\mathbb{N} \cup \{\infty\}, +, 0, \infty)$ of non-negative integers,
- the monoid $(\mathbb{N}, \cdot, 1, 0)$ of non-negative integers,
- the monoid $(\mathbb{R}_+ \cup \{0, \infty\}, +, 0, \infty)$ of non-negative reals,
- the monoid $(\Delta^*, \cdot, \varepsilon, \infty)$ of words with the absorbing element ∞,
- the monoid $(\mathbb{N}[\sqrt{2}], \cdot, 1, 0)$ of real numbers of the form $n_1 + n_2 \cdot \sqrt{2}$ with $n_1, n_2 \in \mathbb{N}$ (cf. [11, 12]), and
- generally every commutative factorial monoid with a single unit element is a unique GCD-monoid [11, 12].

Lemma 1. *Let $\mathcal{A} = (A, \odot, \mathbf{1}, \mathbf{0})$ be a cancellation monoid and $a, b, c \in A$ such that $a \neq \mathbf{0} \neq b$.*

(i) *Then* $a^{-1} \odot (b^{-1} \odot c) = (b \odot a)^{-1} \odot c$.
(ii) *If* $b|a$, *then* $b^{-1} \odot (a \odot c) = (b^{-1} \odot a) \odot c$.
(iii) *If* $b|a$, *then* $a^{-1} \odot (b \odot c) = (b^{-1} \odot a)^{-1} \odot c$.
(iv) *The monoid* $\mathcal{A}$ *is zero-divisor free.*

Formal Power Series and Sequential Transducers Any mapping $S \colon \Sigma^* \longrightarrow A$ is also called *(formal) power series* [14, 4]. The set of all such power series is denoted by $A\langle\langle \Sigma^* \rangle\rangle$. We write (S, w) instead of $S(w)$ for $S \in A\langle\langle \Sigma^* \rangle\rangle$ and $w \in \Sigma^*$. The *support* $\operatorname{supp}(S)$ of S is defined by $\operatorname{supp}(S) = \{ w \in \Sigma^* \mid (S, w) \neq \mathbf{0} \}$.

A *sequential transducer* [7, 22] is a tuple $M = (Q, q_0, F, \Sigma, \delta, \mathcal{A}, a_0, \mu)$ where (i) Q is a non-empty, finite set of *states*, (ii) $q_0 \in Q$ is an *initial state*, (iii) $F \subseteq Q$ is a set of *final states*, (iv) Σ is an alphabet, (v) $\delta \colon Q \times \Sigma \longrightarrow Q$ is a *transition mapping*, (vi) $\mathcal{A} = (A, \odot, \mathbf{1}, \mathbf{0})$ is a monoid, (vii) $a_0 \in A \setminus \{\mathbf{0}\}$ is a non-zero *initial weight*, and (viii) $\mu \colon Q \times \Sigma \longrightarrow A$ is a *weight mapping*. 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 $\sigma \in \Sigma$ (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 $S = S_M$.

Example 2. Let $\mathcal{A} = (\mathbb{N} \cup \{\infty\}, +, 0, \infty)$ and $\Sigma = \{a, b\}$. Then the sequential transducer $M = (\{\star\}, \star, \{\star\}, \Sigma, \delta, \mathcal{A}, 0, \mu)$ with $\delta(\star, a) = \delta(\star, b) = \star$ and $\mu(\star, a) = \mu(\star, b) = 1$ recognizes the power series S, which maps each word to its length.

3 Myhill-Nerode Equivalence Relation and Minimization

In this section we construct an equivalence relation $\equiv_S$ for a given power series $S \in A\langle\langle \Sigma^* \rangle\rangle$, where $\mathcal{A} = (A, \odot, \mathbf{1}, \mathbf{0})$ is a unique GCD-monoid and Σ is an arbitrary alphabet. Moreover, whenever S is sequential then the index of $\equiv_S$ will be finite. Therefore, we firstly define a certain normal form of sequential transducers and show that each sequential transducer M with k states can be transformed into a normalized sequential transducer M' with at most $(k + 1)$ states such that $S_M = S_{M'}$. Roughly speaking, a sequential transducer is normalized, if there exists a distinguished *dead state* $\bot \in Q \setminus F$ such that every transition from $q \in Q$ using $\sigma \in \Sigma$ with weight $\mu(q, \sigma) = \mathbf{0}$ leads to $\bot$, *i.e.*, $\delta(q, \sigma) = \bot$.

