Weighted Tree Automata with Constraints

Andreas Maletti[1] · Andreea-Teodora Nász[1]

Abstract

The HOM problem, which asks whether the image of a regular tree language under a given tree homomorphism is again regular, is known to be decidable [Godoy & Giménez: The HOM problem is decidable. JACM 60(4), 2013]. However, the problem remains open for regular weighted tree languages. It is demonstrated that the main notion used in the unweighted setting, the tree automaton with equality and inequality constraints , can straightforwardly be generalized to the weighted setting and can represent the image of any regular weighted tree language under any nondeleting and nonerasing tree homomorphism. Several closure properties as well as decision problems are also investigated for the weighted tree languages generated by weighted tree automata with constraints.

1 Introduction

Numerous extensions of nondeterministic finite-state string automata have been proposed in the past few decades. On the one hand, the qualitative evaluation of inputs was extended to a quantitative evaluation in the weighted automata of [26]. This devel-

---

✉ Andreea-Teodora Nász
   nasz@informatik.uni-leipzig.de

   Andreas Maletti
   maletti@informatik.uni-leipzig.de

[1] Universität Leipzig, Faculty of Mathematics and Computer Science, PO box 100 920, Leipzig 04009, Germany

opment led to the fruitful study of recognizable formal power series [25], which are well-suited for representing factors such as costs, consumption of resources, or time and probabilities related to the processed input. The main algebraic structure for the weight calculations are semirings [16, 17], which offer a nice compromise between generality and efficiency of computation (due to their distributivity). On the other hand, finite-state automata have been generalized to other input structures such as infinite words [24] and trees [4]. Finite-state tree automata were introduced independently in [7, 27, 28], and they and the tree languages they generate, called regular tree languages, have been intensively studied since their inception [4]. They are successfully utilized in various applications in many diverse areas like natural language processing [18], picture generation [8], and compiler construction [31]. Indeed several applications require the combination of the two mentioned generalizations, and a broad range of weighted tree automaton (WTA) models has been studied (see [13, Chapter 9] for an overview).

It is well-known that finite-state tree automata cannot ensure that two subtrees (of potentially arbitrary size) are always equal in an accepted tree [14]. An extension proposed in [22] aims to remedy this problem and introduces a tree automaton model that explicitly can require certain subtrees to be equal or different. Such models are very useful when investigating (tree) transformation models (see [13] for an overview) that can copy subtrees (thus resulting in equal subtrees in the output), and they are the main tool used in the seminal paper [15] that proved that the HOM problem is decidable. The HOM problem was a long-standing open problem in the theory of tree languages and recently solved in [15]. It asks whether the image of an (effectively presented) regular tree language under a given tree homomorphism is again regular. This is not necessarily the case as tree homomorphisms can create copies of subtrees. Indeed removing this ability from the tree homomorphism, obtaining a linear tree homomorphism, yields that the mentioned image is always regular [14]. In the solution to the HOM problem provided in [15] the image is first represented by a tree automaton with constraints, and then it is investigated whether this tree automaton actually generates a regular tree language.

The HOM problem is also interesting in the weighted setting as it once again provides an answer whether a given homomorphic image of a regular weighted tree language can be represented efficiently. While preservation of regularity has been investigated [3, 10–12] also in the weighted setting, the decidability of the HOM problem remains wide open. With the goal of investigating this problem, we introduce weighted tree grammars with constraints (WTGc for short) in this contribution. We demonstrate that those WTGc can again represent all (nondeleting and nonerasing) homomorphic images of the regular weighted tree languages. Thus, in principle, it only remains to provide a decision procedure for determining whether a given WTGc generates a regular weighted tree language. We approach this task by providing some common closure properties following essentially the steps also taken in [15]. For zero-sum free semirings we can also show that decidability of support emptiness and finiteness are directly inherited from the unweighted case [15].

The present work is a revised and extended version of [20] presented at the 26th Int. Conf. Developments in Language Theory (DLT 2022). We provide additional proof details and examples, as well as a new pumping lemma for the class of (nondeleting and nonerasing) homomorphic images of regular weighted tree languages. We utilize this pumping lemma to show that for any zero-sum free semiring, the class of homomorphic images of regular weighted tree languages is properly contained in the class of weighted tree languages generated by all positive WTGc, which are WTGc that utilize only equality constraints.

2 Preliminaries

We denote the set of nonnegative integers by $\mathbb{N}$, and we let $[k] = \{i \in \mathbb{N} \mid 1 \leq i \leq k\}$ for every $k \in \mathbb{N}$. For all sets T and Z let T^Z be the set of all mappings $\varphi \colon Z \to T$, and correspondingly we sometimes write φ_z instead of $\varphi(z)$ for every $\varphi \in T^Z$. The inverse image $\varphi^{-1}(S)$ of φ for a subset $S \subseteq T$ is $\varphi^{-1}(S) = \{z \in Z \mid \varphi(z) \in S\}$, and we write $\varphi^{-1}(t)$ instead of $\varphi^{-1}(\{t\})$ for every $t \in T$. The *range* of φ is

$$\text{ran}(\varphi) = \{\varphi(z) \mid z \in Z\} \ .$$

Let $R \subseteq T \times Z$ be a relation. We denote its inverse relation $\{(z, t) \mid (t, z) \in R\}$ by R^{-1}. The identity relation on T is $\text{id}_T = \{(t, t) \mid t \in T\}$, or simply 'id' if the set T is clear from the context. Finally, the cardinality of Z is denoted by $|Z|$.

A *ranked alphabet* (Σ, rk) is a pair consisting of a finite set Σ and a map $\text{rk} \in \mathbb{N}^\Sigma$ that assigns a rank to each symbol of Σ. If there is no risk of confusion, we denote a ranked alphabet (Σ, rk) by Σ. We write $\sigma^{(k)}$ to indicate that $\text{rk}(\sigma) = k$. Moreover, for every $k \in \mathbb{N}$ we let $\Sigma_k = \text{rk}^{-1}(k)$. Let $X = \{x_i \mid i \geq 1\}$ be a countable set of (formal) variables. For each $k \in \mathbb{N}$ we let $X_k = \{x_i \mid i \in [k]\}$. Given a ranked alphabet Σ and a set Z, the set $T_\Sigma(Z)$ of Σ'' *trees indexed by* Z is the smallest set such that $Z \subseteq T_\Sigma(Z)$ and $\sigma(t_1, \ldots, t_k) \in T_\Sigma(Z)$ for every $k \in \mathbb{N}$, $\sigma \in \Sigma_k$, and $t_1, \ldots, t_k \in T_\Sigma(Z)$. We abbreviate $T_\Sigma(\emptyset)$ simply to T_Σ, and any subset $L \subseteq T_\Sigma$ is called a *tree language*.

