# Determinization of Weighted Tree Automata using Factorizations

Matthias Büchse⋆,[a], Jonathan May⋆⋆,[b], and Heiko Vogler[a]

[a] Faculty of Computer Science
Technische Universität Dresden
01062 Dresden, Germany
{buechse,vogler}@tcs.inf.tu-dresden.de

[b] Information Sciences Institute
4676 Admiralty Way
Marina del Ray, CA 90292
jonmay@isi.edu

**Abstract.** Ranked lists of output trees from syntactic statistical NLP applications often suffer from multiple repeated entries, leading to misrepresentation and loss of information. This is chiefly due to nondeterminism in the structure from which the lists are generated. Since weighted tree automata are an appropriate representation of this kind of structure, we show a technique derived from Kirsten and Mäurer (2005) to determinize them. This eliminates repeated entries from the resulting list while preserving proper weights. Our technique verifies the results conjectured in May and Knight (2006) for nonrecursive wta and outlines requirements for generalization to recursive wta for certain weight domains.

## 1 Introduction

Natural language processing (NLP) systems that seek to answer some difficult questions such as "What is the English translation of this Chinese sentence?" or "What is the syntactic parse of this sentence?" typically propose a large or even infinite set of potential answers, each equipped with a weight. In order to make reasoning on such a large set of answers feasible, often only a subset of $k$ best answers (highest ranking according to their weights) is considered.

Frequently, these NLP systems are nondeterministic, i.e., one answer may be derived in a (large) number of ways; and the weight of an answer is accumulated from the weights of its derivations. For instance, many models produce answers by combining partial results of various sizes [1–5]; others use varying degrees of context to generate answers [6].

As noted in [1, 7, 8], many implementations in fact do not output a list of $k$ best answers, but of $k$ best derivations. This hurts performance because these lists can differ greatly, depending on the degree of nondeterminism. We conducted an experiment on a state-of-the-art syntax-based machine translation

⋆ Corresponding author.
⋆⋆ supported by NSF grant IIS-0428020

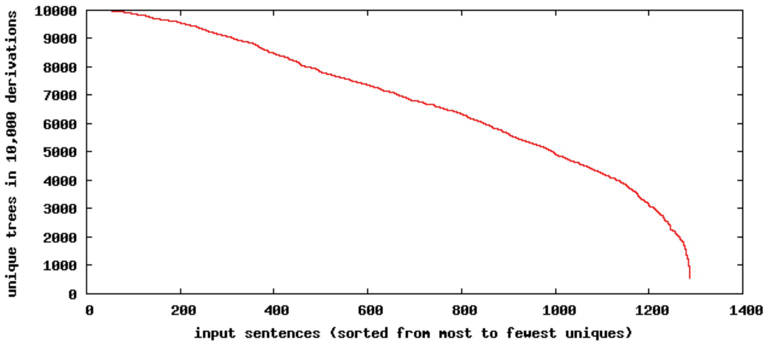



**Fig. 1.** For each of 1,287 Arabic test sentences the number of different English syntax trees obtained from the top 10,000 derivations is shown. The Arabic test sentences were sorted by this number in descending order.

system, similar to that described in [9]. The system is trained on 170,000 sentences of English-Arabic bitext, the English side parsed [6], and the sentences word-aligned with standard semi-supervised methods [10]. Fig. 1 indicates that nondeterminism was pervasive.

Let us briefly indicate three ways of obtaining a list of best answers, or an approximation thereof, from these implementations. The obvious noninvasive way for an approximation is to merge derivations in the output list according to the answer they belong to. This approach can necessitate a parameter increase: some $l \gg k$ derivations may have to be processed in order to obtain a fixed number $k$ of unique answers. Clearly, the actual weight and thus the model's preferential order of the answers will not be known, as the true weight may be spread across many derivations.

In order to open our discussion for invasive approaches, we consider operational models which can be used to describe NLP implementations. When the answers are drawn from a linear sequential model, such as a speech recognition lattice or an $n$-gram language model, an appropriate operational model is the weighted finite-state automaton (wfsa). For tree-based systems, which can be found in syntactic parsing and recently in machine translation [6, 11, 2], a suitable operational model is the weighted (finite-state) tree automaton (wta) [12, 13], a generalization of wfsa[1]. A list of $k$ best accepting paths (corresponding to derivations of answers) can be obtained efficiently for these automata [16, 17]. Naturally, the nondeterminism of the NLP systems is reflected in the automata, i.e., they may contain several paths that represent the same answer.

A second approach for approximating a list of $k$ best answers may now be obtained by supplementing the first approach with a method of pruning the

[1] Other widely used representations of a potentially infinite tree space in syntactic NLP literature are packed forests [14] and regular tree grammars [15], however weighted forms of all of these representations are isomorphic (under acceptable conditions) and wta are more appropriate for the notions presented here.

automaton. This can reduce the parameter increase required, but may entirely remove some unique answers. A third, more principled approach is determinization. Determinization of an automaton generates an equivalent deterministic automaton, i.e., one that recognizes the same object with the same weight, but by at most one accepting path. Thus, for deterministic automata, the list of $k$ best paths coincides with the list of $k$ best answers. While determinization can in general be costly (having a worst-case time complexity which is exponential in the number of states), it comes without the downsides of an approximation.

Starting from the well-known powerset construction, Mohri [18] developed an algorithm for determinizing wfsa over the tropical semiring and showed sufficient conditions for its termination. Kirsten and Mäurer [19] generalized this approach to a large class of semirings by introducing factorizations. Determinization of wta was first described by Borchardt [20], however he required the semiring to be locally finite. May and Knight [8] applied Mohri's approach to nonrecursive wta over the semiring of nonnegative reals, and they provided empirical evidence that their algorithm was effective on repetitive output from machine translation and parsing systems, but gave no formal proof of correctness.

In this paper, we generalize the factorization approach of [19] from wfsa to wta. We give a formal construction and a proof of its correctness, and we show sufficient conditions for its termination. Our work generalizes [8] by also considering the case of recursive wta. Following [19], we require in this case that (i) the wta have the twins property and (ii) the semiring be extremal. Since Property (i) is undecidable, further research is needed to turn our result for recursive wta into an algorithm. Property (ii) implies that only the best derivation of each answer is taken into account.

## 2 Preliminaries

A pair $(\Sigma, rk)$ is a *ranked alphabet* if $\Sigma$ is an alphabet and $rk : \Sigma \to \mathbb{N}$. Let $(\Sigma, rk)$ be a ranked alphabet. If $rk$ is clear from the context, then for every $k \in \mathbb{N}$ we write $\Sigma^{(k)}$ instead of $rk^{-1}(k)$. Henceforth, we identify $(\Sigma, rk)$ with $\Sigma$.

Let $V$ be a set. We define the *set of trees over $\Sigma$ indexed by $V$*, denoted by $T_\Sigma(V)$, to be the smallest set $T$ such that $V \subseteq T$ and for every $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, and $\xi_1, \ldots, \xi_k \in T$, also $\sigma(\xi_1, \ldots, \xi_k) \in T$. We set $T_\Sigma = T_\Sigma(\emptyset)$. Let $\xi \in T_\Sigma(V)$. We define the set of *positions of* $\xi$ as follows. If $\xi \in V$, then $pos(\xi) = \{\varepsilon\}$. If $\xi = \sigma(\xi_1, \ldots, \xi_k)$, then $pos(\xi) = \{\varepsilon\} \cup \{iw \mid i \in \{1, \ldots, k\}, w \in pos(\xi_i)\}$. Let $w \in pos(\xi)$. The label of $\xi$ at $w$ is denoted by $\xi(w)$ and the subtree of $\xi$ rooted at $w$ is denoted by $\xi|_w$.

A tuple $\mathcal{A} = (A, +, \cdot, 0, 1)$ is a *semiring* if $A$ is a set, $+$ and $\cdot$ are binary, associative operations on $A$, $+$ is commutative, $\cdot$ distributes over $+$, 0 and 1 are elements of $A$, 0 is neutral with respect to $+$, 1 is neutral with respect to $\cdot$, and 0 is absorbing with respect to $\cdot$ (i.e., $a \cdot 0 = 0 = 0 \cdot a$).

Let $\mathcal{A} = (A, +, \cdot, 0, 1)$ be a semiring. In notation, we will identify $\mathcal{A}$ with $A$. We say that $A$ is *commutative* if the operation $\cdot$ is commutative, and $A$ is *zero-divisor free* if $a \cdot b = 0$ implies that $a = 0$ or $b = 0$.