Definition 3. *Let* $M = (Q, q_0, F, \Sigma, \delta, \mathcal{A}, a_0, \mu)$ *be a sequential transducer. We say that* M *is* normalized*, if there exists a state* $\perp \in Q \setminus F$ *with* $\perp \neq q_0$ *such that for every* $\sigma \in \Sigma$ *we have* $\delta(\perp, \sigma) = \perp$ *and for every* $q \in Q$ *we have* $\mu(q, \sigma) = \mathbf{0} \iff \delta(q, \sigma) = \perp$.

Proposition 4. *For every non-normalized sequential transducer* $M = (Q, q_0, F, \Sigma, \delta, \mathcal{A}, a_0, \mu)$ *there exists a normalized sequential transducer* M' *with at most* $\big(\mathrm{card}(Q) + 1\big)$ *states such that* $S_M = S_{M'}$.

Proof (of Proposition 4). Let $\perp \notin Q$ and $Q' = Q \cup \{\perp\}$. The mappings $\delta' \colon Q' \times \Sigma \longrightarrow Q'$ and $\mu' \colon Q' \times \Sigma \longrightarrow A$ are defined for every $q \in Q'$ and $\sigma \in \Sigma$ by

(i) $\delta'(q, \sigma) = \delta(q, \sigma)$ and $\mu'(q, \sigma) = \mu(q, \sigma)$ whenever $q \in Q$, $\mu(q, \sigma) \neq \mathbf{0}$,
(ii) $\delta'(q, \sigma) = \perp$ and $\mu'(q, \sigma) = \mathbf{0}$ otherwise.

Then $M' = (Q', q_0, F, \Sigma, \delta, \mathcal{A}, a_0, \mu')$ is a normalized sequential transducer such that $S_{M'} = S_M$. The construction is standard, so we leave the proof details to the reader. □

The main beneficial property of normalized sequential transducers is stated in the following lemma. Since cancellation monoids are zero-divisor free (cf. Proposition 1(iv)), we have that the accumulated weight is zero, if and only if the sequential transducer is in a dead state. Henceforth, we will use $\perp$ to stand for a dead state.

Lemma 5. *Let* $M = (Q, q_0, F, \Sigma, \delta, \mathcal{A}, a_0, \mu)$ *be a normalized sequential transducer. Then for every* $w \in \Sigma^*$ *and* $q \in Q \setminus \{\perp\}$

$$\widehat{\mu}_q(w) = \mathbf{0} \quad \iff \quad \widehat{\delta}_q(w) = \perp \ . \tag{2}$$

Inspired by the Myhill-Nerode congruence relation [18, 19], we now define a similar relation for sequential transducers over unique GCD-monoids. We let $S \in A\langle\!\langle \Sigma^* \rangle\!\rangle$ be a power series in the sequel. Moreover, we will simply write that there exists $a \in A \setminus \{\mathbf{0}\}$ such that $a^{-1} \odot b = c$ to mean that there exists an a such that $a|b$ and $a \odot c = b$. Finally, for every $w \in \Sigma^*$ let $g(w) = \gcd_{u \in \Sigma^*,\, w \cdot u \in \mathrm{supp}(S)}(S, w \cdot u)$.

Definition 6. *We define the* Myhill-Nerode *relation* $\equiv_S \subseteq \Sigma^* \times \Sigma^*$ *for every* $w_1, w_2 \in \Sigma^*$ *as follows. We let* $w_1 \equiv_S w_2$*, if and only if there exist* $a_1, a_2 \in A \setminus \{\mathbf{0}\}$ *such that the following statements are well-formed and satisfied for every* $w \in \Sigma^*$.

$$w_1 \cdot w \in \mathrm{supp}(S) \iff w_2 \cdot w \in \mathrm{supp}(S) \tag{3}$$

$$a_1^{-1} \odot g(w_1 \cdot w) \quad = \quad a_2^{-1} \odot g(w_2 \cdot w) \tag{4}$$

Having defined $\equiv_S$ we now turn to its properties. Firstly, we observe that $\equiv_S$ is an equivalence relation on Σ^* (cf. Proposition 7) and secondly, whenever two words w_1 and w_2 are equivalent, then for every word w also $w_1{\cdot}w$ and $w_2{\cdot}w$ are equivalent (cf. Lemma 8).