Let Σ be a ranked alphabet, Z a set, and $t \in T_\Sigma(Z)$. The set $\text{pos}(t)$ of *positions of* t is inductively defined by $\text{pos}(z) = \{\varepsilon\}$ for all $z \in Z$ and by

$$\text{pos}\big(\sigma(t_1, \ldots, t_k)\big) = \{\varepsilon\} \cup \bigcup_{i \in [k]} \{iw \mid w \in \text{pos}(t_i)\}$$

for all $k \in \mathbb{N}$, $\sigma \in \Sigma_k$, and $t_1, \ldots, t_k \in T_\Sigma(Z)$. The size $|t|$ of t is defined as $|t| = |\text{pos}(t)|$, and its height $\text{ht}(t)$ is $\text{ht}(t) = \max_{w \in \text{pos}(t)} |w|$. For $w \in \text{pos}(t)$ and $t' \in T_\Sigma(Z)$, the *label* $t(w)$ of t at w, the *subtree* $t|_w$ of t at w, and the *substitution* $t[t']_w$ of t' into t at w are defined by $z(\varepsilon) = z|_\varepsilon = z$ and $z[t']_\varepsilon = t'$ for all $z \in Z$ and for $t = \sigma(t_1, \ldots, t_k)$ by $t(\varepsilon) = \sigma$, $t(iw') = t_i(w')$, $t|_\varepsilon = t$, $t|_{iw'} = t_i|_{w'}$, $t[t']_\varepsilon = t'$, and

$$t[t']_{iw'} = \sigma\big(t_1, \ldots, t_{i-1}, t_i[t']_{w'}, t_{i+1}, \ldots, t_k\big)$$

for all $k \in \mathbb{N}, \sigma \in \Sigma_k, t_1, \ldots, t_k \in T_\Sigma(Z), i \in [k]$, and $w' \in \text{pos}(t_i)$. For all $S \subseteq \Sigma \cup Z$, we let $\text{pos}_S(t) = \{w \in \text{pos}(t) \mid t(w) \in S\}$ and $\text{var}(t) = \{x \in X \mid \text{pos}_x(t) \neq \emptyset\}$. For a single $\sigma \in \Sigma \cup Z$ we abbreviate $\text{pos}_{\{\sigma\}}(t)$ simply by $\text{pos}_\sigma(t)$.

The yield mapping yield: $T_\Sigma(Z) \to Z^*$ is recursively defined by

$$\text{yield}(z) = z \quad \text{and} \quad \text{yield}\big(\sigma(t_1, \ldots, t_k)\big) = \text{yield}(t_1) \cdots \text{yield}(t_k)$$

for every $z \in Z, k \in \mathbb{N}, \sigma \in \Sigma_k$, and trees $t_1, \ldots, t_k \in T_\Sigma(Z)$. A tree $t \in T_\Sigma(Z)$ is called *context* if $|\text{pos}_z(t)| = 1$ for every $z \in Z$. We write $C_\Sigma(Z)$ for the set of such contexts and $\widehat{C}_\Sigma(X_k) = \{c \in C_\Sigma(X_k) \mid \text{yield}(c) = x_1 \cdots x_k\}$. Finally, for every $t \in T_\Sigma(Z)$, finite $V \subseteq Z$, and $\theta \in T_\Sigma(Z)^V$, the substitution θ applied to t is written as $t\theta$ and defined by $v\theta = \theta_v$ for every $v \in V$, $z\theta = z$ for every $z \in Z \setminus V$, and

$$\sigma(t_1, \ldots, t_k)\theta = \sigma(t_1\theta, \ldots, t_k\theta)$$

for all $k \in \mathbb{N}, \sigma \in \Sigma_k$, and $t_1, \ldots, t_k \in T_\Sigma(Z)$. We also write the substitution $\theta \in T_\Sigma(Z)^V$ as $[v_1 \leftarrow \theta_{v_1}, \ldots, v_n \leftarrow \theta_{v_n}]$ if $V = \{v_1, \ldots, v_n\}$. Finally, we abbreviate it further to just $[\theta_{v_1}, \ldots, \theta_{v_n}]$ if $V = X_n$.

A *commutative semiring* [16, 17] is a tuple $(\mathbb{S}, +, \cdot, 0, 1)$ such that $(\mathbb{S}, +, 0)$ and $(\mathbb{S}, \cdot, 1)$ are commutative monoids, $\cdot$ distributes over $+$, and $0 \cdot s = 0$ for all $s \in \mathbb{S}$. Examples include (i) the Boolean semiring $\mathbb{B} = \big(\{0, 1\}, \vee, \wedge, 0, 1\big)$, (ii) the semiring $\mathbb{N} = \big(\mathbb{N}, +, \cdot, 0, 1\big)$, (iii) the tropical semiring $\mathbb{T} = \big(\mathbb{N} \cup \{\infty\}, \min, +, \infty, 0\big)$, and (iv) the arctic semiring $\mathbb{A} = \big(\mathbb{N} \cup \{-\infty\}, \max, +, -\infty, 0\big)$. Given two semirings

$$(\mathbb{S}, +, \cdot, 0, 1) \quad \text{and} \quad (\mathbb{T}, \oplus, \odot, \bot, \top) \,,$$

a *semiring homomorphism* is a mapping $h \in \mathbb{T}^\mathbb{S}$ such that $h(0) = \bot$, $h(1) = \top$, and $h(s_1 + s_2) = h(s_1) \oplus h(s_2)$ as well as $h(s_1 \cdot s_2) = h(s_1) \odot h(s_2)$ for all $s_1, s_2 \in \mathbb{S}$. When there is no risk of confusion, we refer to a semiring $(\mathbb{S}, +, \cdot, 0, 1)$ simply by its carrier set $\mathbb{S}$. A semiring $\mathbb{S}$ is a *ring* if there exists $-1 \in \mathbb{S}$ such that $-1 + 1 = 0$. Let Σ be a ranked alphabet. Any mapping $A \in \mathbb{S}^{T_\Sigma}$ is called a *weighted tree language* over $\mathbb{S}$, and its support is $\text{supp}(A) = \{t \in T_\Sigma \mid A_t \neq 0\}$.

Let Σ and Δ be ranked alphabets and $h' \in T_\Delta(X)^\Sigma$ a map such that $h'_\sigma \in T_\Delta(X_k)$ for all $k \in \mathbb{N}$ and $\sigma \in \Sigma_k$. We extend h' to $h \in T_\Delta^{T_\Sigma}$ by (i) $h(\alpha) = h'_\alpha \in T_\Delta(X_0) = T_\Delta$ for all $\alpha \in \Sigma_0$ and (ii) $h\big(\sigma(t_1, \ldots, t_k)\big) = h'_\sigma\big[h(t_1), \ldots, h(t_k)\big]$ for all $k \in \mathbb{N}, \sigma \in \Sigma_k$, and $t_1, \ldots, t_k \in T_\Sigma$. The mapping h is called the *tree homomorphism induced by* h'. For complexity arguments, we define the *size* $|h|$ *of* h as $|h| = \sum_{\sigma \in \Sigma} |\text{pos}(h'_\sigma)|$. The tree homomorphism h is *nonerasing* if $h'_\sigma \notin X$ for all $k \in \mathbb{N}$ and $\sigma \in \Sigma_k$, and it is *nondeleting* if $\text{var}(h'_\sigma) = X_k$ for all $k \in \mathbb{N}$ and $\sigma \in \Sigma_k$. For simplicity, we denote both h' and its induced tree homomorphism by h.

Let $h \in T_\Delta^{T_\Sigma}$ be a nonerasing and nondeleting homomorphism. Then h is *input finitary*; i.e., the set $h^{-1}(u)$ is finite for every $u \in T_\Delta$ because $|t| \leq |u|$ for each $t \in h^{-1}(u)$. Additionally, let $A \in \mathbb{S}^{T_\Sigma}$ be a weighted tree language. We define the weighted tree language $h(A) \in \mathbb{S}^{T_\Delta}$ for every $u \in T_\Delta$ by $h(A)_u = \sum_{t \in h^{-1}(u)} A_t$.