*Example 1.* We give examples of semirings (commutative except no. 5):

1. the *Boolean semiring* $(\mathbb{B}, \vee, \wedge, 0, 1)$ where $\mathbb{B} = \{0, 1\}$, and $\vee$ and $\wedge$ denote disjunction and conjunction, respectively;
2. the *semiring of nonnegative real numbers* $(\mathbb{R}^{\geq 0}, +, \cdot, 0, 1)$,
3. the *Viterbi semiring* $([0, 1], \max, \cdot, 0, 1)$;
4. the *tropical semiring* $(\mathbb{R}^{\geq 0}_{\infty}, \min, +, \infty, 0)$ where $\mathbb{R}^{\geq 0}_{\infty} = \mathbb{R}^{\geq 0} \cup \{\infty\}$;
5. the *formal language semiring* $(\mathcal{P}(\Delta^*), \cup, \cdot, \emptyset, \{\varepsilon\})$ over an alphabet $\Delta$ where $\cdot$ denotes the usual concatenation of languages. □

*For the rest of this paper, let $\Sigma$ be an arbitrary ranked alphabet, and let $(A, +, \cdot, 0, 1)$ be an arbitrary commutative semiring.*

Let $Q$ be a set. The set of all mappings from $Q$ into $A$ (in some cases also called *$Q$-vectors over $A$*) is denoted by $A^Q$. For every $u \in A^Q$ and $q \in Q$, we denote $u(q)$ by $u_q$, the *$q$-component of $u$*. The $Q$-vector over $A$ which maps every $q \in Q$ to 0 is denoted by $\tilde{0}$. A *(formal) tree series (over $\Sigma$ and $A$)* is a mapping $\varphi : T_\Sigma \to A$. For every $\xi \in T_\Sigma$, we call $\varphi(\xi)$ the *coefficient of $\xi$ with respect to $\varphi$* and denote it by $(\varphi, \xi)$. The set of all tree series over $\Sigma$ and $A$ is denoted by $A\langle\langle T_\Sigma \rangle\rangle$. The tree series which maps every tree to 0 is denoted by $\tilde{0}$.

## 3 Weighted Tree Automata

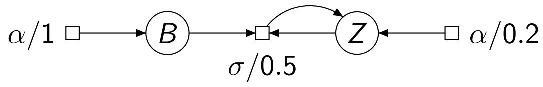
A wta [12, 13, 21, 22] is a finite-state device which specifies a tree series. The core of a wta is its transition mapping. Imagine a situation where the automaton processes some node $w$ of some input tree $\xi$. Then a transition of the automaton is described by the states at the direct descendants of $w$, the label at $w$, and the state at $w$ itself. The transition mapping assigns a weight to every possible transition. The transition weights are used later on to define the weight of $\xi$. Intuitively, assigning a weight of 0 means that a transition is not possible.

A triple $M = (Q, \mu, \nu)$ is a *wta (over $\Sigma$ and $A$)* if $Q$ is a nonempty finite set (of *states*), $\mu$ is a family $(\mu_k \mid k \in \mathbb{N})$ (of *transition mappings*), $\mu_k : \Sigma^{(k)} \to A^{Q^k \times Q}$ for every $k \in \mathbb{N}$, and $\nu \in A^Q$ (*final weights*). Let $M = (Q, \mu, \nu)$ be a wta. We say that $M$ is *bottom-up deterministic* (for short: *bu-det*) if for every $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, and $(q_1, \ldots, q_k) \in Q^k$, there is at most one $q \in Q$ such that $\mu_k(\sigma)_{q_1 \ldots q_k, q} \neq 0$.

*Example 2.* Let $A$ be the Viterbi semiring $([0, 1], \max, \cdot, 0, 1)$, $\Sigma = \{\sigma^{(2)}, \alpha^{(0)}\}$, and $M = (Q, \mu, \nu)$ the wta over $\Sigma$ and $A$ where

- $Q = \{Z, B\}$,
- $\mu_0(\alpha)_{\varepsilon, B} = 1$, $\mu_0(\alpha)_{\varepsilon, Z} = 0.2$, $\mu_2(\sigma)_{BZ, Z} = 0.5$, and $\mu_2(\sigma)_{w, q} = 0$ for every $(w, q) \in (Q^2 \times Q) \setminus \{(BZ, Z)\}$,
- $\nu_Z = 1$, $\nu_B = 0$.

The mapping $\mu$ can be visualized by a hypergraph as shown in Fig. 2(a), which omits any transitions with weight 0. Another equivalent representation of $\mu$ is given by the rules and rule weights of a *weighted regular tree grammar* as in Fig. 2(b). Note that $M$ is not bu-det because $\mu_0(\alpha)_{\varepsilon, q} \neq 0$ holds for every $q \in \{B, Z\}$. □




(a)

$Z \to \sigma(B, Z)$ 0.5
$Z \to \alpha$ 0.2
$B \to \alpha$ 1

(b)

**Fig. 2.** (a) Visualization of a transition mapping using a hypergraph. (b) The same transition mapping shown as a weighted regular tree grammar.

Now we define two different semantics for wta. We begin with the *initial algebra semantics*, which provides the framework for both our determinization construction and the proof of its correctness. In this approach, in order to compute the weight for a given input tree $\xi = \sigma(\xi_1, \ldots, \xi_k)$, the wta recursively computes for each state a separate weight of $\xi$. Let us denote the weight of $\xi$ in state $q$ by $h_\mu(\xi)_q$. Then $h_\mu(\xi)$ is a $Q$-vector over $A$, i.e., $h_\mu(\xi) \in A^Q$. The vector $h_\mu(\xi)$ is obtained by combining the weights from the vectors $h_\mu(\xi_1), \ldots, h_\mu(\xi_k)$ in a specific way, using the transition weights from $\mu$. Before we define $h_\mu$ formally, we show an auxiliary definition which is an extension of $\mu$ and which captures this kind of combination.

Let $M = (Q, \mu, \nu)$. Then $\mu$ induces a family $(\mu_M(\sigma) \mid \sigma \in \Sigma)$ of mappings where for every $k \in \mathbb{N}$ and $\sigma \in \Sigma^{(k)}$ we have $\mu_M(\sigma) : \underbrace{A^Q \times \cdots \times A^Q}_{k} \to A^Q$ and for every $u_1, \ldots, u_k \in A^Q$

$$\mu_M(\sigma)(u_1, \ldots, u_k)_q = \sum_{(q_1,\ldots,q_k) \in Q^k} (u_1)_{q_1} \cdot \ldots \cdot (u_k)_{q_k} \cdot \mu_k(\sigma)_{q_1 \ldots q_k, q} \ .$$

Then $h_\mu : T_\Sigma \to A^Q$ is defined recursively by letting for every $\xi = \sigma(\xi_1, \ldots, \xi_k)$: $h_\mu(\xi) = \mu_M(\sigma)(h_\mu(\xi_1), \ldots, h_\mu(\xi_k))$. Finally, the *i-behavior of $M$*, denoted by $[\![M]\!]_i$, is the tree series in $A\langle\!\langle T_\Sigma \rangle\!\rangle$ such that for every $\xi \in T_\Sigma$:

$$([\![M]\!]_i, \xi) = \sum_{q \in Q} h_\mu(\xi)_q \cdot \nu_q \ .$$

In fact, $h_\mu$ is the unique $\Sigma$-homomorphism from the initial $\Sigma$-algebra $T_\Sigma$ to the $\Sigma$-algebra $(A^Q, \mu_M)$; this gives the name to this semantics [23].

*Example 3.* We show the i-behavior of $M$ from Example 2. For notational convenience, we will write the elements of $A^Q$ as column vectors, where the first row is the $B$-component. Then by elementary computation

$$h_\mu(\alpha) = \begin{pmatrix} 1 \\ 0.2 \end{pmatrix} \ , \qquad h_\mu(\sigma(\alpha, \alpha)) = \begin{pmatrix} 0 \\ h_\mu(\alpha)_B \cdot h_\mu(\alpha)_Z \cdot 0.5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.1 \end{pmatrix} \ .$$

Now we form a general hypothesis. To this end, we define for every $n \in \mathbb{N}$ the tree $\rho_n \in T_\Sigma$ by letting $\rho_0 = \alpha$ and $\rho_{n+1} = \sigma(\alpha, \rho_n)$ for every $n \in \mathbb{N}$. It is easy to prove by induction on $\xi$ that

$$h_\mu(\xi)_Z = \begin{cases} 0.2 \cdot 0.5^n & \text{if } \xi = \rho_n \text{ for some } n \in \mathbb{N} \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

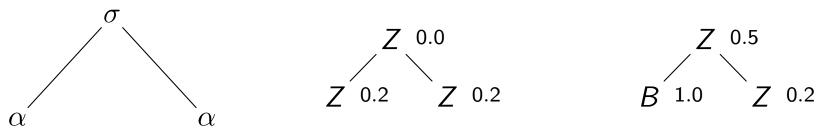



**Fig. 3.** Visualization of an input tree $\xi$ and runs $\kappa_1$ and $\kappa_2$ from Example 4.

By the nature of $\nu$, it is easy to see that $([\![M]\!]_i, \xi) = h_\mu(\xi)_Z$. □

In the *run semantics*, which provides the framework for our termination result, every node of the input tree is decorated with a state. Using the information from $\mu$, such a decoration (called a *run*) is given a weight. The weight of the input tree is then the sum of the weights of all possible runs. Preparing for our result, we consider trees with zero or more occurrences of the nullary symbol $z$.