Proposition 7. *The relation $\equiv_S$ is an equivalence relation on Σ^*.*

Lemma 8. *Let $w_1, w_2, w' \in \Sigma^*$. If $w_1 \equiv_S w_2$ then also $w_1{\cdot}w' \equiv_S w_2{\cdot}w'$.*

As in the classical case we obtain that the number of equivalence classes of $\equiv_S$, where S is a sequential power series recognized by a sequential transducer with k states, is at most $(k+1)$. Later on, we will show how to construct a sequential transducer from $\equiv_S$ provided that $\equiv_S$ has finite index and for every $w \in \mathrm{supp}(S)$ we have $(S, w) = g(w)$. Moreover, the constructed sequential transducer will have $\mathrm{index}(\equiv_S)$ many states, so together will the following proposition this shows that we can construct a minimal sequential transducer.

Proposition 9. *Let M be a sequential transducer with k states. If M is non-normalized, then $\mathrm{index}(\equiv_{S_M}) \leq k+1$, whereas $\mathrm{index}(\equiv_{S_M}) \leq k$, if M is normalized.*

Next we define directed power series, which are power series in which a support element w is assigned a weight which is the gcd of the weight of all support elements which have w as prefix. Clearly, every sequential power series is directed, which is stated in Lemma 11, and Proposition 12 shows that no gcd in Definition 6 is flawed, if $\equiv_S$ has finite index and S is directed. In particular, together with the previous proposition this means that any power series S, in which such a gcd is flawed, cannot be sequential.

Definition 10. *We call a power series $S \in A\langle\langle \Sigma^* \rangle\rangle$* directed*, if for every $w \in \mathrm{supp}(S)$ we have $(S, w) = g(w)$.*

Lemma 11. *Every sequential power series is directed.*

Proposition 12. *If S is directed and $\equiv_S$ has finite index, then there exists no $w \in \Sigma^*$ such that $g(w)$ is flawed.*

In the last proposition we show that we can actually implement $\equiv_S$ as a sequential transducer M, provided that S is directed and $\equiv_S$ has finite index. As in the classical construction, the state set of M will be the set of equivalence classes of $\equiv_S$. Our construction basically follows the construction of [17].

Proposition 13. *If $S \in A\langle\langle \Sigma^* \rangle\rangle$ 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 (of Proposition 13). In the proof we write $[w]$ and $[\Sigma^*]$ instead of $[w]_{\equiv_S}$ and $[\Sigma^*]_{\equiv_S}$, respectively, for every $w \in \Sigma^*$ in order to avoid too many subscripts. We construct $M = (Q, q_0, F, \Sigma, \delta, \mathcal{A}, a_0, \mu)$ by setting for every $w \in \Sigma^*$ and $\sigma \in \Sigma$

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

The proof of well-definedness and correctness, *i.e.*, $S_M = S$, can be found in the appendix. □

Finally, we are ready to state the main theorem. Note that in case $\mathcal{A} = (\{0, 1\}, \wedge, 1, 0)$ the classical Myhill-Nerode theorem coincides with our theorem.

Theorem 14. *Let $\mathcal{A} = (A, \odot, \mathbf{1}, \mathbf{0})$ be a unique GCD-monoid, Σ be an alphabet, and $S \in A\langle\langle \Sigma^* \rangle\rangle$. Then the following are equivalent.*

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

Proof (of Theorem 14). Proposition 13 proves the direction (i) $\Rightarrow$ (ii), whereas (ii) $\Rightarrow$ (i) can be concluded from Proposition 9 and Lemma 11.

The minimal sequential transducer can be obtained from the construction presented in the proof of Proposition 13, which is formalized in the final theorem.

Theorem 15. *Let $\mathcal{A} = (A, \odot, \mathbf{1}, \mathbf{0})$ be a unique GCD-monoid, Σ be an alphabet, and $S \in A\langle\langle \Sigma^* \rangle\rangle$ be directed. If the index of $\equiv_S$ is finite, then the sequential transducer M constructed in the proof of Proposition 13 is minimal with respect to the number of states amongst all normalized sequential transducers recognizing S.*

Proof (of Theorem 15). Note that M itself is not necessarily normalized, but the statement that every normalized sequential transducer recognizing S has at least $\mathrm{index}(\equiv_S)$ states was shown in Proposition 9.