3 Weighted Tree Grammars with Constraints

Let us start with the formal definition of our weighted tree grammars. They are a weighted variant of the tree automata with equality and inequality constraints originally introduced in [1, 5]. Compared to [1, 5] our model is slightly more expressive as we allow arbitrary constraints, whereas constraints were restricted to subtrees occurring in the productions in [1, 5]. This more restricted version will be called classic in the following. An overview of further developments for these automata can be found in [29]. We essentially use the version recently utilized to solve the HOM problem [15, Definition 4.1]. For the rest of this section, let $(\mathbb{S}, +, \cdot, 0, 1)$ be a commutative semiring.

Definition 1 (see [15, Definition 4.1]) A *weighted tree grammar with constraints* (WTGc) is a tuple $G = (Q, \Sigma, F, P, \text{wt})$ such that

- Q is a finite set of nonterminals and $F \in \mathbb{S}^Q$ assigns final weights,
- Σ is a ranked alphabet of input symbols,
- P is a finite set of productions of the form (ℓ, q, E, I), where $\ell \in T_\Sigma(Q) \backslash Q$, $q \in Q$, and $E, I \subseteq \mathbb{N}^* \times \mathbb{N}^*$ are finite sets, and
- $\text{wt} \in \mathbb{S}^P$ assigns a weight to each production.   □

In the following, let $G = (Q, \Sigma, F, P, \text{wt})$ be a WTGc. The components of a production $p = (\ell, q, E, I) \in P$ are the left-hand side ℓ, the target nonterminal q, the set E of equality constraints, and the set I of inequality constraints. Correspondingly, the production p is also written $\ell \xrightarrow{E,I} q$ or even $\ell \xrightarrow{E,I}_{\text{wt}_p} q$ if we want to indicate its weight. Additionally, we simply list an equality constraint $(v, v') \in E$ as $v = v'$ and an inequality constraint $(v, v') \in I$ as $v \neq v'$. A production $\ell \xrightarrow{E,I} q \in P$ is *normalized* if $\ell = \sigma(q_1, \ldots, q_k)$ for some $k \in \mathbb{N}$, $\sigma \in \Sigma_k$, and $q_1, \ldots, q_k \in Q$. It is *positive* if $I = \emptyset$; i.e., it has no inequality constraints, and it is *unconstrained* if $E = \emptyset = I$; i.e., the production has no constraints at all. Instead of $\ell \xrightarrow{\emptyset,\emptyset} q$ we also write just $\ell \to q$. The production is *classic* if $\{v, v'\} \subseteq \text{pos}_Q(\ell)$ for all constraints $(v, v') \in E \cup I$. In other words, in a classic production the constraints can only refer to nonterminal-labeled subtrees of the left-hand side. The WTGc G is a *weighted tree automaton with constraints* (WTAc) if all productions $p \in P$ are normalized, and it is a *weighted tree grammar* (WTG) [14] if all productions $p \in P$ are unconstrained. If G is both a WTAc as well as a WTG, then it is a *weighted tree automaton* (WTA) [14]. All these devices have *Boolean final weights* if $F \in \{0, 1\}^Q$, they are *positive* if every $p \in P$ is positive, and they are *classic* if every production $p \in P$ is classic. Finally, if we utilize the Boolean semiring $\mathbb{B}$, then we reobtain the unweighted versions and omit the 'W' in the abbreviations and the mapping 'wt' from the tuple.

The semantics for our WTGc G is a slightly non-standard *derivation semantics* when compared to [15, Definitions 4.3 & 4.4]. Let $(v, v') \in \mathbb{N}^* \times \mathbb{N}^*$ and $t \in T_\Sigma$. If $v, v' \in \text{pos}(t)$ and $t|_v = t|_{v'}$, we say that t satisfies (v, v'), otherwise t dissatisfies (v, v'). Let now $C \subseteq \mathbb{N}^* \times \mathbb{N}^*$ be a finite set of constraints. We write $t \models C$ if t satisfies all $(v, v') \in C$, and $t \overset{\forall}{\not\models} C$ if t dissatisfies all $(v, v') \in C$. Universally dissatisfying C is generally stronger than simply not satisfying C.

Definition 2 A *sentential form (for G)* is simply a tree of $\xi \in T_\Sigma(Q)$. Given an input tree $t \in T_\Sigma$, sentential forms $\xi, \zeta \in T_\Sigma(Q)$, a production $p = \ell \xrightarrow{E,I} q \in P$, and a position $w \in \mathrm{pos}(\xi)$, we write $\xi \Rightarrow^{p,w}_{G,t} \zeta$ if $\xi|_w = \ell$, $\zeta = \xi[q]_w$, and the constraints E and I are fulfilled on $t|_w$; i.e., $t|_w \models E$ and $t|_w \not\models^{\forall} I$. A sequence

$$d = (p_1, w_1) \cdots (p_n, w_n) \in (P \times \mathbb{N}^*)^*$$

is a *derivation of G for t* if there exist $\xi_1, \ldots, \xi_n \in T_\Sigma(Q)$ such that

$$t \Rightarrow^{p_1,w_1}_{G,t} \xi_1 \Rightarrow^{p_2,w_2}_{G,t} \cdots \Rightarrow^{p_n,w_n}_{G,t} \xi_n \ .$$

It is *left-most* if additionally $w_1 \prec w_2 \prec \cdots \prec w_n$, where $\preceq$ is the lexicographic order on $\mathbb{N}^*$ in which prefixes are larger, so ε is the largest element. □

Note that the sentential forms $\xi_1, \ldots, \xi_n$ are uniquely determined if they exist, and for any derivation d for t there exists a unique permutation of d that is a left-most derivation for t. The derivation d is *complete* if $\xi_n \in Q$, and in that case it is also called a *derivation to* ξ_n. The set of all complete left-most derivations for t to $q \in Q$ is denoted by $D^q_G(t)$. The WTGc G is *unambiguous* if $\sum_{q \in \mathrm{supp}(F)} |D^q_G(t)| \leq 1$ for every $t \in T_\Sigma$.

Let $p = \ell \xrightarrow{E,I} q \in P$ be a production. Since there exist unique $k = |\mathrm{pos}_Q(\ell)|$, context $c \in \widehat{C}_\Sigma(X_k)$, and $q_1, \ldots, q_k \in Q$ such that $\ell = c[q_1, \ldots, q_k]$, we also simply write

$$c[q_1, \ldots, q_k] \xrightarrow{E,I} q$$

instead of p. Using this notation, we can present a recursion for the set $D^q_G(t)$ of complete left-most derivations for $t \in T_\Sigma$ to $q \in Q$.

$$D^q_G(t) = \left\{ d_1 \cdots d_k (p, \varepsilon) \;\middle|\; k \in \mathbb{N},\ p = c[q_1, \ldots, q_k] \xrightarrow{E,I} q \in P,\ t \models E,\ t \not\models^{\forall} I \right.$$
$$\left. t_1, \ldots, t_k \in T_\Sigma,\ t = c[t_1, \ldots, t_k],\ \forall i \in [k]\colon d_i \in D^{q_i}_G(t_i) \right\}$$

Specifically, let $d = (p_1, w_1) \cdots (p_n, w_n)$ be a complete derivation for some tree $t \in T_\Sigma$. For a given position $w \in \{w_1, \ldots, w_n\}$, we let $k \in \mathbb{N}$ and $1 \leq i_1 < \cdots < i_k \leq n$ be the indices such that $\{i_1, \ldots, i_k\} = \{i \in [n] \mid w_i = ww'_i\}$; i.e., the indices of the derivation steps applied to positions below w with w'_i being the suffix of w_i following the prefix w for all $i \in \{i_1, \ldots, i_k\}$. The *derivation for $t|_w$ incorporated in d* is the derivation $(p_{i_1}, w'_{i_1}), \ldots, (p_{i_k}, w'_{i_k})$. Conversely, for every $w \in \mathbb{N}^*$ we abbreviate the derivation $(p_1, ww_1) \cdots (p_n, ww_n)$ by simply wd.