Let $\xi \in T_\Sigma(\{z\})$ and $p, q \in Q$. We define the set $R_M^{p,q}(\xi)$ of *runs of $M$ on $\xi$ beginning at occurrences of $z$ in state $p$ and ending at the root in state $q$* by letting $R_M^{p,q}(\xi) = \{\kappa : pos(\xi) \to Q \mid \kappa(\varepsilon) = q, \forall w \in pos(\xi) \colon \xi(w) = z$ implies $\kappa(w) = p\}$. We set $R_M^q(\xi) = \bigcup_{p \in Q} R_M^{p,q}(\xi)$ and $R_M(\xi) = \bigcup_{q \in Q} R_M^q(\xi)$.

Moreover, we define the mapping $wt_{M,\xi} : R_M(\xi) \to A$ by letting for every $\kappa \in R_M(\xi)$: $wt_{M,\xi}(\kappa) = \prod_{w \in pos(\xi)} wt_{M,\xi}(\kappa, w)$ where $wt_{M,\xi}(\kappa, w) = 1$ if $\xi(w) = z$ and $wt_{M,\xi}(\kappa, w) = \mu_k(\sigma)_{\kappa(w1)\ldots\kappa(wk),\kappa(w)}$ if $\xi|_w = \sigma(\xi_1, \ldots, \xi_k)$. Furthermore, we will use the following abbreviation: for every finite subset $R \subseteq R_M(\xi)$ we let $wt_{M,\xi}(R) = \sum_{\kappa \in R} wt_{M,\xi}(\kappa)$. In the sequel, we will omit the subscript $\xi$ if clear from the context.

The *r-behavior of $M$*, denoted by $[\![M]\!]_r$, is the tree series in $A\langle\!\langle T_\Sigma \rangle\!\rangle$ such that for every $\xi \in T_\Sigma$:

$$([\![M]\!]_r, \xi) = \sum_{q \in Q, \kappa \in R_M^q(\xi)} wt_M(\kappa) \cdot \nu_q \ .$$

*Example 4.* We reconsider the wta $M$ from Example 2. Fig. 3 depicts the tree $\xi = \sigma(\alpha, \alpha)$ and two runs $\kappa_1, \kappa_2 \in R_M^Z(\xi)$, augmented by the transition weights at each position. It is easy to see that $wt_M(\kappa_1) = 0$ and $wt_M(\kappa_2) = 0.1$. □

In [22, Sect. 3.2], it has been shown that $[\![M]\!]_i = [\![M]\!]_r$. Hence, we drop the subscript from these notations. We say that two wta $M$ and $M'$ over $\Sigma$ and $A$ are *equivalent* if $[\![M]\!] = [\![M']\!]$.

**Lemma 5.** *Let $(Q, \mu, \nu)$ be a bu-det wta and $\xi \in T_\Sigma$. There is at most one $q \in Q$ such that $\mathrm{h}_\mu(\xi)_q \neq 0$, and there is at most one $\kappa \in R_M(\xi)$ such that $\mathrm{wt}_M(\kappa) \neq 0$.*

Throughout this paper, we assume for every wta $M = (Q, \mu, \nu)$ that there is an $\alpha \in \Sigma^{(0)}$ such that $\mu_M(\alpha)() \neq \tilde{0}$. We can do this without loss of generality because otherwise, $[\![M]\!] = \tilde{0}$, and this tree series can also be computed by a wta which has the aforementioned restriction.

*In the sequel, let $M = (Q, \mu, \nu)$ be an arbitrary wta over $\Sigma$ and $A$ such that $\mu_M(\alpha)() \neq \tilde{0}$ for some $\alpha \in \Sigma^{(0)}$.*

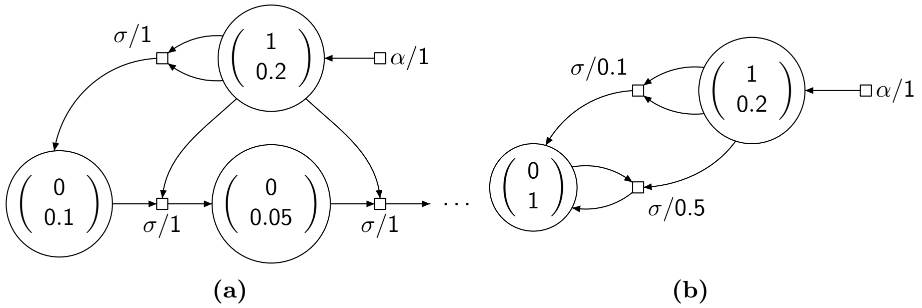



**Fig. 4.** Determinization of the wta $M$ from Example 2 via (a) the approach of Borchardt [20] and (b) the factorization approach.

## 4 Factorizations and Determinization

In the unweighted case, determinization is accomplished using the well-known powerset construction. This construction can be generalized to the weighted case in a straightforward manner by keeping the weights in the new states, however at the price that the new state set might be infinite [20].

*Example 6.* We apply the technique of [20] to the wta $M$ from Example 2. The new state set is obtained as follows:

$$\begin{aligned} Q' &= \{h_\mu(\xi) \mid \xi \in T_\Sigma\} \\ &= \{\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1\\0.2\end{pmatrix}, \begin{pmatrix}0\\0.1\end{pmatrix}, \begin{pmatrix}0\\0.05\end{pmatrix}, \begin{pmatrix}0\\0.025\end{pmatrix}, \ldots\} . \end{aligned}$$

The first element corresponds to trees $\xi$ with no $i$ such that $\xi = \rho_i$, the second to $\rho_0$, the third to $\rho_1$, and so on (cf. Example 3). Part of the resulting transition mapping of the construction is shown in Fig. 4(a). Notice how the determinized automaton mimics the behavior of $\mu_M$ using its states. □

By adapting the factorization technique of [19] from wfsa to wta, we are able to improve the previous approach. Instead of moving the complete computation of weights from the transition mapping into the new states, we factor the elements of $A^Q$ in such a way that the transition mapping in the new automaton is equipped with the factor common to all components. For instance, applying this technique to the wta of Fig. 2 yields the wta in Fig. 4(b). Now we develop this approach in more detail.

Let $Q$ be a finite set. A pair $(f, g)$ is a *factorization (of dimension $Q$)* if $f : A^Q \setminus \{\tilde{0}\} \to A^Q$, $g : A^Q \setminus \{\tilde{0}\} \to A$, and $u = g(u) \cdot f(u)$ for every $u \in A^Q \setminus \{\tilde{0}\}$. A factorization $(f, g)$ is called *maximal* if for every $u \in A^Q$ and $a \in A$, we have that $a \cdot u \neq \tilde{0}$ implies $f(u) = f(a \cdot u)$. Note that if we had permitted $a \cdot u = \tilde{0}$ in the definition of a maximal factorization, and if $f(a \cdot u)$ were defined, we would obtain that $f(u) = f(0 \cdot u) = f(\tilde{0})$ for every $u \in A^Q$. Also note that, for every finite set $Q$, we have the uniquely defined *trivial* factorization $(f, g)$ where $f(u) = u$ and $g(u) = 1$ for every $u \in A^Q \setminus \{\tilde{0}\}$.

*Example 7.* Let $Q$ be a nonempty finite set and $(f, g)$ a factorization of dimension $Q$. We show some cases concerning $A$ and $(f, g)$ where $(f, g)$ is maximal:

1. if $A$ is the semiring of nonnegative reals $(\mathbb{R}^{\geq 0}, +, \cdot, 0, 1)$, $g(u) = \sum_{q \in Q} u_q$, and $f(u) = \frac{1}{g(u)} \cdot u$;
2. if $A$ is the semiring $(\mathbb{R}^{\geq 0}, \max, \cdot, 0, 1)$, $g(u) = \sum_{q \in Q} u_q$, $f(u) = \frac{1}{g(u)} \cdot u$; or alternatively, $g(u) = \max\{u_q \mid q \in Q\}$;
3. if $A$ is the Viterbi semiring $([0, 1], \max, \cdot, 0, 1)$, $g(u) = \max\{u_q \mid q \in Q\}$, and $f(u) = \frac{1}{g(u)} \cdot u$;
4. if $A$ is the tropical semiring $(\mathbb{R}^{\geq 0}_{\infty}, \min, +, \infty, 0)$, $g(u) = \min\{u_q \mid q \in Q\}$, and $f(u) = -g(u) + u$.

Note that division by zero does not occur because $u \neq \tilde{0}$. May and Knight [8] deal with Case (1) and nonrecursive wta. Assuming a wfsa with the twins property, Mohri [18] deals with Case (4), and Kirsten and Mäurer [19] can handle Cases (2)–(4), in which $A$ is extremal (for details see Sect. 5). □

We now adapt the determinization by factorization method of [19, Sect. 3.3] from wfsa to wta. Let $(f, g)$ be a factorization of dimension $Q$. The *determinization of $M$ by $(f, g)$* is the triple $det_{(f,g)}(M) = (Q', \mu', \nu')$ where

– $Q'$ is the smallest set $P \subseteq A^Q$ such that for every $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, $u_1, \ldots, u_k \in P$: if $\mu_M(\sigma)(u_1, \ldots, u_k) \neq \tilde{0}$, then $f(\mu_M(\sigma)(u_1, \ldots, u_k)) \in P$.