Finally, we present an example showing an application of the above theorems. The example is simplistic on purpose; realistic examples can be found, *e.g.*, in [16, 17].

Example 16. Let $\mathcal{A} = (\mathbb{N}, \cdot, 1, 0)$ be the unique GCD-monoid of the non-negative integers and $\Sigma = \{a, b\}$. The power series $S \in \mathbb{N}\langle\langle \Sigma^* \rangle\rangle$ is defined for every $w \in \Sigma^*$ by $(S, w) = 2^{|w|_a} \cdot 3^{|w|_b}$ where $|w|_\sigma$ denotes the number of σ's occuring in w. One easily verifies that $\mathrm{supp}(S) = \Sigma^*$ and that S is directed. Moreover, we observe that $w_1 \equiv_S w_2$ for every $w_1, w_2 \in \Sigma^*$. Hence $\equiv_S$ has index 1. Note we again drop the actual equivalence relation from the equivalence classes. According to the construction of Proposition 13 we obtain the sequential transducer $M = (\{[\varepsilon]\}, [\varepsilon], \{[\varepsilon]\}, \Sigma, \delta, \mathcal{A}, g(\varepsilon), \mu)$ with

(i) $\delta([\varepsilon], a) = [a] = [\varepsilon]$ and $\delta([\varepsilon], b) = [b] = [\varepsilon]$,
(ii) $g(\varepsilon) = (S, \varepsilon) = 2^{|\varepsilon|_a} \cdot 3^{|\varepsilon|_b} = 1$, and
(iii) $\mu([\varepsilon], a) = g(\varepsilon)^{-1} \cdot g(a) = 2$ and $\mu([\varepsilon], b) = g(\varepsilon)^{-1} \cdot g(b) = 3$,

which according to Proposition 13 recognizes S and is furthermore minimal by Theorem 15.

4 Conclusions

We have presented a generalization of the classical Myhill-Nerode congruence relation. Moreover, we proved that the properties of $\equiv_S$ having finite index and S being directed exactly characterize the sequential property. As in the classical case, we also obtained a minimization for sequential transducers over unique GCD-monoids. We believe it worthwhile to generalize these results to the class of GCD-monoids, which would include all groups. Furthermore, a similar approach can also be applied to deterministic bottom-up weighted finite tree automata (cf. [5]) and we would like to see a generalized result also for formal tree series.

References

1. Alfred V. Aho, John E. Hopcroft, and Jeffrey D. Ullmann. *The Design and Analysis of Computer Algorithms.* Addison-Wesley, Reading, 1974.
2. Alfred V. Aho, Ravi Sethi, and Jeffrey D. Ullmann. *Compilers, Principles, Techniques and Tools.* Addison-Wesley, Reading, 1986.
3. Jean Berstel. *Transductions and Context-Free Languages.* Teubner Studienbücher, Stuttgart, 1979.
4. Jean Berstel and Christophe Reutenauer. *Rational Series and Their Languages*, volume 12 of *EATCS Monographs on Theoretical Computer Science.* Springer, Heidelberg, 1988.

