Myhill-Nerode Theorem for Sequential Transducers over GCD-Semirings

Andreas Maletti
Faculty of Computer Science; Dresden University of Technology; 01062 Dresden; GERMANY

Theorem

Sequential Transducers — Definition

A *sequential transducer* [3, 4] is a *weighted automaton* $(Q, \Sigma, \mathcal{A}, I, F, \mu)$ such that

- $I_q \neq \mathbf{0}$ for at most one $q \in Q$,
- $F \in \{\mathbf{0}, \mathbf{1}\}^Q$, and
- for every $q \in Q$ and $a \in \Sigma$ there exists at most one $p \in Q$ such that $\mu(a)_{q,p} \neq \mathbf{0}$.

Motivation

Sequential transducers are applied, e.g., in

- Text processing (pattern matching, indexing, compression),
- Natural language processing (recognition, synthesis),
- Image processing (filtering, compression)

Myhill-Nerode Congruence — Definition

Let $\mathcal{A}$ be a *unique GCD-semiring* and S be a *formal power series*. We define the Myhill-Nerode congruence 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 $w \in \Sigma^*$

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

Directedness — Definition

S is called *directed*, if $(S, w) = g(w)$ for all $w \in \operatorname{supp}(S)$ where

$$g(w) = \gcd_{u \in \Sigma^*,\, w \cdot u \in \operatorname{supp}(S)} (S, wu).$$

Minimal Sequential Transducer — Construction

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

Proof:
Let $M = (Q, \Sigma, \mathcal{A}, I, F, \mu)$ where for every $w \in \Sigma^*$ and $a \in \Sigma$

- $Q = [\Sigma^*]$,
- $I([w]) = g(\varepsilon)$, if $[w] = [\varepsilon]$, otherwise $I([w]) = \mathbf{0}$,
- $F([w]) = \mathbf{1}$, if $w \in \operatorname{supp}(S)$, otherwise $F([w]) = \mathbf{0}$, and
- $\mu(a)_{[w],[w \cdot a]} = g(w)^{-1} \odot g(w \cdot a)$, otherwise $\mu(a)_{q,p} = \mathbf{0}$.

Main Theorem

Theorem: The following are equivalent.
(i) S is *directed* and $\equiv_S$ has *finite index*.
(ii) S is *sequential*, i.e., there exists a *sequential transducer* M such that $S(M) = S$.