– $$\mu'_k(\sigma)_{u_1 \ldots u_k, u} = \begin{cases} g(\mu_M(\sigma)(u_1, \ldots, u_k)) & \text{if } \mu_M(\sigma)(u_1, \ldots, u_k) \neq \tilde{0} \text{ and} \\ & \quad u = f(\mu_M(\sigma)(u_1, \ldots, u_k)) \ , \\ 0 & \text{otherwise ,} \end{cases}$$

  for every $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, $u_1, \ldots, u_k \in Q'$, and $u \in Q'$,

– $\nu'_u = \sum_{q \in Q} u_q \cdot \nu_q$ for every $u \in Q'$.

Note that $Q'$ is uniquely determined because it is chosen from a set which is closed under intersection. Moreover, $Q'$ is not empty because of our assumption that $\mu_M(\alpha)() \neq \tilde{0}$ for some $\alpha \in \Sigma^{(0)}$. Hence, $det_{(f,g)}(M)$ is a wta iff $Q'$ is finite. It is easy to see that, if $det_{(f,g)}(M)$ is a wta, then it is bu-det.

The following observation, which can be proved using the fixpoint theorem of Tarski and Kleene [24, Sect. 1.5.2], shows a stratification of $Q'$; this basically gives an algorithm for computing $Q'$ (in case it is finite).

**Observation 8** *Let $\det_{(f,g)}(M) = (Q', \mu', \nu')$. Then there is a family $(Q'_i \mid i \in \mathbb{N})$ such that $Q'_0 = \emptyset$, for every $i \in \mathbb{N}$ we have $Q'_{i+1} = \{f(\mu_M(\sigma)(u_1, \ldots, u_k)) \mid k \in \mathbb{N}, \sigma \in \Sigma^{(k)}, u_1, \ldots, u_k \in Q'_i, \mu_M(\sigma)(u_1, \ldots, u_k) \neq \tilde{0}\}$ and $Q' = \bigcup_{i \in \mathbb{N}} Q'_i$; moreover, $Q'$ is finite iff there is an $n \in \mathbb{N}$ with $Q' = Q'_n$.*

*Example 9.* We show how $det_{(f,g)}(M) = (Q', \mu', \nu')$ is computed for the wta $M$ from Example 2 and the maximal factorization $(f, g)$ given for the Viterbi semiring in Example 7. First, we compute $Q'$ using Observation 8. We use the following abbreviations:

$$u_1 = \begin{pmatrix} 1 \\ 0.2 \end{pmatrix} \qquad \text{and} \qquad u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ .$$

Then $Q'_0 = \emptyset$ and

$$\begin{aligned} Q'_1 &= \{f(\mu_M(\alpha)())\} = \{f(u_1)\} = \{u_1\} \ , \\ Q'_2 &= \{f(\mu_M(\alpha)()), f(\mu_M(\sigma)(u_1, u_1))\} = Q'_1 \cup \{f(0.1 \cdot u_2)\} = \{u_1, u_2\} \ , \\ Q'_3 &= Q'_2 \cup \{f(\mu_M(\sigma)(u_1, u_2))\} = Q'_2 \cup \{f(0.5 \cdot u_2)\} = \{u_1, u_2\} \ . \qquad (\star) \end{aligned}$$

For $(\star)$, we note that $\mu_M(\sigma)(u_2, u_1) = \tilde{0} = \mu_M(\sigma)(u_2, u_2)$.

Clearly, we have $Q'_i = Q'_2$ for every $i \in \mathbb{N}$, $i \geq 2$. Hence, $Q' = Q_2$. Figure 4(b) shows $\mu'$. In particular, $g(\mu_M(\sigma)(u_1, u_1)) = g(0.1 \cdot u_2) = 0.1$ and hence $\mu'_2(\sigma)_{u_1 u_1, u_2} = 0.1$. Finally, $\nu'_{u_1} = 0.2$ and $\nu'_{u_2} = 1$. □

The following theorem, which corresponds to Theorem 3 of [19], shows that maximal factorizations generate the smallest bu-det wta among all bu-det wta which are obtained by factorization. Note that this does not mean that there is no smaller equivalent bu-det wta.

**Theorem 10.** *Let $A$ be zero-divisor free and let $(f, g)$ and $(\tilde{f}, \tilde{g})$ be factorizations of dimension $Q$ such that $(f, g)$ is maximal. Moreover, let $\det_{(f,g)}(M) = (Q', \mu', \nu')$ and $\det_{(\tilde{f},\tilde{g})}(M) = (\tilde{Q}, \tilde{\mu}, \tilde{\nu})$. Then $Q' = f(\tilde{Q})$; hence $|Q'| \leq |\tilde{Q}|$, and if $\det_{(\tilde{f},\tilde{g})}(M)$ is a wta, then so is $\det_{(f,g)}(M)$.*

## 5 Correctness and Termination

In this section, we show that our determinization construction is correct; more precisely: if $det_{(f,g)}(M)$ is a wta, then $det_{(f,g)}(M)$ is equivalent to $M$. Moreover, we give two sufficient conditions under which $det_{(f,g)}(M)$ is a wta. We begin with correctness. The following theorem corresponds to Theorem 1 of [19].

**Theorem 11.** *If $\det_{(f,g)}(M)$ is a wta, then $[\![M]\!] = [\![det_{(f,g)}(M)]\!]$.*

*Proof.* Let $det_{(f,g)}(M) = (Q', \mu', \nu')$ be a wta, i.e., $Q'$ is finite. We abbreviate $det_{(f,g)}(M)$ by $M'$. Let $\xi \in T_\Sigma$. Since $M'$ is bu-det., Lemma 5 yields that $h_{\mu'}(\xi)_u \neq 0$ for at most one $u \in Q'$. We conjecture that (i) if there is no such $u$, then $h_\mu(\xi) = \tilde{0}$ (intuitively, both automata are stuck) and (ii) if there is such a $u$, then $h_{\mu'}(\xi)_u \cdot u = h_\mu(\xi)$ (intuitively, $M'$ arrived at $u$, and we reverse the factorization applied in the construction). A proof by induction, using our conjecture as hypothesis, yields that the conjecture is correct. We now show that $([\![M']\!], \xi) = ([\![M]\!], \xi)$. It is easy to see that in Case (i) both sides evaluate to 0. For Case (ii), we calculate $([\![M']\!], \xi) = \sum_{u \in Q'} h_{\mu'}(\xi)_u \cdot \nu'_u = \sum_{u \in Q'} h_{\mu'}(\xi)_u \cdot \sum_{q \in Q} u_q \cdot \nu_q = \sum_{u \in Q'} \sum_{q \in Q} h_{\mu'}(\xi)_u \cdot u_q \cdot \nu_q = \sum_{q \in Q} h_\mu(\xi)_q \cdot \nu_q = ([\![M]\!], \xi)$. □

Before we proceed with the sufficient conditions under which $det_{(f,g)}(M)$ is a wta, we show another property. As can be seen from our construction, the state set $Q'$ of $det_{(f,g)}(M)$ is obtained by alternatingly applying $\mu_M(\sigma)$ and $f$. This alternation can be avoided if $(f, g)$ is a maximal factorization. In that case, one can first apply $h_\mu$, which evaluates operations of the form $\mu_M(\sigma)$, and then apply $f$ only once. This is expressed in the following lemma, which corresponds to Lemma 2 of [19].

**Lemma 12.** *Let* $(f,g)$ *be a maximal factorization, and* $\det_{(f,g)}(M) = (Q', \mu', \nu')$. *Then* $Q' \subseteq f(h_\mu(T_\Sigma) \setminus \{\tilde{0}\})$.

Now we show two sufficient conditions for $det_{(f,g)}(M)$ to be a wta (cf. Theorem 15). The first condition is adapted from [19, Theorem 5], that is, we require $A$ to be *extremal*[2] [25] and $M$ to have the *twins property* [26]; the second condition requires $h_\mu(T_\Sigma)$ to be finite, in particular this holds if $M$ is nonrecursive. Now we develop the ingredients.

The semiring $A$ is called *extremal* if $a+b \in \{a,b\}$ for every $a,b \in A$. We note that the semirings listed in Example 7 are extremal, except for the first one.

To define the twins property we need the concept of a context and the concept of substition into a context. A *$\Sigma$-context* is a tree in $T_\Sigma(\{z\})$ in which $z$ occurs exactly once. The set of all $\Sigma$-contexts is denoted by $C_\Sigma$. Given a context $\zeta$ and a tree $\xi$, we denote by $\zeta \cdot \xi$ the tree which is obtained from $\zeta$ by replacing the occurrence of $z$ by $\xi$.

A wta $M = (Q, \mu, \nu)$ is said to have the *twins property* if for every $p, q \in Q$, $\xi \in T_\Sigma$, and $\zeta \in C_\Sigma$, we have that if $wt_M(R_M^p(\zeta\cdot\xi)) \neq 0$ and $wt_M(R_M^q(\zeta\cdot\xi)) \neq 0$, then $wt_M(R_M^{p,p}(\zeta)) = wt_M(R_M^{q,q}(\zeta))$.