5. Björn Borchardt. The Myhill-Nerode theorem for recognizable tree series. In *Seventh International Conference on Developments in Language Theory, Proceedings*, volume 2710 of *Lecture Notes in Computer Science*, pages 146–158. Springer, 2003.
6. Jack W. Carlyle and Azaria Paz. Realizations by stochastic finite automaton. *Journal of Computer and System Sciences*, 5(1):26–40, 1971.
7. Christian Choffrut. Une caractérisation des fonctions séquentielles et des fonctions sous-séquentielles en tant que relations rationelles. *Theoretical Computer Science*, 5(3):325–337, 1977.
8. Christian Choffrut. A generalization of Ginsburg and Rose's characterization of g-s-m mappings. In *Sixth International Colloquium on Automata, Languages, and Programming, Proceedings*, volume 71 of *Lecture Notes in Computer Science*, pages 88–103, Heidelberg, 1979. Springer.
9. Karel Culik II and Jarkko Kari. Digital images and formal languages. In Grzegorz Rozenberg and Arto Salomaa, editors, *Beyond Words*, volume 3 of *Handbook of Formal Languages*, chapter 10, pages 599–616. Springer, Heidelberg, 1997.
10. Samuel Eilenberg. *Automata, Languages, and Machines – Volume A*. Number 59 in Pure and Applied Mathematics. Academic Press, New York, 1974.
11. Nathan Jacobsen. *Basic Algebra I*. W. H. Freeman and Company, New York, second edition, 1985.
12. Nathan Jacobsen. *Basic Algebra II*. W. H. Freeman and Company, New York, second edition, 1989.
13. Dexter Kozen. On the Myhill-Nerode theorem for trees. *EATCS Bulletin*, 47:170–173, 1992.
14. Werner Kuich and Arto Salomaa. *Semirings, Automata, Languages*. EATCS Monographs on Theoretical Computer Science. Springer, 1986.
15. Mehryar Mohri. Minimization of sequential transducers. In *Fifth International Symposium on Combinatorial Pattern Matching, Proceedings*, volume 807 of *Lecture Notes in Computer Science*, pages 151–163, Heidelberg, 1994. Springer.
16. Mehryar Mohri. Finite-state transducers in language and speech processing. *Computational Linguistics*, 23(2):269–311, 1997.
17. Mehryar Mohri. Minimization algorithms for sequential transducers. *Theoretical Computer Science*, 234(1–2):177–201, 2000.
18. John Myhill. Finite automata and the representation of events. Technical Report 57-624, Wright Air Development Division, Ohio, 1957.
19. Anil Nerode. Linear automaton transformations. In *Proceedings of the AMS*, volume 9, pages 541–544. AMS, 1958.
20. Dominique Perrin. Finite automata. In Jan Van Leuwen, editor, *Formal Models and Semantics*, volume B of *Handbook of Theoretical Computer Science*, chapter 1, pages 1–57. Elsevier, Amsterdam, 1990.
21. Christophe Reutenauer. Subsequential functions: Characterizations, minimization, examples. In *Sixth International Meeting of Young Computer Scientists, Proceedings*, volume 464 of *Lecture Notes in Computer Science*, pages 62–79, Heidelberg, 1990. Springer.
22. Marcel P. Schützenberger. Sur une variante des fonctions séquentielles. *Theoretical Computer Science*, 4(1):47–57, 1977.
23. Marcel P. Schützenberger and Christophe Reutenauer. Minimization of rational word functions. *SIAM Journal of Computing*, 20(4):669–685, 1991.
24. Sheng Yu. Regular languages. In Grzegorz Rozenberg and Arto Salomaa, editors, *Word, Language, Grammar*, volume 1 of *Handbook of Formal Languages*, chapter 2, pages 41–110. Springer, Heidelberg, 1997.

Appendix

Proof (of Lemma 1). We prove the items separately.

(i) The following chain of equivalent statements shows the claim.

$$x = a^{-1} \odot (b^{-1} \odot c) \iff a \odot x = b^{-1} \odot c \iff b \odot a \odot x = c \tag{5}$$
$$\iff x = (b \odot a)^{-1} \odot c \tag{6}$$

(ii) This statement is trivial.

(iii) In the second line (8) we cancel b from the left.

$$x = a^{-1} \odot (b \odot c) \iff b \odot c = b \odot (b^{-1} \odot a) \odot x \tag{7}$$
$$\iff c = (b^{-1} \odot a) \odot x \tag{8}$$
$$\iff x = (b^{-1} \odot a)^{-1} \odot c \tag{9}$$

(iv) Let $a_1 \odot a_2 = \mathbf{0}$ for some $a_1, a_2 \in A \setminus \{\mathbf{0}\}$. Then $a_1 \odot a_2 = a_1 \odot \mathbf{0}$ and by the cancellation property also $a_2 = \mathbf{0}$, which is a contradiction to the assumption.