Definition 3 The *weight* of a derivation $d = (p_1, w_1) \cdots (p_n, w_n)$ is defined to be

$$\mathrm{wt}_G(d) = \prod_{i=1}^{n} \mathrm{wt}(p_i) \ .$$

The weighted tree language generated by G, written simply $G \in \mathbb{S}^{T_\Sigma}$, is defined for every $t \in T_\Sigma$ by

$$G_t = \sum_{q \in Q, \, d \in D_G^q(t)} F_q \cdot \mathrm{wt}_G(d) \; . \qquad \qquad \square$$

Two WTGc are *equivalent* if they generate the same weighted tree language. Finally, a weighted tree language is

- *regular* if it is generated by some WTG,
- *positive constraint-regular* if it is generated by some positive WTGc,
- *classic constraint-regular* if it is generated by some classic WTGc, and
- *constraint-regular* if it is generated by some WTGc.

Since the weights of productions are multiplied, we can assume without loss of generality that $\mathrm{wt}_p \neq 0$ for all $p \in P$.

Example 1 Consider the WTGc $G = (Q, \Sigma, F, P, \mathrm{wt})$ over the arctic semiring $\mathbb{A}$ with nonterminals $Q = \{q, q'\}$, $\Sigma = \{\alpha^{(0)}, \gamma^{(1)}, \sigma^{(2)}\}$, $F_q = -\infty$, $F_{q'} = 0$, and P and 'wt' given by the productions $p_1 = \alpha \to_0 q$, $p_2 = \gamma(q) \to_1 q$, and $p_3 = \sigma(\gamma(q), q) \xrightarrow{11=2}_1 q'$. Clearly, G is positive and classic, but not a WTAc. The tree $t = \sigma(\gamma(\gamma(\alpha)), \gamma(\alpha))$ has the unique left-most derivation

$$d = (p_1, 111) \, (p_2, 11) \, (p_1, 21) \, (p_2, 2) \, (p_3, \varepsilon)$$

to the nonterminal q', which is illustrated in Fig. 1. Overall, we have

$$\mathrm{supp}(G) = \left\{ \sigma\left(\gamma^{i+1}(\alpha), \gamma^i(\alpha)\right) \mid i \in \mathbb{N} \right\}$$

and $G_t = |\mathrm{pos}_\gamma(t)|$ for every $t \in \mathrm{supp}(G)$, where $\gamma^i(t)$ abbreviates $\gamma(\cdots\gamma(t)\cdots)$ containing i" times the unary symbol γ atop t. $\qquad \square$

Next, we introduce another semantics, called initial algebra semantics, which is based on the presented recursive presentation of derivations and often more convenient in proofs.
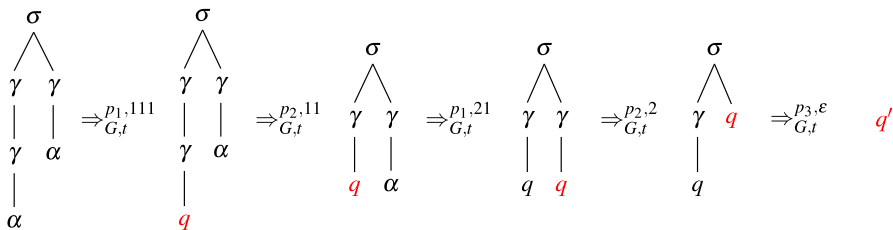

Fig. 1 Illustration of the derivation mentioned in Example 1

Definition 4 For every nonterminal $q \in Q$ we recursively define the map $\mathrm{wt}_G^q \in \mathbb{S}^{T_\Sigma}$ such that for every $t \in T_\Sigma$ by

$$\mathrm{wt}_G^q(t) = \sum_{\substack{p=c[q_1,\ldots,q_k] \xrightarrow{E,I} q \in P \\ t_1,\ldots,t_k \in T_\Sigma \\ t=c[t_1,\ldots,t_k] \\ t \models E,\, t \not\models I}} \mathrm{wt}_p \cdot \prod_{i=1}^{k} \mathrm{wt}_G^{q_i}(t_i) \ . \tag{1}$$

□

It is a routine matter to verify that $\mathrm{wt}_G^q(t) = \sum_{d \in D_G^q(t)} \mathrm{wt}_G(d)$ for every $q \in Q$ and $t \in T_\Sigma$. This utilizes the presented recursive decomposition of complete derivations as well as distributivity of the semiring $\mathbb{S}$.

As for WTG and WTA [13], also every (positive) WTGc can be turned into an equivalent (positive) WTAc at the expense of additional nonterminals by decomposing the left-hand sides.

Lemma 1 (cf. [15, Lemma 4.8]) *WTGc and WTAc are equally expressive. This also applies to positive WTGc.*

Proof Let $G = (Q, \Sigma, F, P, \mathrm{wt})$ be a WTGc with a non-normalized production

$$p = \sigma(\ell_1, \ldots, \ell_k) \xrightarrow{E,I} q \in P \ ,$$

let $U \supseteq Q$ and let $\varphi \in U^{T_\Sigma(Q)}$ be an injective map such that $\varphi_q = q$ for all $q \in Q$. We define the WTGc $G' = (Q', \Sigma, F', P', \mathrm{wt}')$ such that $Q' = Q \cup \{\varphi_{\ell_1}, \ldots, \varphi_{\ell_k}\}$, $F_q' = F_q$ for all $q \in Q$ and $F_{q'}' = 0$ for all $q' \in Q' \backslash Q$, and

$$P' = \left(P \backslash \{p\}\right) \cup \left\{\sigma(\varphi_{\ell_1}, \ldots, \varphi_{\ell_k}) \xrightarrow{E,I} q\right\} \cup \left\{\ell_i \to \varphi_{\ell_i} \mid i \in [k], \ell_i \notin Q\right\} \ ,$$

and for every $p' \in P'$

$$\mathrm{wt}_{p'}' = \begin{cases} \mathrm{wt}_{p'} & \text{if } p' \in P \backslash \{p\} \\ \mathrm{wt}_p & \text{if } p' = \sigma(\varphi_{\ell_1}, \ldots, \varphi_{\ell_k}) \xrightarrow{E,I} q \\ 1 & \text{otherwise.} \end{cases}$$

To prove that G' is equivalent to G we observe that for every left-most derivation

$$d = (p_1, w_1) \cdots (p_n, w_n)$$

of G, there exists a corresponding derivation d' of G', which is obtained by replacing each derivation step (p_a, w_a) with $p_a = p$ by the sequence

$$(\ell_i \to \varphi_{\ell_i}, w_a i)_{i \in [k], \ell_i \notin Q} \left(\sigma(\varphi_{\ell_1}, \ldots, \varphi_{\ell_k}) \xrightarrow{E,I} q, w_a\right)$$

of derivation steps of G' (yielding also a unique corresponding left-most derivation). This replacement preserves the weight of the derivation. Vice versa any left-most derivation of G' that utilizes the production $\sigma(\varphi_{\ell_1}, \ldots, \varphi_{\ell_k}) \xrightarrow{E,I} q \in P'$ at w needs to previously utilize the productions $\ell_i \to \varphi_{\ell_i} \in P'$ at wi for all $i \in [k]$ with $\ell_i \notin Q$ since these are the only productions that generate the nonterminal φ_{ℓ_i}. Thus, we established a weight-preserving bijection between the left-most derivations of G and G', so it is obvious that $G' = G$. Repeated application of the normalization yields an equivalent WTAc. Assuming a fixed ranked alphabet Σ, every step of the normalization adds at most $\max_{\sigma \in \Sigma} \text{rk}(\sigma)$ new productions and states that can be computed in constant time. In total $\sum_{c[q_1, \ldots, q_k] \xrightarrow{E,I} q \in P} (|\text{pos}_\Sigma(c)| - 1)$ steps will be executed, so the overall construction runs in linear time and returns an output WTAc whose size is linear in the size of the input WTGc. Finally, we note that the constructed WTAc is positive if the original WTGc is positive. □

As we will see in the next example, the construction used in the proof of Lemma 1 does not preserve the classic property.