*Example 13.* We show that the wta $M$ from Example 2 has the twins property. Let $p, q \in Q$, $\xi \in T_\Sigma$, and $\zeta \in C_\Sigma$ such that $wt_M(R_M^p(\zeta\cdot\xi)) \neq 0$ and $wt_M(R_M^q(\zeta\cdot\xi)) \neq 0$. We show that $wt_M(R_M^{p,p}(\zeta)) = wt_M(R_M^{q,q}(\zeta))$. If $p = q$, this is trivial. For reasons of symmetry, it suffices to consider the case that $p = B$ and $q = Z$. Since $wt_M(R_M^B(\zeta \cdot \xi)) \neq 0$, we conclude that $\zeta = z$ and $\xi = \alpha$. Thus we obtain $wt_M(R_M^{B,B}(z)) = 1 = wt_M(R_M^{Z,Z}(z))$. □

We note that the twins property is undecidable. This follows from the facts that equality of recognizable series over the Tropical semiring is undecidable [27], and that the latter problem can be reduced to the former.

**Lemma 14.** *If* $A$ *is a commutative, extremal semiring and* $M$ *has the twins property, then there is a finite set* $P \subseteq A^Q$ *with* $h_\mu(T_\Sigma) \subseteq \{a \cdot u \mid a \in A, u \in P\}$.

*Proof.* We only sketch the proof idea. Let $\xi_1 \in T_\Sigma$. Since $A$ is extremal, for every $q \in Q$ there is a $\kappa_q \in R_M^q(\xi_1)$ such that $h_\mu(\xi_1)_q = wt(\kappa_q)$. If $\xi_1$ is sufficiently "large", then we find positions $w_1$ and $w_2$ such that $\kappa_q(w_1) = \kappa_q(w_2)$ for every $q \in Q$. Provided that we have chosen the runs $\kappa_q$ in a suitable manner, the twins property guarantees that the "slices" of $\kappa_q$ starting at position $w_1$ and ending at position $w_2$ have the same weight, say $a_1$, for every $q \in Q$. We can remove this slice from $\xi_1$, obtaining the smaller tree $\xi_2$ and runs $\kappa'_q$ on $\xi_2$ with $wt_M(\kappa_q) = a_1 \cdot wt_M(\kappa'_q)$. This can be iterated a finite number of times, yielding the trees $\xi_1, \ldots, \xi_n$ and weights $a_1, \ldots, a_{n-1}$, where $\xi_n$ is in a finite set of "small" trees (giving rise to a finite set $P$ of vectors). □

Part (i) of the following theorem generalizes Theorem 5 of [19] from wfsa to wta. Part (ii) provides a formal verification of [8].

[2] called *min-semiring* in [19]

**Theorem 15.** *Let $(f, g)$ be a maximal factorization. Then* $\det_{(f,g)}(M)$ *is a wta if (i) $A$ is extremal and $M$ has the twins property, or (ii) $h_\mu(T_\Sigma)$ is finite.*

*Proof.* Let $det_{(f,g)}(M) = (Q', \mu', \nu')$. By Lemma 12, $Q' \subseteq f(h_\mu(T_\Sigma) \setminus \{0\}^Q)$. Hence, if $h_\mu(T_\Sigma)$ is finite, so is $Q'$, and $det_{(f,g)}(M)$ is a wta. If $A$ is extremal and $M$ has the twins property, Lemma 14 yields that there is a finite set $P \subseteq A^Q$ such that $h_\mu(T_\Sigma) \subseteq \{a \cdot u \mid a \in A, u \in P\}$. We calculate $Q' \subseteq f(h_\mu(T_\Sigma) \setminus \{0\}^Q) \subseteq f(\{a \cdot u \mid a \in A, u \in P\} \setminus \{0\}^Q) \subseteq f(P \setminus \{0\}^Q)$ because $(f, g)$ is maximal. Hence, $Q'$ is finite, and $det_{(f,g)}(M)$ is a wta. □

## 6 Conclusion and Open Problems

We have generalized the determinization by factorization of [19] from wfsa to wta. We have proved that the construction is correct and shown sufficient conditions under which it results in a finite automaton. We have indicated how our results transfer to the algorithm of [8] for nonrecursive automata.

Further research is needed for the case of recursive automata. Allauzen and Mohri [28] show that the twins property for wfsa is decidable for a decidable subclass called cycle-unambiguous. In addition, there are applications where determinization for wfsa has polynomial complexity [18]. It would be interesting to see how these results transfer to our setting.

**Acknowledgements** The authors are grateful to the referees for critical, but valuable remarks, and to Andreas Maletti for fruitful discussions.

## References

1. Bod, R.: A computational model of language performance: data oriented parsing. In: Proc. COLING, Nantes, France (1992) 855–859
2. Galley, M., Hopkins, M., Knight, K., Marcu, D.: What's in a translation rule? In: Proc. HLT-NAACL, Boston, ACL (2004) 273–280
3. Langkilde, I., Knight, K.: The practical value of n-grams in generation. In: Proc. of the Ninth International Workshop on Natural Language Generation, Niagra-on-the-lake, Ontario, Canada (1998) 248–255
4. Liang, P., Bouchard-Côté, A., Klein, D., Taskar, B.: An end-to-end discriminative approach to machine translation. In: Proc. ACL, Sydney, Australia, Association for Computational Linguistics (July 2006) 761–768
5. Tillmann, C., Zhang, T.: A block bigram prediction model for statistical machine translation. ACM Trans. Speech Lang. Process. **4**(3) (2007) 1–31
6. Collins, M.: Three generative, lexicalised models for statistical parsing. In: Proc. ACL, Madrid, ACL (1997) 16–23
7. Blunsom, P., Cohn, T., Osborne, M.: A discriminative latent variable model for statistical machine translation. In: Proc. HLT, Columbus, Ohio, ACL (June 2008) 200–208
8. May, J., Knight, K.: A better N-best list: practical determinization of weighted finite tree automata. In: Proc. NAACL-HLT. (2006)

9. DeNeefe, S., Knight, K., Wang, W., Marcu, D.: What can syntax-based MT learn from phrase-based MT? In: Proc. EMNLP-CoNLL, Prague, ACL (2007) 755–763
10. Fraser, A., Marcu, D.: Getting the structure right for word alignment: LEAF. In: Proc. EMNLP-CoNLL, Prague, ACL (2007) 51–60
11. Charniak, E.: Immediate-head parsing for language models. In: Proc. ACL, Toulouse, France, ACL (2001) 116–123
12. Berstel, J., Reutenauer, C.: Recognizable formal power series on trees. Theoret. Comput. Sci. **18**(2) (1982) 115–148
13. Alexandrakis, A., Bozapalidis, S.: Weighted grammars and Kleene's theorem. Inform. Process. Lett. **24**(1) (1987) 1–4
14. Billot, S., Lang, B.: The structure of shared forests in ambiguous parsing. In: Proc. ACL, Vancouver, ACL (1989) 143–151
15. Brainerd, W.S.: Tree generating regular systems. Information and Control **14** (1969) 217–231
16. Eppstein, D.: Finding the k shortest paths. SIAM Journal on Computing **28**(2) (1998) 652–673
17. Huang, L., Chiang, D.: Better k-best parsing. In: Proc. IWPT, Vancouver, ACL (2005) 53–64
18. Mohri, M.: Finite-state transducers in language and speech processing. Computational Linguistics **23**(2) (1997) 269–311
19. Kirsten, D., Mäurer, I.: On the determinization of weighted automata. J. Autom. Lang. Comb. **10** (2005) 287–312
20. Borchardt, B.: A pumping lemma and decidability problems for recognizable tree series. Acta Cybernet. **16**(4) (2004) 509–544
21. Ésik, Z., Kuich, W.: Formal tree series. J. Autom. Lang. Comb. **8**(2) (2003) 219–285
22. Fülöp, Z., Vogler, H.: Weighted tree automata and tree transducers. In Droste, M., Kuich, W., Vogler, H., eds.: Handbook of Weighted Automata. Springer (2009) to appear.
23. Goguen, J., Thatcher, J., Wagner, E., Wright, J.: Initial algebra semantics and continuous algebras. J. ACM **24** (1977) 68–95
24. Wechler, W.: Universal Algebra for Computer Scientists. Volume 25 of Monogr. Theoret. Comput. Sci. EATCS Ser. Springer (1992)
25. Mahr, B.: Iteration and summability in semirings. Annals of Discrete Mathematics **19** (1984) 229–256
26. Choffrut, C.: Une caracterisation des fonctions sequentielles et des fonctions sous-sequentielles en tant que relations rationnelles. Theoret. Comput. Science **5** (1977) 325–337
27. Krob, D.: The equality problem for rational series with multiplicities in the tropical semiring is undecidable. Intern. J. of Inform. and Comput. **4** (1994) 405–425
28. Allauzen, C., Mohri, M.: Efficient Algorithms for Testing the Twins Property. J. of Automata, Languages and Combinatorics **8**(2) (2003) 117–144

# A Formal proofs

We will often use the following observation, which can be shown by elementary computations.

**Observation 16** *For every $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, $u_1, \ldots, u_k \in A^Q$, and $a_1, \ldots, a_k \in A$, we have $\mu_M(\sigma)(a_1 \cdot u_1, \ldots, a_k \cdot u_k) = a_1 \cdot \ldots \cdot a_k \cdot \mu_M(\sigma)(u_1, \ldots, u_k)$.*

## A.1 Proof of Theorem 10

For the proof of Theorem 10 we need the following lemma.

