Information Processing Letters 140 (2018) 42–47

Contents lists available at ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

# On deterministic weighted automata ☆

Peter Kostolányi

*Department of Computer Science, Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Mlynská dolina, 842 48 Bratislava, Slovakia*

ARTICLE INFO

*Article history:*
Received 30 January 2018
Received in revised form 15 July 2018
Accepted 20 August 2018
Available online 27 August 2018
Communicated by Krishnendu Chatterjee

*Keywords:*
Weighted automaton
Determinism
Sequentiality
Pure sequentiality
Formal languages

ABSTRACT

Two families of input-deterministic weighted automata over semirings are considered: purely sequential automata, in which terminal weights of states are either zero or unity, and sequential automata, in which states can have arbitrary terminal weights. The class of semirings over which all weighted automata admit purely sequential equivalents is fully characterised. A similar characterisation is proved for sequential automata under an assumption that all elements of the underlying semiring have finitely many multiplicative left inverses, which is in particular true for all commutative semirings and all division semirings.



## 1. Introduction

Classical nondeterministic finite automata admit a well known extension, in which each transition is weighted (typically) by an element of some semiring [7,22]. Such automata – usually known as *weighted automata* – realise formal power series instead of recognising languages, and they have been studied extensively both from the theoretical point of view and in connection with their practical applications. The reader might consult [7] for an overview of some of the most important research directions.

For applications such as natural language processing [16,20], input-determinism of weighted automata often turns out to be crucial. This basically means that the automaton has precisely one state with nonzero initial weight and there is at most one transition for each input symbol leading from each state. However, it is well known that not all weighted automata can be determinised [19,21]. The research has thus focused mainly on providing sufficient conditions – both on the automaton and the underlying semiring – under which a weighted automaton admits a deterministic equivalent, and on devising efficient determinisation algorithms for automata satisfying such conditions [1,15,20,21].

We shall focus here on a slightly different question: over which semirings *all* weighted automata can be determinised? This in fact amounts to the study of deterministic weighted automata from a negative point of view, as we shall see that the class of such semirings is fairly constrained.

More precisely, we shall deal with this question for two classes of deterministic weighted automata: for *purely sequential* weighted automata, in which terminal weights of states might only be chosen as zero or unity of the underlying semiring, and for *sequential* weighted automata, in which terminal weights can be arbitrary (the term "deterministic weighted automata" usually refers to the latter [21]). This terminology follows Lombardy and Sakarovitch [19]; it may differ significantly in other sources (in particular, purely sequential automata are often called sequential, while sequential automata are called subsequential [20]).

☆ Partially supported by the grant VEGA 2/0165/16.
*E-mail address:* kostolanyi@fmph.uniba.sk.

https://doi.org/10.1016/j.ipl.2018.08.005
0020-0190/

We shall prove that weighted automata over $S$ always admit *purely sequential* equivalents if and only if $S$ is a locally finite division semiring. Moreover, local finiteness of $S$ is known to be sufficient to guarantee that all weighted automata over $S$ have *sequential* equivalents [19]. We shall prove that if $S$ has no element with infinitely many multiplicative left inverses, then this is also a necessary condition. In particular, if $S$ is commutative or a division semiring, then weighted automata over $S$ can always be sequentialised if and only if $S$ is locally finite.

Finally, let us mention that there is a branch of research motivated by quantitative formal verification dealing with weighted – or quantitative – automata over various structures beyond semirings [2–6,8–11,18]. We shall nevertheless confine ourselves to the classical setting of semirings in this article, making theoretical analysis more tractable. Possible extensions of the results presented herein to structures more general than semirings are left for further research.

## 2. Preliminaries

A *monoid* is a triple $(M, \cdot, 1)$, where $M$ is a set, $\cdot$ is an associative binary operation on $M$, and 1 is a neutral element with respect to $\cdot$. A *commutative monoid* is a monoid $(M, \cdot, 1)$ such that $\cdot$ is commutative. A *semiring* is a quintuple $(S, +, \cdot, 0, 1)$ such that $(S, +, 0)$ is a commutative monoid, $(S, \cdot, 1)$ is a monoid, the operation $\cdot$ distributes over $+$ both from left and from right, and $0 \cdot a = a \cdot 0 = 0$ holds for all $a$ in $S$. A *commutative semiring* is a semiring such that the monoid $(S, \cdot, 1)$ is commutative. A *division semiring* [12] is a (not necessarily commutative) semiring such that $(S - \{0\}, \cdot, 1)$ is a group, i.e., each nonzero element of $S$ has a multiplicative inverse. We shall often simply write $S$ for a semiring $(S, +, \cdot, 0, 1)$.

A *subsemiring* of a semiring $(S, +, \cdot, 0, 1)$ is a semiring $(S', +, \cdot, 0, 1)$ such that $S' \subseteq S$ (and $+, \cdot$ are restricted to $S' \times S'$). If moreover $X \subseteq S$ is a set, the *subsemiring of S generated by X* is the intersection of all subsemirings of $S$ containing $X$. A semiring $S$ is *finitely generated* if it is generated by some finite subset of $S$. A semiring $S$ is *locally finite* if every finitely generated subsemiring of $S$ is finite. Submonoids, finitely generated monoids, and locally finite monoids are defined similarly.

A *formal power series* over a semiring $S$ and over an alphabet $\Sigma$ is a mapping $r \colon \Sigma^* \to S$. It is customary to write $(r, w)$ instead of $r(w)$ for the value of $r$ on a word $w$ in $\Sigma^*$; the formal power series $r$ itself is then written as

$$r = \sum_{w \in \Sigma^*} (r, w) w.$$

The set of all formal power series over $S$ and $\Sigma$ is denoted by $S\langle\langle \Sigma^* \rangle\rangle$.