Example 2 Consider the classic and positive WTGc G of Example 1 and its non-normalized production $p = \sigma(\gamma(q), q) \xrightarrow{11=2}_1 q'$. Applying the construction in the proof of Lemma 1 we replace p by the productions $\sigma(q'', q) \xrightarrow{11=2}_1 q$, which is not classic, and $\gamma(q) \to_0 q''$, where q'' is some new nonterminal. The WTGc obtained this way is already a positive WTAc. □

Another routine normalization turns the final weights into Boolean final weights following the approach of [2, Lemma 6.1.1]. This is achieved by adding special copies of all nonterminals that terminate the derivation and pre-apply the final weight.

Lemma 2 *WTGc and WTGc with Boolean final weights are equally expressive. This also applies to positive WTGc, classic WTGc, and classic positive WTGc as well as the same WTAc.*

Proof Let $G = (Q, \Sigma, F, P, \text{wt})$ be a WTGc. Let $f \in C^Q$ be bijective with $C \cap Q = \emptyset$. We construct the WTGc $G' = (Q \cup C, \Sigma, F', P \cup P', \text{wt} \cup \text{wt}')$ such that $p' = \ell \xrightarrow{E,I} f_q$ belongs to P' and $\text{wt}'_{p'} = \text{wt}_p \cdot F_q$ for every $p = \ell \xrightarrow{E,I} q \in P$. No other productions belong to P'. Finally, $F'_q = 0$ for all $q \in Q$ and $F_c = 1$ for all $c \in C$. The proof of equivalence is straightforward showing for every $t \in T_\Sigma$ and $q \in Q$ that

$$\text{wt}^q_{G'}(t) = \text{wt}^q_G(t) \quad \text{and} \quad \text{wt}^{f(q)}_{G'}(t) = \text{wt}^q_G(t) \cdot F_q \ .$$

The construction trivially preserves the properties normalized, positive, and classic. □

Let $d \in D^q_G(t)$ be a derivation for some $q \in Q$ and $t \in T_\Sigma$. Since we often argue with the help of such derivations d, it is a nuisance that we might have $\text{wt}_G(d) = 0$. This anomaly can occur even if $\text{wt}_p \neq 0$ for all $p \in P$ due to the presence of zero-divisors, which are elements $s, s' \in \mathbb{S} \setminus \{0\}$ such that $s \cdot s' = 0$. However, we can

fortunately avoid such anomalies altogether utilizing a construction of [19], which has been lifted to tree automata in [9].

Lemma 3 *For every WTGc G there exists a WTGc $G' = (Q', \Sigma, F', P', \text{wt}')$ that is equivalent and $\text{wt}'_{G'}(d') \neq 0$ for all $q' \in Q'$, $t' \in T_\Sigma$, and $d' \in D^{q'}_{G'}(t')$. This also applies to positive WTGc, classic WTGc, and classic positive WTGc as well as the same WTAc. The construction also preserves Boolean final weights.*

Proof Let $G = (Q, \Sigma, F, P, \text{wt})$. Obviously, $(\mathbb{S}, \cdot, 1, 0)$ is a commutative monoid with zero. Let $(s_1, \dots, s_n)$ be an enumeration of the finite set $\text{wt}(P) \setminus \{1\} \subseteq \mathbb{S}$. We consider the monoid homomorphism $h \colon \mathbb{N}^n \to \mathbb{S}$, which is given by

$$h(m_1, \dots, m_n) = \prod_{i=1}^{n} s_i^{m_i}$$

for every $m_1, \dots, m_n \in \mathbb{N}$. According to DICKSON's lemma [6] the set $\min h^{-1}(0)$ is finite, where the partial order is the standard pointwise order on $\mathbb{N}^n$. Hence there is $u \in \mathbb{N}$ such that $\min h^{-1}(0) \subseteq \{0, \dots, u\}^n = U$. We define the operation $\oplus \colon U^2 \to U$ by $(v \oplus v')_i = \min(v_i + v'_i, u)$ for every $v, v' \in U$ and $i \in [n]$. Moreover, for every $i \in [n]$ we let $1_{s_i} \in U$ be the vector such that $(1_{s_i})_i = 1$ and $(1_{s_i})_a = 0$ for all $a \in [n] \setminus \{i\}$. Let $V = U \setminus h^{-1}(0)$. We construct the equivalent WTGc G' such that $Q' = Q \times V$, $F'_{\langle q, v \rangle} = F_q$ for all $\langle q, v \rangle \in Q'$, and P' and wt' are given as follows. For every production

$$p = c[q_1, \dots, q_k] \xrightarrow{E, I} q \in P$$

and all $v_1, \dots, v_k \in V$ such that $v = 1_{\text{wt}_p} \oplus \bigoplus_{i=1}^{k} v_i \in V$ the production

$$c\big[\langle q_1, v_1 \rangle, \dots, \langle q_k, v_k \rangle\big] \xrightarrow{E, I} \langle q, v \rangle$$

belongs to P' and its weight is $\text{wt}'_{p'} = \text{wt}_p$. No further productions are in P'. The construction trivially preserves the properties positive, classic, and normalized. For correctness, let $q' = \langle q, v \rangle \in Q'$, $t' \in T_\Sigma$, and $d' \in D^{q'}_{G'}(t')$. We suitably (for the purpose of zero-divisors) track the weight of the derivation in v and $h_v \neq 0$ by definition. Consequently, $\text{wt}'_{G'}(d') \neq 0$ as required. We note that possibly $\text{wt}_{G'}(d') \neq h_v$. $\qquad \square$

For zero-sum free semirings [16, 17] we obtain that the support $\text{supp}(G)$ of an WTGc can be generated by a TGc. A semiring is *zero-sum free* if $s = 0 = s'$ for every $s, s' \in \mathbb{S}$ such that $s + s' = 0$. Clearly, rings are never zero-sum free, but the mentioned semirings $\mathbb{B}$, $\mathbb{N}$, $\mathbb{T}$, and $\mathbb{A}$ are all zero-sum free.