$\square$

Proof (of Lemma 5). The direction $\widehat{\delta}_q(w) = \perp$ implies $\widehat{\mu}_q(w) = \mathbf{0}$ is trivial. We only note that $\widehat{\delta}_q(\varepsilon) = q \neq \perp$. Now in order to prove the converse, let $\widehat{\mu}_q(w) = \mathbf{0}$. Clearly, $w \neq \varepsilon$, because $\widehat{\mu}_q(\varepsilon) = \mathbf{1}$ and we generally assumed that $\mathbf{0} \neq \mathbf{1}$. We prove the statement by induction on the length of w. Let w be a sequence of length 1, then $\widehat{\mu}_q(w) = \mu(q, w)$ and $\mu(q, w) = \mathbf{0}$, if and only if $\delta(q, w) = \perp$ by Definition 3. Hence in this case $\widehat{\delta}_q(w) = \perp$. Now let $w = u{\cdot}\sigma$ for some $u \in \Sigma^*$ and $\sigma \in \Sigma$. Then $\widehat{\mu}_q(u{\cdot}\sigma) = \widehat{\mu}_q(u) \odot \mu(\widehat{\delta}_q(u), \sigma) = \mathbf{0}$. By zero-divisor freeness we conclude that (i) $\widehat{\mu}_q(u) = \mathbf{0}$ or (ii) $\mu(\widehat{\delta}_q(u), \sigma) = \mathbf{0}$. The former yields by induction hypothesis that $\widehat{\delta}_q(u) = \perp$ and thus $\widehat{\delta}_q(u{\cdot}\sigma) = \delta(\widehat{\delta}_q(u), \sigma) = \delta(\perp, \sigma) = \perp$ by Definition 3. In case (ii) we conclude $\widehat{\delta}_q(u{\cdot}\sigma) = \delta(\widehat{\delta}_q(u), \sigma) = \perp$, because $\delta(\widehat{\delta}_q(u), \sigma) = \perp$, if and only if $\mu(\widehat{\delta}_q(u), \sigma) = \mathbf{0}$ according to Definition 3.

$\square$

Proof (of Proposition 7). Clearly, $\equiv_S$ is reflexive (set $a_1 = \mathbf{1} = a_2$) and symmetric. Moreover, transitivity of (3) is also trivial, so it remains to prove transitivity of (4). Let $w_1, w_2, w_3 \in \Sigma^*$ be such that $w_1 \equiv_S w_2$ and $w_2 \equiv_S w_3$. Consequently, there exist $a_1, a_2, a_2', a_3' \in A \setminus \{\mathbf{0}\}$ such that for every $w \in \Sigma^*$ we have $a_1^{-1} \odot g(w_1{\cdot}w) = a_2^{-1} \odot g(w_2{\cdot}w)$ and $(a_2')^{-1} \odot g(w_2{\cdot}w) = (a_3')^{-1} \odot g(w_3{\cdot}w)$. Since $\mathcal{A}$ is a unique GCD-semiring, we obtain that $\mathrm{lcm}(a_2, a'_2)$ exists because $g(w_2 \cdot w)$ is a common multiple of a_2 and a'_2. We deduce

$$a_2 \odot \bigl(a_1^{-1} \odot g(w_1 \cdot w)\bigr) = (a'_2) \odot \bigl((a'_3)^{-1} \odot g(w_3 \cdot w)\bigr) = a \qquad (10)$$

from the previous two equalities and observe that $a_2 | a$ and $a'_2 | a$. Hence also $\mathrm{lcm}(a_2, a'_2) | a$. Let $b_2, b'_2 \in A$ be such that $b_2 = a_2^{-1} \odot \mathrm{lcm}(a_2, a'_2)$ and $b'_2 = (a'_2)^{-1} \odot \mathrm{lcm}(a_2, a'_2)$. Consequently, multiplying (10) from the left with $\mathrm{lcm}(a_2, a'_2)$ we obtain

$$b_2^{-1} \odot \bigl(a_1^{-1} \odot g(w_1 \cdot w)\bigr) = (b'_2)^{-1} \odot \bigl((a'_3)^{-1} \odot g(w_3 \cdot w)\bigr) \qquad (11)$$

$$(a_1 \odot b_2)^{-1} \odot g(w_1 \cdot w) = (a'_3 \odot b'_2)^{-1} \odot g(w_3 \cdot w) \ , \qquad (12)$$

which establishes transitivity. Hence we have proved that $\equiv_S$ is an equivalence relation. □