Let $r_1$ and $r_2$ be in $S\langle\langle \Sigma^* \rangle\rangle$. The series $r_1 + r_2$ is then defined by $(r_1 + r_2, w) = (r_1, w) + (r_2, w)$ for all $w$ in $\Sigma^*$ and the series $r_1 \cdot r_2$ is defined by

$$(r_1 \cdot r_2, w) = \sum_{\substack{u, v \in \Sigma^* \\ uv = w}} (r_1, u)(r_2, v)$$

for all $w$ in $\Sigma^*$. The set $S\langle\langle \Sigma^* \rangle\rangle$ constitutes a semiring together with these two operations [7].

Let $S$ be a semiring. A *proper weighted automaton* [7] over $S$ is a sextuple $\mathcal{A} = (Q, \Sigma, T, \nu, \iota, \tau)$, where $Q$ is a nonempty finite set of states, $\Sigma$ is a (finite) alphabet, $T \subseteq Q \times \Sigma \times Q$ is a set of transitions, $\nu \colon T \to S$ is a transition weighting function, $\iota \colon Q \to S$ is an initial weighting function, and $\tau \colon Q \to S$ is a terminal weighting function.

Moreover, a *run* in $\mathcal{A} = (Q, \Sigma, T, \nu, \iota, \tau)$ is a word $\gamma = p_1[p_1, c_1, q_1]p_2[p_2, c_2, q_2] \ldots p_n[p_n, c_n, q_n]p_{n+1}$ in $(QT)^*Q$ such that $n$ is a nonnegative integer and $q_i = p_{i+1}$ for $i = 1, \ldots, n$. We shall denote by $S_{\mathcal{A}}(\gamma) := p_1$ the *source* of the run $\gamma$, and by $D_{\mathcal{A}}(\gamma) := p_{n+1}$ the *destination* of $\gamma$. The *label* of a run $\gamma$ in $\mathcal{A}$ is given by $\lambda_{\mathcal{A}}(\gamma)$, where $\lambda_{\mathcal{A}} \colon (Q \cup T)^* \to \Sigma^*$ is a homomorphism such that $\lambda_{\mathcal{A}}(q) = \varepsilon$ for all $q$ in $Q$ and $\lambda_{\mathcal{A}}([p, c, q]) = c$ for all $[p, c, q]$ in $T$. The *weight* of a run $\gamma$ in $\mathcal{A}$ is given by $W_{\mathcal{A}}(\gamma)$, where $W_{\mathcal{A}} \colon (Q \cup T)^* \to (S, \cdot, 1)$ is a monoid homomorphism such that $W_{\mathcal{A}}(q) = 1$ for all $q$ in $Q$ and $W_{\mathcal{A}}([p, c, q]) = \nu([p, c, q])$ for all $[p, c, q]$ in $T$. If $\mathcal{A}$ is clear from the context, we shall usually write $S(\gamma)$, $D(\gamma)$, $\lambda(\gamma)$, and $W(\gamma)$ instead of $S_{\mathcal{A}}(\gamma)$, $D_{\mathcal{A}}(\gamma)$, $\lambda_{\mathcal{A}}(\gamma)$, and $W_{\mathcal{A}}(\gamma)$, respectively.

Let us denote by $R(\mathcal{A})$ the set of all runs in $\mathcal{A}$ and by $R(\mathcal{A}, w)$, where $w$ is in $\Sigma^*$, the set of all $\gamma$ in $R(\mathcal{A})$ such that $\lambda(\gamma) = w$. The *behaviour* of $\mathcal{A}$ then is a power series $\|\mathcal{A}\|$ defined by

$$(\|\mathcal{A}\|, w) = \sum_{\gamma \in R(\mathcal{A}, w)} \iota(S(\gamma)) W(\gamma) \tau(D(\gamma))$$

for all $w$ in $\Sigma^*$. Note that this sum is always finite, and thus well defined.

We shall always assume that $\nu([p, c, q]) \neq 0$ for all $[p, c, q]$ in $T$ – this is without loss of generality, as having a transition with zero weight is clearly equivalent to having no transition at all.

Let $\mathcal{A} = (Q, \Sigma, T, \nu, \iota, \tau)$ be a proper weighted automaton over a semiring $S$. The automaton $\mathcal{A}$ is *sequential* if $q_0$ in $Q$ exists such that $\iota(q_0) \neq 0$ and $\iota(q) = 0$ for all $q \neq q_0$ in $Q$, and if *at most one* $q$ in $Q$ with $[p, c, q]$ in $T$ exists for each $p$ in $Q$ and $c$ in $\Sigma$. Moreover, $\mathcal{A}$ is *purely sequential* if it is sequential and if $\tau(q)$ is in $\{0, 1\}$ for each $q$ in $Q$. Finally, $\mathcal{A}$ is *unambiguous* if $R(\mathcal{A}, w)$ contains at most one run $\gamma$ with $\iota(S(\gamma)) W(\gamma) \tau(D(\gamma)) \neq 0$ for each $w$ in $\Sigma^*$.

**Remark 2.1.** A purely sequential automaton *may* contain a state with *initial* weight equal neither to 0, nor to 1. However, it is easy to see that when it comes to series realised by purely sequential automata, restricting initial weights to 0 and 1 would only affect possible weights of the empty word.

**Remark 2.2.** We shall confine ourselves to the study of *proper* weighted automata, i.e., automata without transitions labelled by the empty word. This clearly has no effect on our results. Moreover, *all weighted automata are understood to be proper in what follows.*

It is well known that each weighted automaton over a locally finite semiring $S$ admits a sequential equiva-