**Lemma 17.** *Let $(f, g)$ be a maximal factorization. Furthermore, let $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, $u_1, \ldots, u_k \in A^Q$, and $u'_1, \ldots, u'_k \in A^Q$ such that $u'_i \in \{u_i, f(u_i)\}$ for every $i \in \{1, \ldots, k\}$. Then $\mu_M(\sigma)(u_1, \ldots, u_k) \neq \tilde{0}$ implies $\mu_M(\sigma)(u'_1, \ldots, u'_k) \neq \tilde{0}$, and the converse holds if $A$ zero-divisor free. Furthermore, if $\mu_M(\sigma)(u_1, \ldots, u_k) \neq \tilde{0}$, then $f(\mu_M(\sigma)(u_1, \ldots, u_k)) = f(\mu_M(\sigma)(u'_1, \ldots, u'_k))$.*

*Proof.* We construct the sequence $a_1, \ldots, a_k \in A$ by letting for every $i \in \{1, \ldots, k\}$

$$a_i = \begin{cases} g(u_i) & \text{if } u'_i = f(u_i)\ , \\ 1 & \text{otherwise}\ . \end{cases}$$

Clearly then $(\star)$

$$\begin{aligned} \mu_M(\sigma)(u_1, \ldots, u_k) &= \mu_M(\sigma)(a_1 \cdot u'_1, \ldots, a_k \cdot u'_k) && ((f,g) \text{ factorization}) \\ &= a_1 \cdot \ldots \cdot a_k \cdot \mu_M(\sigma)(u'_1, \ldots, u'_k)\ . && (\text{Obs. } 16) \end{aligned}$$

First, let $\mu_M(\sigma)(u_1, \ldots, u_k) \neq \tilde{0}$. By $(\star)$ also $\mu_M(\sigma)(u'_1, \ldots, u'_k) \neq \tilde{0}$. Applying $f$ to $(\star)$ and using that $(f, g)$ is maximal, we obtain that $f(\mu_M(\sigma)(u_1, \ldots, u_k)) = f(\mu_M(\sigma)(u'_1, \ldots, u'_k))$. Second, assume that $\mu_M(\sigma)(u'_1, \ldots, u'_k) \neq \tilde{0}$. Since $a_i \neq 0$ for every $i \in \{1, \ldots, k\}$, and since $A$ is zero-divisor free, $(\star)$ yields that $\mu_M(\sigma)(u_1, \ldots, u_k) \neq \tilde{0}$. ☐

The following proof uses the stratification of $Q'$ from Observation 8.

*Proof (of Theorem 10).* Note that for every $u \in A^Q \setminus \{\tilde{0}\}$, we have $\tilde{g}(u) \cdot \tilde{f}(u) = u = g(u) \cdot f(u)$, and by applying $f$ we obtain $f(\tilde{f}(u)) = f(u) = f(f(u))$ because $(f, g)$ is maximal.

We begin with the proof of $f(\tilde{Q}) \subseteq Q'$. We show the following statement by induction on $i$: for every $i \in \mathbb{N}$, $f(\tilde{Q}_i) \subseteq Q'$. To this end, let $i \in \mathbb{N}$ and $\tilde{u} \in f(\tilde{Q}_{i+1})$. Then there are $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, and $\tilde{u}_1, \ldots, \tilde{u}_k \in \tilde{Q}_i$ such that $\tilde{u} = f(\tilde{f}(\mu_M(\sigma)(\tilde{u}_1, \ldots, \tilde{u_k})))$. Hence

$$\begin{aligned} \tilde{u} &= f(\tilde{f}(\mu_M(\sigma)(\tilde{u}_1, \ldots, \tilde{u_k}))) \\ &= f(\mu_M(\sigma)(\tilde{u}_1, \ldots, \tilde{u_k})) \\ &= f(\mu_M(\sigma)(f(\tilde{u}_1), \ldots, f(\tilde{u_k}))) && (\text{Lemma } 17) \\ &\in Q'\ . && (\text{induction hypothesis, def. of } Q') \end{aligned}$$

Now we prove $Q' \subseteq f(\tilde{Q})$. We show the following statement by induction on $i$: for every $i \in \mathbb{N}$, $Q'_i \subseteq f(\tilde{Q})$. To this end, let $i \in \mathbb{N}$ and $u' \in Q'_{i+1}$. Then there are $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, and $u'_1, \ldots, u'_k \in Q'_i$ such that $u' = f(\mu_M(\sigma)(u'_1, \ldots, u'_k))$. By induction hypothesis, there are $\tilde{u}_1, \ldots, \tilde{u}_k \in \tilde{Q}$ such that $u'_i = f(\tilde{u}_i)$ for every $i \in \{1, \ldots, k\}$. Hence

$$\begin{aligned} u' &= f(\mu_M(\sigma)(f(\tilde{u}_1), \ldots, f(\tilde{u_k}))) \\ &= f(\mu_M(\sigma)(\tilde{u}_1, \ldots, \tilde{u_k})) && \text{(Lemma 17)} \\ &= f(\tilde{f}(\mu_M(\sigma)(\tilde{u}_1, \ldots, \tilde{u_k}))) \\ &\in f(\tilde{Q}) \ . \end{aligned}$$

□

### A.2 Proof of Theorem 11, full version

*Proof.* Let $det_{(f,g)}(M) = (Q', \mu', \nu')$ be a wta, i.e., $Q'$ is finite. Let us abbreviate $det_{(f,g)}(M)$ by $M'$. The claim follows from the following statement ($\star$), which we show by induction on $\xi$: for every $\xi \in T_\Sigma$ there is a $u' \in Q'$ such that for every $u \in Q'$ and $q \in Q$, we have

$$h_{\mu'}(\xi)_u \cdot u_q = \begin{cases} h_\mu(\xi)_q & \text{if } u = u' \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

Note that ($\star$) implies that for every $\xi \in T_\Sigma$ there is at most one $u \in Q'$ such that there is a $q \in Q$ with $h_{\mu'}(\xi)_u \cdot u_q \neq 0$.

Let $\xi = \sigma(\xi_1, \ldots, \xi_k)$. By induction hypothesis, there are $u'_1, \ldots, u'_k \in Q'$ such that for every $i \in \{1, \ldots, k\}$, $u \in Q'$, and $q \in Q$

$$h_{\mu'}(\xi_i)_u \cdot u_q = \begin{cases} h_\mu(\xi_i)_q & \text{if } u = u'_i \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

We construct $u'$ as follows: if $\mu_M(\sigma)(u'_1, \ldots, u'_k) = \tilde{0}$, we fix an arbitrary $u' \in Q'$; otherwise, we set $u' = f(\mu_M(\sigma)(u'_1, \ldots, u'_k))$. Now let $u \in Q'$ and $q \in Q$. We will use the following abbreviation:

$$ind = \{(u_1, \ldots, u_k) \in (Q')^k \mid \mu_M(\sigma)(u_1, \ldots, u_k) \neq \tilde{0} \wedge u = f(\mu_M(\sigma)(u_1, \ldots, u_k))\} \ .$$

Then

$$\begin{aligned} & h_{\mu'}(\xi)_u \cdot u_q \\ &= \mu_{M'}(\sigma)(h_{\mu'}(\xi_1), \ldots, h_{\mu'}(\xi_k))_u \cdot u_q \\ &= \Big(\textstyle\sum_{u_1,\ldots,u_k \in Q'} h_{\mu'}(\xi_1)_{u_1} \cdot \ldots \cdot h_{\mu'}(\xi_k)_{u_k} \cdot \mu'_k(\sigma)_{u_1 \ldots u_k, u}\Big) \cdot u_q && \text{(def. of } \mu_{M'}) \\ &= \textstyle\sum_{u_1,\ldots,u_k \in Q'} h_{\mu'}(\xi_1)_{u_1} \cdot \ldots \cdot h_{\mu'}(\xi_k)_{u_k} \cdot \mu'_k(\sigma)_{u_1 \ldots u_k, u} \cdot u_q && (\dagger) \\ &= \textstyle\sum_{(u_1,\ldots,u_k) \in ind} h_{\mu'}(\xi_1)_{u_1} \cdot \ldots \cdot h_{\mu'}(\xi_k)_{u_k} \cdot g(\mu_M(\sigma)(u_1, \ldots, u_k)) \end{aligned}$$

$$
\begin{aligned}
&\cdot f(\mu_M(\sigma)(u_1, \ldots, u_k))_q && \text{(def. of } \mu'\text{)}\\
&= \textstyle\sum_{(u_1,\ldots,u_k)\in ind} h_{\mu'}(\xi_1)_{u_1} \cdot \ldots \cdot h_{\mu'}(\xi_k)_{u_k} \cdot \mu_M(\sigma)(u_1, \ldots, u_k)_q \\
& && \text{((}f, g\text{) factorization)}\\
&= \textstyle\sum_{(u_1,\ldots,u_k)\in ind} \mu_M(\sigma)(h_{\mu'}(\xi_1)_{u_1} \cdot u_1, \ldots, h_{\mu'}(\xi_k)_{u_k} \cdot u_k)_q && \text{(Obs. 16)}
\end{aligned}
$$

Note that by Lemma 5, the sum on each side of (†) has at most one nonzero summand, which proves (†). Now we continue our calculation. To this end, we distinguish two cases.