Proof (of Lemma 8). If $w_1 \equiv_S w_2$ then there exist $a_1, a_2 \in A \setminus \{\mathbf{0}\}$ such that for every $w \in \Sigma^*$ Equations (3) and (4) hold. Consequently, also $w_1 \cdot w' \equiv_S w_2 \cdot w'$. □

Proof (of Proposition 9). We will only prove the case for non-normalized sequential transducers. The proof for normalized sequential transducers simply omits the first step in the proof. Henceforth, let M be a non-normalized sequential transducer. According to Proposition 4 there exists a normalized sequential transducer $M' = (Q, q_0, F, \Sigma, \delta, \mathcal{A}, a_0, \mu)$ such that $\mathrm{card}(Q) \leq k + 1$ and $S_{M'} = S_M$. Clearly, the relation $\equiv \subseteq \Sigma^* \times \Sigma^*$ defined for every $w_1, w_2 \in \Sigma^*$ by $w_1 \equiv w_2$, if and only if $\widehat{\delta}_{q_0}(w_1) = \widehat{\delta}_{q_0}(w_2)$ is an equivalence relation on Σ^*. We observe that $\mathrm{index}(\equiv) \leq \mathrm{card}(Q)$. So it is sufficient to prove $\equiv \subseteq \equiv_{S_M}$ in order to prove the statement. Therefore, let $w_1 \equiv w_2$, *i.e.*, $\widehat{\delta}_{q_0}(w_1) = \widehat{\delta}_{q_0}(w_2)$. Furthermore, let $a_1 = \widehat{\mu}_{q_0}(w_1)$ and $a_2 = \widehat{\mu}_{q_0}(w_2)$.

Case 1: Let $a_1 = \mathbf{0}$. By Lemma 5 we conclude that $\widehat{\delta}_{q_0}(w_1) = \perp = \widehat{\delta}_{q_0}(w_2)$ where $\perp$ is a dead state. Consequently, also $a_2 = \mathbf{0}$, which yields that for every $w \in \Sigma^*$ we have $(S_M, w_1 \cdot w) = \mathbf{0} = (S_M, w_2 \cdot w)$ and hence $w_1 \equiv_{S_M} w_2$.

Case 2: Let $a_1 \neq \mathbf{0}$ and $q = \widehat{\delta}_{q_0}(w_1) \neq \perp$. Immediately, we observe that also $a_2 \neq \mathbf{0}$ by Lemma 5. Let $g_M(u) = \mathrm{gcd}_{v \in \Sigma^*, u \cdot v \in \mathrm{supp}(S_M)}(S_M, u \cdot v)$ for every $u \in \{w_1, w_2\}$. Then clearly $\widehat{\mu}_{q_0}(u) | g_M(u)$ and for every $i \in \{1, 2\}$

$$a_i^{-1} \odot g_M(w_i) = \mathop{\mathrm{gcd}}_{w \in \Sigma^*, \widehat{\mu}_q(w) \neq \mathbf{0}} \widehat{\mu}_q(w) \ , \qquad (13)$$

which yields $a_1^{-1} \odot g_M(w_1 \cdot w) = a_2^{-1} \odot g_M(w_2 \cdot w)$. Moreover, since $\widehat{\delta}_q(w)$ is independent of w_1 and w_2 also $w_1 \cdot w \in \operatorname{supp}(S_M) \iff w_2 \cdot w \in \operatorname{supp}(S_M)$. Thus $w_1 \equiv_{S_M} w_2$ and we have proved the statement. $\square$

Proof (of Lemma 11). Let $M = (Q, q_0, F, \Sigma, \delta, \mathcal{A}, a_0, \mu)$ be a sequential transducer recognizing S, *i.e.*, $S_M = S$. Clearly, if $w \in \operatorname{supp}(S)$, then $q = \widehat{\delta}_{q_0}(w) \in F$. Then for every $u \in \Sigma^*$ such that $w \cdot u \in \operatorname{supp}(S)$ we observe that

$$
\begin{aligned}
(S, w \cdot u) &= a_0 \odot \widehat{\mu}_{q_0}(w \cdot u) && (14)\\
&= a_0 \odot \widehat{\mu}_{q_0}(w) \odot \widehat{\mu}_q(u) && (15)\\
&= (S, w) \odot \widehat{\mu}_q(u) \enspace , && (16)
\end{aligned}
$$