Corollary 1 (of Lemmas 2 and 3) *If $\mathbb{S}$ is zero-sum free, then $\text{supp}(G)$ is (positive, classic) constraint-regular for every (respectively, positive, classic) WTGc G.*

Proof We apply Lemma 2 to obtain an equivalent WTGc with Boolean final weights and then Lemma 3 to obtain the WTGc $G' = (Q', \Sigma, F', P', \text{wt}')$ with Boolean final

weights. As mentioned we can assume that $\text{wt}'_{p'} \neq 0$ for all $p' \in P'$. Let $q' \in \text{supp}(F')$ and $t' \in T_\Sigma$ with $D_{G'}^{q'}(t') \neq \emptyset$. Since $\text{wt}'_{G'}(d') \neq 0$ for every derivation $d' \in D_{G'}^{q'}(t')$ and $s + s' \neq 0$ for all $s, s' \in \mathbb{S} \setminus \{0\}$ due to zero-sum freeness, we obtain $t' \in \text{supp}(G')$. Thus, the existence of a complete derivation for t' to an accepting nonterminal (i.e., one with final weight 1) characterizes whether we have $t' \in \text{supp}(G')$. Consequently, the TGc $(Q', \Sigma, \text{supp}(F'), P')$ generates the tree language $\text{supp}(G')$, which is thus constraint-regular. The properties positive and classic are preserved in all the constructions. $\qquad\square$

4 Closure Properties

Next we investigate several closure properties of the constraint-regular weighted tree languages. We start with the (point-wise) sum, which is given by $(A + A')_t = A_t + A'_t$ for every $t \in T_\Sigma$ and $A, A' \in \mathbb{S}^{T_\Sigma}$. Given WTGc G and G' generating A and A' we can trivially use a disjoint union construction to obtain a WTGc generating $A + A'$. We omit the details.

Proposition 1 *The (positive, classical) constraint-regular weighted tree languages (over a fixed ranked alphabet) are closed under sums.* $\qquad\square$

The corresponding (point-wise) product is the HADAMARD product, which is given by $(A \cdot A')_t = A_t \cdot A'_t$ for every $t \in T_\Sigma$ and $A, A' \in \mathbb{S}^{T_\Sigma}$. With the help of a standard product construction we show that the (positive) constraint-regular weighted tree languages are also closed under HADAMARD product. As preparation we introduce a special normal form. A WTAc $G = (Q, \Sigma, F, P, \text{wt})$ is *constraint-determined* if $E = E'$ and $I = I'$ for all productions

$$\sigma(q_1, \ldots, q_k) \xrightarrow{E, I} q \in P \quad \text{and} \quad \sigma(q_1, \ldots, q_k) \xrightarrow{E', I'} q \in P \ .$$

In other words, two productions cannot differ only in the sets of constraints. It is straightforward to turn any (positive) WTAc into an equivalent constraint-determined (positive) WTAc by introducing additional nonterminals (e.g. annotate the constraints to the nonterminal on the right-hand side).

Theorem 1 *The (positive) constraint-regular weighted tree languages (over a fixed ranked alphabet) are closed under HADAMARD products.*

Proof Let $A, A' \in \mathbb{S}^{T_\Sigma}$ be constraint-regular. Without loss of generality (see Lemma 1) we can assume constraint-determined WTAc

$$G = (Q, \Sigma, F, P, \text{wt}) \quad \text{and} \quad G' = (Q', \Sigma, F', P', \text{wt}')$$

that generate A and A', respectively. We construct the direct product WTAc

$$G \times G' = (Q \times Q', \Sigma, F'', P'', \text{wt}'')$$

such that $F''_{\langle q, q' \rangle} = F_q \cdot F'_{q'}$ for every $q \in Q$ and $q' \in Q'$ and for every production $p = \sigma(q_1, \ldots, q_k) \xrightarrow{E,I} q \in P$ and production $p' = \sigma(q'_1, \ldots, q'_k) \xrightarrow{E',I'} q' \in P'$ the production

$$p'' = \sigma\big(\langle q_1, q'_1 \rangle, \ldots, \langle q_k, q'_k \rangle\big) \xrightarrow{E \cup E', I \cup I'} \langle q, q' \rangle$$

belongs to P'' and its weight is $\text{wt}''_{p''} = \text{wt}_p \cdot \text{wt}'_{p'}$. No other productions belong to P''. It is straightforward to see that the property positive is preserved. The correctness proof that $G \times G' = A \cdot A'$ is a straightforward induction proving

$$\text{wt}_{G \times G'}^{\langle q, q' \rangle}(t) = \text{wt}_G^q(t) \cdot \text{wt}_{G'}^{q'}(t)$$

for all $t \in T_\Sigma$ using the initial algebra semantics. The WTAc G and G' are required to be constraint-determined, so that we can uniquely identify the basic productions $p \in P$ and $p' \in P'$ that construct a newly formed production $p'' \in P''$.

We can obtain a constraint-determined WTAc at the expense of a polynomial increase in the number of productions (assuming that the ranked alphabet of input symbols is fixed). Let $r = \max_{\sigma \in \Sigma} \text{rk}(\sigma)$ be the maximal rank of an input symbol and $c = |P|$ be the number of productions of the given WTAc $G = (Q, \Sigma, F, P, \text{wt})$. First, we modify the target nonterminal q of each production $\rho = (\ell, q, E, I) \in P$ to additionally include the identifier ρ, which yields the production $(\ell, \langle q, \rho \rangle, E, I)$. This effectively yields the new nonterminal set $Q \times P$, which has size $|Q| \cdot c$. Then the copies of the production $(\sigma(q_1, \ldots, q_k), \langle q, \rho \rangle, E, I)$ form the set

$$\left\{ \big(\sigma(\langle q_1, \rho_1 \rangle, \ldots, \langle q_k, \rho_k \rangle), \langle q, \rho \rangle, E, I\big) \;\middle|\; \rho_1, \ldots, \rho_k \in P \right\} .$$

Clearly, this turns each production into at most c^r productions since $k \leq r$, so the overall number of productions after all replacements is at most c^{r+1}. The product construction itself is then quadratic. □

We note that the previous construction also works for classic WTAc.

Example 3 Let $G = \big(\{q\}, \Sigma, F, P, \text{wt}\big)$ and $G' = \big(\{z\}, \Sigma, F', P', \text{wt}'\big)$ be WTAc over $\mathbb{A}$ and $\Sigma = \{\alpha^{(0)}, \gamma^{(1)}, \sigma^{(2)}\}$, $F_q = F'_z = 0$, and the productions

$$\alpha \to_0 q \qquad \gamma(q) \to_2 q \qquad \sigma(q,q) \xrightarrow{1=2}_0 q \qquad (P)$$

$$\alpha \to_0 z \qquad \gamma(z) \xrightarrow{11 \neq 12}_1 z \qquad \sigma(z,z) \to_1 z . \qquad (P')$$

We observe that

$$\text{supp}(G) = \big\{ t \in T_\Sigma \mid \forall w \in \text{pos}_\sigma(t) \colon t|_{w1} = t|_{w2} \big\}$$

$$\text{supp}(G') = \big\{ t \in T_\Sigma \mid \forall w \in \text{pos}_\gamma(t) \colon \text{ if } t(w1) = \sigma \text{ then } t|_{w11} \neq t|_{w12} \big\}$$

and $G_t = 2|\text{pos}_\gamma(t)|$ as well as $G'_{t'} = |\text{pos}_\gamma(t')| + |\text{pos}_\sigma(t')|$ for every tree $t \in \text{supp}(G)$ and tree $t' \in \text{supp}(G')$. We obtain the WTAc $G \times G' = (\{\langle q, z \rangle\}, \Sigma, F'', P'', \text{wt}'')$ with $F''_{\langle q,z \rangle} = 0$ and the following productions.

$$\alpha \to_0 \langle q, z \rangle \qquad \gamma(\langle q, z \rangle) \xrightarrow{11 \neq 12}_3 \langle q, z \rangle \qquad \sigma(\langle q, z \rangle, \langle q, z \rangle) \xrightarrow{1=2}_1 \langle q, z \rangle$$

Hence we obtain the equality $(G \times G')_t = 3|\text{pos}_\gamma(t)| + |\text{pos}_\sigma(t)| = G_t \cdot G'_t$ for every tree $t \in \text{supp}(G) \cap \text{supp}(G')$. $\qquad \square$

Next, we use an extended version of the classical power set construction to obtain an unambiguous WTAc that keeps track of the reachable nonterminals, but preserves only the homomorphic image of its weight. The unweighted part of the construction mimics a power-set construction and the handling of constraints roughly follows [15, Definition 3.1].