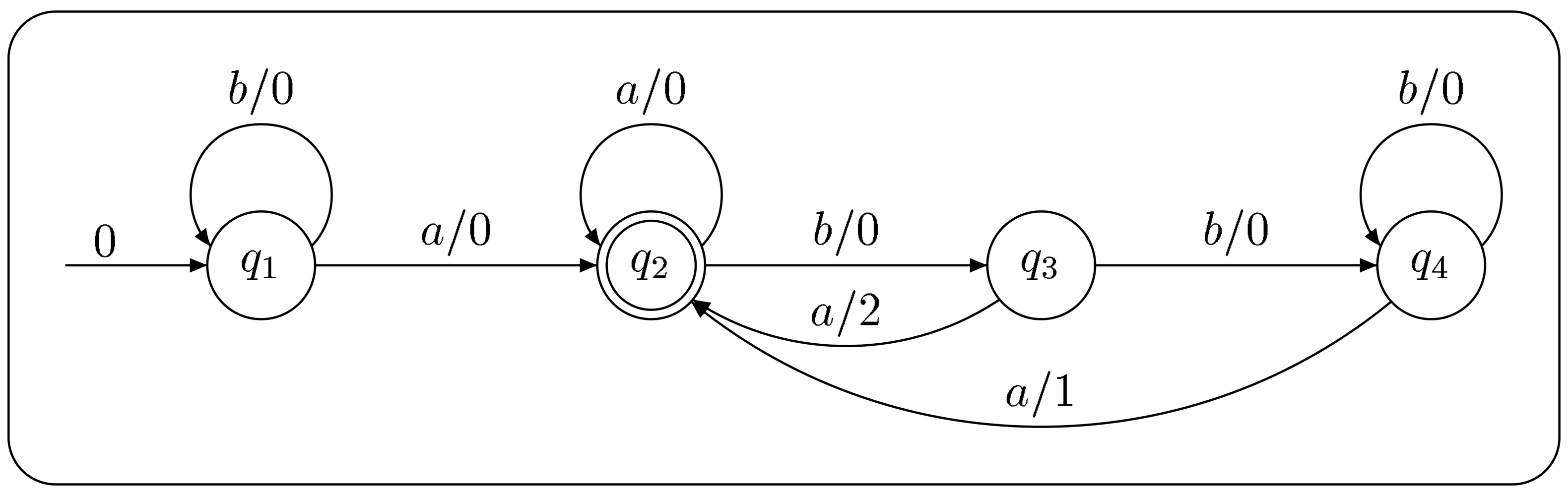
Example

Extension to GCD-Semirings

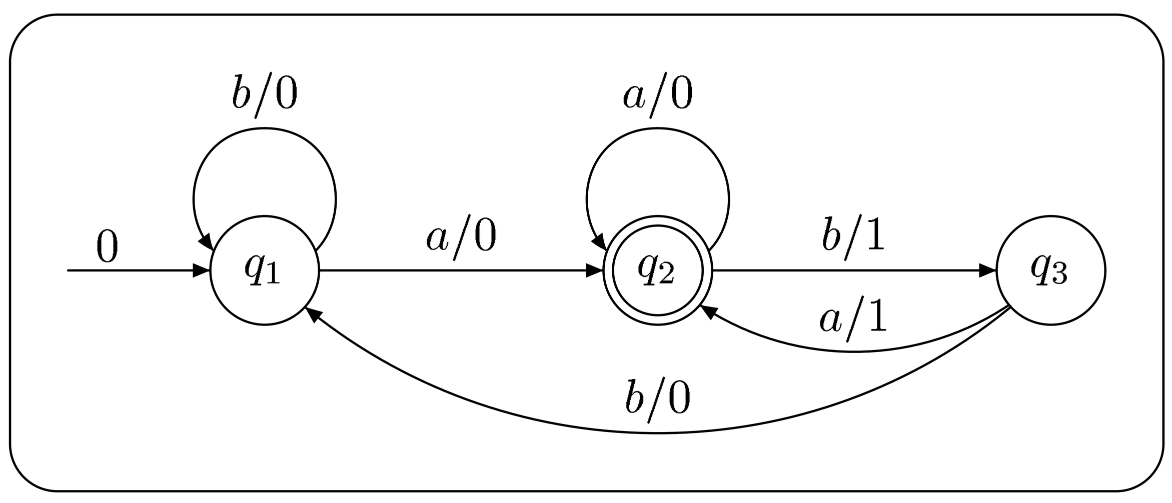
Corollary: Let $\mathcal{A}$ be a *GCD-semiring*. The following are equivalent.
(i) S is *directed*, and $\equiv_{[S]_\sim}$ and $\equiv_{S'}$ with $S' = ([S]_\sim)^{-1} \odot S$ have *finite index* [1, 5].
(ii) S is *sequential*.

Sequential Transducers — Examples

Non-minimal sequential transducer for $(S, w) = |w|_{aba} + |w|_{ab^+a}$, if $w = w' \cdot a$, otherwise $(S, w) = -\infty$:



Minimal sequential transducer:



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] Nathan Jacobsen. *Basic Algebra I*. W. H. Freeman and Company, New York, second edition, 1985.

[3] Mehryar Mohri. Finite-state transducers in language and speech processing. *Computational Linguistics*, 23(2):269–311, 1997.

[4] Mehryar Mohri. Minimization algorithms for sequential transducers. *Theoretical Computer Science*, 234(1–2):177–201, 2000.

[5] Marcel P. Schützenberger and Christophe Reutenauer. Minimization of rational word functions. *SIAM Journal of Computing*, 20(4):669–685, 1991.

Contact

Postal address: Andreas Maletti
Fakultät Informatik
Technische Universität Dresden
01062 Dresden
GERMANY

Email address: `maletti@tcs.inf.tu-dresden.de`