which shows $(S, w) | (S, w \cdot u)$ and hence directedness follows. $\square$

Proof (of Proposition 12). In order to derive a contradiction, assume that $\equiv_S$ has finite index and there exists a word $w \in \Sigma^*$ such that $g(w)$ is flawed. Then immediately the corresponding gcd for all words $w' \in \Sigma^*$ such that for some $u \in \Sigma^*$ we have $w' = w \cdot u$ is also flawed. Thus we obtain an infinite set $W = \{ w \cdot u \in \operatorname{supp}(S) \mid u \in \Sigma^* \}$ of words for which the gcd is flawed. Note that for all $w \in W$ we have $(S, w) \neq \mathbf{1}$. Now we consider a minimal subset $W' \subseteq W$ which has the property that for every $w \in W$ there exists a $w' \in W'$ such that $(S, w') | (S, w)$. Hence for every two distinct elements $w_1, w_2 \in W'$ we have that (S, w_1) is not a divisor of (S, w_2) and W' must apparently be infinite, else $g(w)$ is not flawed.

Since $\equiv_S$ has finite index, we have that for two distinct $w_1, w_2 \in W'$ it holds that $w_1 \equiv_S w_2$ by the pigeon-hole principle. Consequently, by Definition 6 there exist $a_1, a_2 \in A$ such that for every $w \in \Sigma^*$

$$
a_1^{-1} \odot g(w_1 \cdot w) = a_2^{-1} \odot g(w_2 \cdot w) \enspace . \tag{17}
$$

Since both gcd's are flawed, we deduce $a_1^{-1} \odot \mathbf{1} = a_2^{-1} \odot \mathbf{1}$, which yields that $a_1 = \mathbf{1} = a_2$. Moreover, also $a_1^{-1} \odot (S, w_1) = a_2^{-1} \odot (S, w_2)$ due to the directedness, which shows that $(S, w_1) = (S, w_2)$. However, this is contradictory, because $w_1, w_2 \in W'$. Hence for no $w \in \Sigma^*$ the gcd $g(w)$ is flawed which proves the statement. $\square$

Proof (of Proposition 13, continued). Firstly, let us prove well-definedness. Therefore, let $w_1, w_2 \in \Sigma^*$ such that $w_1 \equiv_S w_2$. Apparently, F is well-defined by (3) and according to Lemma 8 also $w_1 \cdot \sigma \equiv_S w_2 \cdot \sigma$ for every $\sigma \in \Sigma$, hence δ is well-defined. Since S is directed and $\equiv_S$ has finite index, for no $w \in \Sigma^*$ the gcd $g(w)$ is flawed (cf. Proposition 12), hence $g(w)|g(w\cdot\sigma)$. Thus μ is well-formed and we continue by proving

$$g(w_1)^{-1} \odot g(w_1\cdot\sigma) = g(w_2)^{-1} \odot g(w_2\cdot\sigma) \ . \tag{18}$$

By $w_1 \equiv_S w_2$ there exist $a_1, a_2 \in A \setminus \{\mathbf{0}\}$ such that for every $w \in \Sigma^*$ we have that $g(w_1\cdot w) = a_1 \odot (a_2^{-1} \odot g(w_2\cdot w))$. Consequently,

$$\begin{aligned} g(w_1)^{-1} \odot g(w_1\cdot\sigma) &= \big(a_1 \odot a_2^{-1} \odot g(w_2)\big)^{-1} \odot a_1 \odot a_2^{-1} \odot g(w_2\cdot\sigma) & (19)\\ &= g(w_2)^{-1} \odot g(w_2\cdot\sigma) \ , & (20) \end{aligned}$$

thereby proving that μ is well-defined. Consequently, M is well-defined and it remains to prove that $S_M = S$. For every $w \in \Sigma^*$ with $w \notin \mathrm{supp}(S)$ we immediately obtain $(S_M, w) = \mathbf{0}$, because $[w] \notin F$. On the other hand, let $w \in \mathrm{supp}(S)$, then

$$(S_M, w) = g(\varepsilon) \odot \widehat{\mu}_{[\varepsilon]}(w) = g(w) = (S, w) \ , \tag{21}$$

where the last equality follows from directedness. Hence $S_M = S$. $\square$