Theorem 2 *Let $h \in \mathbb{T}^\mathbb{S}$ be a semiring homomorphism into a finite semiring $\mathbb{T}$. For every (classic) WTAc $G = (Q, \Sigma, F, P, \text{wt})$ over $\mathbb{S}$ there exists an unambiguous (classic) WTAc $G' = (\mathbb{T}^Q, \Sigma, F', P', \text{wt}')$ such that for every tree $t \in T_\Sigma$ and $\varphi \in \mathbb{T}^Q$*

$$\text{wt}^\varphi_{G'}(t) = \begin{cases} 1 & \text{if } \varphi_q = h\big(\text{wt}^q_G(t)\big) \text{ for all } q \in Q \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, $G'_t = h(G_t)$ for every $t \in T_\Sigma$.

Proof For every $\sigma \in \Sigma$, let

$$\mathscr{C}_\sigma = \big\{ E \mid \sigma(q_1, \ldots, q_k) \xrightarrow{E,I} q \in P \big\} \cup \big\{ I \mid \sigma(q_1, \ldots, q_k) \xrightarrow{E,I} q \in P \big\}$$

be the constraints that occur in productions of G whose left-hand side contains σ. We let $F'_\varphi = \sum_{q \in Q} h(F_q) \cdot \varphi_q$ for every $\varphi \in \mathbb{T}^Q$. For all $k \in \mathbb{N}$, $\sigma \in \Sigma_k$, nonterminals $\varphi^1, \ldots, \varphi^k \in \mathbb{T}^Q$, and constraints $\mathscr{E} \subseteq \mathscr{C}_\sigma$ we let $p' = \sigma(\varphi^1, \ldots \varphi^k) \xrightarrow{\mathscr{E}, \mathscr{I}} \varphi \in P'$, where $\mathscr{I} = \mathscr{C}_\sigma \setminus \mathscr{E}$ and for every $q \in Q$

$$\varphi_q = \sum_{\substack{p = \sigma(q_1, \ldots, q_k) \xrightarrow{E,I} q \in P \\ E \subseteq \mathscr{E}, I \subseteq \mathscr{I}}} h(\text{wt}_p) \cdot \varphi^1_{q_1} \cdot \ldots \cdot \varphi^k_{q_k} \; . \tag{2}$$

No additional productions belong to P'. Finally, we set $\text{wt}'_{p'} = 1$ for all $p' \in P'$. In general, the WTAc G' is certainly not deterministic due to the choice of constraints, but G' is unambiguous since the resulting $2^{|\mathscr{C}_\sigma|}$ rules for each left-hand side have mutually exclusive constraint sets. In fact, for each $t \in T_\Sigma$ there is exactly one left-most complete derivation of G' for t, and it derives to $\varphi \in \mathbb{T}^Q$ such that $\varphi_q = h\big(\text{wt}^q_G(t)\big)$ for every $q \in Q$. The weight of that derivation is 1. These statements are proven

inductively. The final statement $G'_t = h(G_t)$ for every $t \in T_\Sigma$ is an easy consequence of the previous statements. If G is classic, then also the constructed WTAc G' is classic.
□

Example 4 Recall the WTAc G and G' from Example 3. Consider the WTAc generating their disjoint union, as well as the semiring homomorphism $h \in \mathbb{B}^\mathbb{A}$ given by $h_a = 1$ for all $a \in \mathbb{A}\backslash\{-\infty\}$ and $h_{-\infty} = 0$. The sets $\mathscr{C}_\gamma$ and $\mathscr{C}_\sigma$ of utilized constraints are $\mathscr{C}_\gamma = \{(11, 12)\}$ and $\mathscr{C}_\sigma = \{(1, 2)\}$, and we write $\varphi \in \mathbb{B}^Q$ simply as subsets of Q. We obtain the unambiguous WTAc G'' with the following sensible (i.e., having satisfiable constraints) productions for all $Q', Q'' \subseteq \{q, z\}$, which all have weight 1.

$$\alpha \longrightarrow \{q, z\}$$

$$\gamma(Q') \xrightarrow{11=12} Q' \cap \{q\} \qquad\qquad \gamma(Q') \xrightarrow{11\neq12} Q'$$

$$\sigma(Q', Q'') \xrightarrow{1=2} Q' \cap Q'' \qquad\qquad \sigma(Q', Q'') \xrightarrow{1\neq2} Q' \cap Q'' \cap \{z\}$$

Each $t \in T_\Sigma$ has exactly one left-most complete derivation in G''; it derives to Q', where (i) $q \in Q'$ iff $t \in \mathrm{supp}(G)$ and (ii) $z \in Q'$ iff $t \in \mathrm{supp}(G')$. It is $F''_\emptyset = 0$ and $F''_Q = 1$ for all non-empty $Q \subseteq \{q, z\}$.
□

Corollary 2 (of Theorem 2) *Let $\mathbb{S}$ be finite. For every (classic) WTAc over $\mathbb{S}$ there exists an equivalent unambiguous (classic) WTAc.*
□

Corollary 3 (of Theorem 2) *Let $\mathbb{S}$ be zero-sum free. For every (classic) WTAc G over $\mathbb{S}$ there exists an unambiguous (classic) TAc generating $\mathrm{supp}(G)$.*

Proof Utilizing Lemma 2 we can first construct an equivalent WTAc with Boolean final weights. If $\mathbb{S}$ is zero-sum free, then there exists a semiring homomorphism $h \in \mathbb{B}^\mathbb{S}$ by [30]. By Lemma 3 we can assume that each derivation of G has non-zero weight and sums of non-zero elements remain non-zero by zero-sum freeness. Thus we can simply replace the factor $h(\mathrm{wt}_p)$ by 1 in (2). The such obtained TAc generates $\mathrm{supp}(G)$. □

Corollary 4 (of Theorem 2) *Let $\mathbb{S}$ be zero-sum free. For every (classic) WTAc G over $\mathbb{S}$ there exists an unambiguous (classic) TAc generating $T_\Sigma \setminus \mathrm{supp}(G)$.*

Proof Let $G' = (Z, \Sigma, Z_0, P')$ be the unambiguous TAc given by Corollary 3. Since G' is also complete in the sense that every input tree has a derivation, the desired unambiguous TAc G'' is simply $G'' = (Z, \Sigma, Z \setminus Z_0, P')$. □

Let $A, A' \in \mathbb{S}^{T_\Sigma}$. It is often useful (see [15, Definition 4.11]) to restrict A to the support of A' but without changing the weights of those trees inside the support. Formally, we define $A|_{\mathrm{supp}(A')} \in \mathbb{S}^{T_\Sigma}$ for every $t \in T_\Sigma$ by $A|_{\mathrm{supp}(A')}(t) = A_t$ if $t \in \mathrm{supp}(A')$ and $A|_{\mathrm{supp}(A')}(t) = 0$ otherwise. Utilizing unambiguous WTAc and the HADAMARD product, we can show that $A|_{\mathrm{supp}(A')}$ is constraint-regular if A and A' are constraint-regular and the semiring $\mathbb{S}$ is zero-sum free.