*Case 1*: Let $(u'_1, \ldots, u'_k) \in \mathit{ind}$. Then we continue with

$$
\begin{aligned}
\ldots &= \mu_M(\sigma)(h_\mu(\xi_1), \ldots, h_\mu(\xi_k))_q && \text{(induction hypothesis)}\\
&= h_\mu(\xi)_q \;.
\end{aligned}
$$

Note that since $(u'_1, \ldots, u'_k) \in \mathit{ind}$, we have $u = u'$.

*Case 2*: Let $(u'_1, \ldots, u'_k) \notin \mathit{ind}$. Then we continue with

$$
\ldots = 0 \;. \qquad \text{(induction hypothesis)}
$$

Since $(u'_1, \ldots, u'_k) \notin \mathit{ind}$, we have $\mu_M(\sigma)(u'_1, \ldots, u'_k) = \tilde{0}$ or $u \neq u'$. For $u \neq u'$, we have nothing more to show. Hence, let $u = u'$. Then $\mu_M(\sigma)(u'_1, \ldots, u'_k) = \tilde{0}$, and we obtain

$$
\begin{aligned}
h_\mu(\xi)_q &= \mu_M(\sigma)(h_\mu(\xi_1), \ldots, h_\mu(\xi_k))_q \\
&= \mu_M(\sigma)(h_{\mu'}(\xi_1)_{u'_1} \cdot u'_1, \ldots, h_{\mu'}(\xi_k)_{u'_k} \cdot u'_k)_q && \text{(induction hypothesis)}\\
&= h_{\mu'}(\xi_1)_{u'_1} \cdot \ldots \cdot h_{\mu'}(\xi_k)_{u'_k} \cdot \mu_M(\sigma)(u'_1, \ldots, u'_k)_q && \text{(Obs. 16)}\\
&= 0 \;.
\end{aligned}
$$

We have thus proved ($\star$) and turn to the proof of $[\![M]\!]_i = [\![M']\!]_i$. To this end, let $\xi \in T_\Sigma$. Then

$$
\begin{aligned}
([\![M']\!]_i, \xi) &= \textstyle\sum_{u\in Q'} h_{\mu'}(\xi)_u \cdot \nu'_u \\
&= \textstyle\sum_{u\in Q'} h_{\mu'}(\xi)_u \cdot \sum_{q\in Q} u_q \cdot \nu_q && \text{(def. of } \nu'\text{)}\\
&= \textstyle\sum_{u\in Q'} \sum_{q\in Q} h_{\mu'}(\xi)_u \cdot u_q \cdot \nu_q \\
&= \textstyle\sum_{q\in Q} h_\mu(\xi)_q \cdot \nu_q && \text{(by } (\star)\text{)}\\
&= ([\![M]\!]_i, \xi) \;.
\end{aligned}
$$

□

### A.3 Proof of Lemma 12

The following proof uses the stratification of $Q'$ from Observation 8.

*Proof (of Lemma 12).* We show the following statement by induction on $i$: for every $i \in \mathbb{N}$, $Q'_i \subseteq f(h_\mu(T_\Sigma) \setminus \{\tilde{0}\})$. For $i = 0$, this is trivially true. Let $i \in \mathbb{N}$

and $u \in Q'_{i+1}$. Then there are $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, and $u_1, \ldots, u_k \in Q'_i$ such that $u = f(\mu_M(\sigma)(u_1, \ldots, u_k))$. By induction hypothesis, there are $\xi_1, \ldots, \xi_k \in T_\Sigma$ such that $u_i = f(h_\mu(\xi_i))$ for every $i \in \{1, \ldots, k\}$. Hence

$$\begin{aligned} u &= f(\mu_M(\sigma)(u_1, \ldots, u_k)) \\ &= f(\mu_M(\sigma)(f(h_\mu(\xi_1)), \ldots, f(h_\mu(\xi_k)))) \\ &= f(\mu_M(\sigma)(h_\mu(\xi_1), \ldots, h_\mu(\xi_k))) && \text{(Lemma 17)} \\ &= f(h_\mu(\sigma(\xi_1, \ldots, \xi_k))) \\ &\in f(h_\mu(T_\Sigma) \setminus \{\tilde{0}\}) \ . \end{aligned}$$

Note that we did not use the part of Lemma 17 which requires $A$ to be zero-divisor free. ◻

### A.4 Proof of Lemma 14, full version

Before we can prove Lemma 14, we need some preliminaries. Let $\xi \in T_\Sigma(\{z\})$ and $w \in pos(\xi)$. For every $\xi' \in T_\Sigma(\{z\})$ we denote by $\xi[\xi']_w$ the tree obtained from $\xi$ by substituting the subtree rooted at $w$ by $\xi'$. For every $\kappa \in R_M(\xi)$, we extend the notions $\xi|_w$ and $\xi[z]_w$ to $\kappa$ in the obvious way, i.e., for every $w' \in pos(\xi|_w)$, we set $\kappa|_w(w') = \kappa(ww')$ and for every $\xi \in pos(\xi[z]_w)$, we set $\kappa[z]_w(w') = \kappa(w')$. Moreover, for every $q \in Q$, $\xi \in T_\Sigma(\{z\})$, and $\kappa \in R^q_M(\xi)$ we set $R_M(\kappa) = R^{q,q}_M(\xi)$.

**Observation 18** *Let $A$ be an extremal semiring, $\xi \in T_\Sigma$, and $R \subseteq R_M(\xi)$ be finite. Then there is a $\kappa \in R$ such that $wt_M(\kappa) = wt_M(R)$.*

A run $\kappa \in R$ with the property that $wt_M(\kappa) = wt_M(R)$ is called *victorious* in [19].

**Observation 19** *Let $A$ be an extremal semiring. Moreover, let $\xi \in T_\Sigma$, $w \in pos(\xi)$, and $\kappa \in R_M(\xi)$ such that $wt_M(\kappa) = wt_M(R_M(\kappa))$. Then $wt_M(\kappa[z]_w) \cdot wt_M(\kappa|_w) = wt_M(R_M(\kappa[z]_w)) \cdot wt_M(\kappa|_w)$.*

*Proof.* We have

$$\begin{aligned} & wt_M(R_M(\kappa[z]_w)) \cdot wt_M(\kappa|_w) \\ &= \left(\textstyle\sum_{\nu \in R_M(\kappa[z]_w)} wt_M(\nu)\right) \cdot wt_M(\kappa|_w) \\ &= \textstyle\sum_{\nu \in R_M(\kappa[z]_w)} wt_M(\nu) \cdot wt_M(\kappa|_w) && \text{(distributivity)} \\ &= \textstyle\sum_{\nu \in R_M(\kappa)} wt_M(\nu) && (\star) \\ &= wt_M(R_M(\kappa)) = wt_M(\kappa) \\ &= wt_M(\kappa[z]_w) \cdot wt_M(\kappa|_w) \ . \end{aligned}$$

It is easy to see that the index set on the right-hand side of $(\star)$ is a superset of that on the left-hand side, which in turn already contains $wt_M(\kappa)$. ◻

**Definition 20.** *Let $Q' \subseteq Q$. Then we define $\mathcal{C}'_{M,Q'} = \{(\xi, \kappa) \mid \xi \in T_\Sigma, \kappa \in R_M(\xi)^{Q'}, \forall q \in Q' \colon \kappa_q \in R^q_M(\xi), wt_M(\kappa_q) \neq 0\}$. We set $\mathcal{C}'_M = \bigcup_{Q' \subseteq Q} \mathcal{C}'_{M,Q'}$. In the following, we define the mappings $\varphi : \mathcal{C}'_M \to A^Q$, the family $(U(\xi,\kappa) \mid (\xi,\kappa) \in \mathcal{C}_{M',Q'})$, and the set $\mathcal{C}_{M,Q'} \subseteq \mathcal{C}'_{M,Q'}$. To this end, let $(\xi,\kappa) \in \mathcal{C}'_{M,Q'}$. We will often identify this pair with $\kappa$. Then*

- *For every $q \in Q$ we set $\varphi(\kappa)_q = wt_M(\kappa_q)$ if $q \in Q'$, otherwise we set $\varphi(\kappa)_q = 0$.*
- *We define $U(\kappa)$ to be the set of all pairs $(w_1, w_2) \in pos(\xi) \times pos(\xi)$ such that $w_1 < w_2$ and for every $q \in Q'$ we have $\kappa_q(w_1) = \kappa_q(w_2)$.*
- *We have $\kappa \in \mathcal{C}_{M,Q'}$ if for every $(w_1,w_2) \in U(\kappa)$ and $q \in Q'$ we have $wt_M(\kappa_q|_{w_1}) = wt_M(R_M(\kappa_q|_{w_1}))$. We set $\mathcal{C}_M = \bigcup_{Q' \subseteq Q} \mathcal{C}_{M,Q'}$.*

**Lemma 21.** *Let $A$ be an extremal semiring and $\xi \in T_\Sigma$. Then there is a $\kappa$ such that $(\xi,\kappa) \in \mathcal{C}_M$ and $\varphi(\kappa) = h_\mu(\xi)$.*

*Proof.* We begin by showing the following, more general statement by induction on $\xi$: for every $\xi \in T_\Sigma$ and $q \in Q$ there is a $\kappa \in R^q_M(\xi)$ such that the statement $P(\xi,\kappa)$ holds, where $P(\xi,\kappa)$ iff for every $w \in pos(\xi)$ we have $wt_M(\kappa|_w) = wt_M(R_M(\kappa|_w))$.

Let $\xi = \sigma(\xi_1,\ldots,\xi_k)$. By Observation 18 there is a $\kappa' \in R^q_M(\xi)$ such that $wt_M(\kappa') = wt_M(R^q_M(\xi))$. By induction hypothesis, there are $\kappa_1 \in R_M(\kappa|_1)$, …, $\kappa_k \in R_M(\kappa|_k)$ such that for every $i \in \{1,\ldots,k\}$ we have $P(\xi_i,\kappa_i)$. Now we construct $\kappa = \kappa'(\varepsilon)(\kappa_1,\ldots,\kappa_k)$. It is easy to see that $P(\xi,\kappa)$ holds.

Finally, let $Q' = \{q \in Q \mid h_\mu(\xi)_q \neq 0\}$. We construct $\kappa \in R_M(\xi)^{Q'}$ as follows: for every $q \in Q'$ let $\kappa_q \in R^q_M(\xi)$ such that $P(\xi,\kappa)$ holds. By [22, Sect. 3.2], we have $\varphi(\kappa) = h_\mu(\xi)$. ☐

**Lemma 22.** *Let $A$ be a commutative, extremal semiring and $M$ have the twins property. Moreover, let $Q' \subseteq Q$ and $(\xi,\kappa) \in \mathcal{C}_{M,Q'}$ such that $U(\kappa) \neq \emptyset$. Then there are $(\xi',\kappa') \in \mathcal{C}_{M,Q'}$ and $a \in A$ such that $\varphi(\kappa) = a \cdot \varphi(\kappa')$ and $|U(\kappa')| < |U(\kappa)|$.*

*Proof (of Lemma 22).* Since $U(\kappa) \neq \emptyset$, there is a pair $(w_1,w_2) \in U(\kappa)$ such that for every $(w'_1,w'_2) \in U(\kappa)$, if $w'_1 \leq w_1$, then $w'_1 = w_1$. We construct $\xi' = \xi[\xi|_{w_2}]_{w_1}$ and for every $q \in Q'$ and $w \in pos(\xi')$, we set $\kappa'_q(w) = \kappa_q(w_2 v)$ if $w = w_1 v$ and $\kappa'_q(w) = \kappa_q(w)$ otherwise. Finally, if $Q' = \emptyset$, we set $a = 0$, otherwise we choose some $q' \in Q'$ and set $a = wt_M(\kappa_{q'}[z]_{w_2}|_{w_1})$. Now we show that (1) $(\xi',\kappa') \in \mathcal{C}_{M,Q'}$, (2) $\varphi(\kappa) = a \cdot \varphi(\kappa')$, and (3) $|U(\kappa')| < |U(\kappa)|$.

We begin with Statement 1. It is easy to see that $(\xi',\kappa') \in \mathcal{C}'_{M,Q'}$. Now let $(w'_1,w'_2) \in U(\kappa')$ and $q \in Q'$. We show that $wt_M(\kappa'_q|_{w'_1}) = wt_M(R_M(\kappa'_q|_{w'_1}))$. Note that $w'_1 \not< w_1$. We distinguish two cases.

Case 1: There are $v_1, v_2 \in \mathbb{N}^*$ such that $w'_1 = w_1 \cdot v_1$ and $w'_2 = w_1 \cdot v_2$. Since $\kappa'_q|_{w_1} = \kappa_q|_{w_2}$ for every $q \in Q'$, we obtain that $(w_2 v_1, w_2 v_2) \in U(\kappa)$. Hence

$$\begin{aligned} wt_M(\kappa'_q|_{w'_1}) &= wt_M(\kappa_q|_{w_2 v_1}) \\ &= wt_M(R_M(\kappa_q|_{w_2 v_1})) \\ &= wt_M(R_M(\kappa'_q|_{w'_1})) \ . \end{aligned}$$

Case 2: Otherwise, $\kappa'_q|_{w'_1} = \kappa_q|_{w_1}$. Thus $(w'_1, w'_2) \in U(\kappa)$. The rest of the proof is very similar to that of Case 1.

Now we show Statement 2 for the non-trivial case, i.e., $Q' \neq \emptyset$. Let $q \in Q'$. Then

$$\begin{aligned}
& wt_M(\kappa_q) \\
&= wt_M(\kappa_q[z]_{w_1}) \cdot wt_M(\kappa_q[z]_{w_2}|_{w_1}) \cdot wt_M(\kappa_q|_{w_2}) \\
&= wt_M(\kappa_q[z]_{w_1}) \cdot wt_M(R_M(\kappa_q[z]_{w_2}|_{w_1})) \cdot wt_M(\kappa_q|_{w_2}) && \text{(Observation 19)} \\
&= wt_M(\kappa_q[z]_{w_1}) \cdot wt_M(R_M(\kappa_{q'}[z]_{w_2}|_{w_1})) \cdot wt_M(\kappa_q|_{w_2}) && \text{(twins property)} \\
&= wt_M(\kappa_q[z]_{w_1}) \cdot wt_M(\kappa_{q'}[z]_{w_2}|_{w_1}) \cdot wt_M(\kappa_q|_{w_2}) && \text{(Observation 19)} \\
&= a \cdot wt_M(\kappa_q[z]_{w_1}) \cdot wt_M(\kappa_q|_{w_2}) && \text{(comm.)} \\
&= a \cdot wt_M(\kappa'_q) \ .
\end{aligned}$$

Finally, Statement 3 is easy to see. □

**Lemma 23.** *There is a finite set $\mathcal{F}_M \subseteq \mathcal{C}'_M$ such that $U^{-1}(\emptyset) \subseteq \mathcal{F}_M$.*

*Proof.* We construct $\mathcal{F}_M = \{\kappa \in \mathcal{C}'_M \mid \forall w \in pos(\kappa)\colon |w| < |Q|^{|Q|}\}$. Let $\kappa \in \mathcal{C}'_M \setminus \mathcal{F}_M$. We show that $U(\kappa) \neq \emptyset$ follows. First of all, there is a $Q' \subseteq Q$ such that $\kappa \in \mathcal{C}'_{M,Q'}$. Since $\kappa \notin \mathcal{F}_M$, there is a $w \in pos(\kappa)$ such that $|w| \geq |Q|^{|Q'|}$. Hence, there are $k \in \mathbb{N}$, $w_1, \ldots, w_k \in \mathbb{N}^*$, and $u_1, \ldots, u_k \in Q^{Q'}$ such that $k > |Q|^{|Q'|}$, $w_i \in pos(\kappa)$ for every $i \in \{1, \ldots, k\}$, $w_1 < w_2 < \ldots < w_k$, and $\kappa_q|_{w_i} \in R_M^{(u_i)_q}$ for every $i \in \{1, \ldots, k\}$ and $q \in Q'$. By the pigeon-hole principle, there are $i, j \in \{1, \ldots, k\}$ such that $i < j$ and $u_i = u_j$. Then, however, $(w_i, w_j) \in U(\kappa)$. □

*Proof (of Lemma 14).* We choose $P = \varphi(\mathcal{F}_M)$ where $\mathcal{F}_M$ is the set from Lemma 23.

Now let $\xi \in T_\Sigma$. Let $n \in \mathbb{N}$ be maximal such that there are $\kappa_1, \ldots, \kappa_n \in \mathcal{C}_M$ and $a_1, \ldots, a_k \in A$ such that $\varphi(\kappa_1) = h_\mu(\xi)$, $U(\kappa_n) = \emptyset$, $\varphi(\kappa_i) = a_i \cdot \varphi(\kappa_1)$ for every $i \in \{1, \ldots, n\}$, and $|U(\kappa_{i+1})| < |U(\kappa_i)|$ for every $i \in \{1, \ldots, n-1\}$. We claim that (1) $n > 0$ and (2) $\kappa_n \in \mathcal{F}_M$, which allows us to derive

$$\begin{aligned}
h_\mu(\xi) &= \varphi(\kappa_1) \\
&= a_n \cdot \varphi(\kappa_n) \\
&\in \{a \cdot u \mid a \in A, u \in P\} \ .
\end{aligned}$$

Statement 1 follows from Lemma 21 if we set $a_1 = 1$.

Finally, we prove Statement 2. Assume that $U(\kappa_n) \neq \emptyset$. By Lemma 22, there are $\kappa'$ and $a'$ such that $\varphi(\kappa_n) = a' \cdot \varphi(\kappa')$ and $|U(\kappa')| < |U(\kappa)|$. Using $\kappa_{n+1} = \kappa'$ and $a_{n+1} = a' \cdot a_n$, we see that $n$ was not maximal. Hence, $U(\kappa_n) = \emptyset$, and by Lemma 23, $\kappa_n \in \mathcal{F}_M$. □