Theorem 3 *Let $\mathbb{S}$ be zero-sum free. For all (classic) WTAc G and G' there exists a (classic) WTAc H such that $H = G|_{\mathrm{supp}(G')}$.*

Proof By Corollary 1 the support $\mathrm{supp}(G')$ is constraint-regular. Hence we can obtain an unambiguous WTAc G'' for $\mathrm{supp}(G')$ using Corollary 3. Without loss of generality we assume that both G and G'' are constraint-determined; we note that the normalization preserves unambiguous WTAc. Finally we construct $G \times G''$, which by Theorem 1 generates exactly $G|_{\mathrm{supp}(G')}$ as required.   □

In the following, we establish a special property for classic WTGc. To this end, we first need another notion. Let $G = (Q, \Sigma, F, P, \mathrm{wt})$ be a WTGc. A nonterminal $\bot \in Q$ is a *sink nonterminal (in G)* if $F_\bot = 0$ and

$$\big\{\sigma(\bot, \ldots, \bot) \to_1 \bot \mid \sigma \in \Sigma\big\} = \big\{\ell \xrightarrow{E,I}_s q \in P \mid q = \bot\big\} \ .$$

In other words, for every sink nonterminal $\bot$ the production $\sigma(\bot, \ldots, \bot) \to \bot$ belongs to P with weight 1 for every symbol $\sigma \in \Sigma$. Additionally, no other productions have the sink nonterminal $\bot$ as target nonterminal. Given a set $E \subseteq \mathbb{N}^* \times \mathbb{N}^*$ of equality constraints, we let $\equiv_E \, = (E \cup E^{-1} \cup \mathrm{id})^+$ be the smallest equivalence relation containing E (where the superscript $^+$ denotes the transitive closure), and $[w]_{\equiv_E}$ be the equivalence class of $w \in \mathbb{N}^*$. Additionally, for every $c[q_1, \ldots, q_k] \xrightarrow{E,I} q \in P$ we let

$$c(E) = \big\{(i, j) \in [k] \times [k] \mid (v, v') \in E, \, c(v) = x_i, \, c(v') = x_j\big\}$$

be a representation of the equality constraints on the indices $[k]$.

Definition 5 A classic WTGc $G = (Q, \Sigma, F, P, \mathrm{wt})$ is *eq″ restricted* if there exists a sink nonterminal $\bot \in Q$ such that for every production $p = c[q_1, \ldots, q_k] \xrightarrow{E,I} q \in P$ and index $i \in [k]$ there exists a nonterminal $q' \in Q$ such that

1. $\{q_j \mid j \in [i]_{\equiv_{c(E)}}\} \subseteq \{q', \bot\}$ and
2. there exists exactly one index $j \in [i]_{\equiv_{c(E)}}$, also called *governing index for i in p*, such that $q_j = q'$.

The mapping $g_p \colon [k] \to [k]$ assigns to each index $i \in [k]$ its governing index for i in p. □

In other words, in an eq″ restricted WTGc one subtree is generated normally by the WTGc and all the subtrees that are required to be equal by means of the equality constraints are generated by the sink nonterminal $\bot$, which can generate any tree with weight 1. In this manner, the restrictions on subtree and weight generation induced by the WTGc are exhibited completely on a single subtree and the "copies" are only provided by the equality constraint, but not further restricted by the WTGc. We will continue to use $\bot$ for the suitable sink nonterminal of an eq″ restricted WTGc.

Finally, we show that the weighted tree languages generated by eq″ restricted positive WTGc are closed under relabelings. A *relabeling* is a tree homomorphism of the form $\pi \in T_\Delta(X)^\Sigma$ such that for every $k \in \mathbb{N}$ and $\sigma \in \Sigma_k$ there exists $\delta \in \Delta_k$ with $\pi_\sigma = \delta(x_1, \ldots, x_k)$. In other words, a relabeling deterministically replaces symbols respecting their rank. We often specify a relabeling just as a mapping $\pi \in \Delta^\Sigma$ such that $\pi_\sigma \in \Delta_k$ for every $k \in \mathbb{N}$ and $\sigma \in \Sigma_k$.

Theorem 4 *The weighted tree languages generated by eq-restricted positive WTGc are closed under relabelings.*

Proof Let WTGc $G = (Q, \Sigma, F, P, \text{wt})$ be an eq" restricted positive WTGc with sink nonterminal $\perp$. Without loss of generality, suppose that $\Sigma \cap X = \emptyset$, and let $\pi \in \Delta^\Sigma$ be a relabeling. We first extend π to a mapping $\pi' \in (\Delta \cup X)^{\Sigma \cup X}$, in which we treat the elements of X as nullary symbols, for every $\sigma \in \Sigma$ and $x \in X$ by $\pi'_\sigma = \pi_\sigma$ and $\pi'_x = x$. Let $G' = (Q, \Delta, F, P', \text{wt}')$ be the eq" restricted positive WTGc such that

$$P' = \left\{ \pi'(c)[q_1, \ldots, q_k] \xrightarrow{E, \emptyset} q \mid c[q_1, \ldots, q_k] \xrightarrow{E, \emptyset} q \in P, \ q \neq \perp \right\}$$
$$\cup \left\{ \delta(\perp, \ldots, \perp) \to \perp \mid \delta \in \Delta \right\}$$

and for every production $p' = c'[q_1, \ldots, q_k] \xrightarrow{E, \emptyset} q \in P'$ with $q \neq \perp$ we let

$$\text{wt}'_{p'} = \sum_{\substack{p = c[q_1, \ldots, q_k] \xrightarrow{E, \emptyset} q \in P \\ c \in (\pi')^{-1}(c')}} \text{wt}_p \ . \tag{3}$$

Finally, $\text{wt}'\big(\delta(\perp, \ldots, \perp) \to \perp\big) = 1$ for all $\delta \in \Delta$. Since the relabeling replaces symbols by symbols, the size of G' is linear in the size of G. For the weight function wt', we must compute the sum (3), which can be achieved by accumulating the weight during the construction of the new productions, so the overall time complexity remains linear. For correctness we prove the following equality for every $u \in T_\Delta$ and $q \in Q$ by induction on u

$$\text{wt}^q_{G'}(u) = \begin{cases} \sum_{t \in \pi^{-1}(u)} \text{wt}^q_G(t) & \text{if } q \neq \perp \\ 1 & \text{otherwise.} \end{cases} \tag{4}$$

The second case is immediate since there is a single derivation, namely the one utilizing only nonterminal $\perp$, for u to $\perp$ and its weight is 1. In the remaining case we have $q \neq \perp$. Then

$$\text{wt}^q_{G'}(u)$$

$$\overset{(1)}{=} \sum_{\substack{p' = c'[q_1, \ldots, q_k] \xrightarrow{E, \emptyset} q \in P' \\ u_1, \ldots, u_k \in T_\Delta \\ u = c'[u_1, \ldots, u_k] \\ u \models E}} \text{wt}'_{p'} \cdot \prod_{i=1}^k \text{wt}^{q_i}_{G'}(u_i)$$

$$\overset{\text{IH}}{=} \sum_{\substack{p' = c'[q_1, \ldots, q_k] \xrightarrow{E, \emptyset} q \in P' \\ u_1, \ldots, u_k \in T_\Delta \\ u = c'[u_1, \ldots, u_k] \\ u \models E}} \text{wt}'_{p'} \cdot \prod_{\substack{i \in [k] \\ q_i \neq \perp}} \left(\sum_{t_i \in \pi^{-1}(u_i)} \text{wt}^{q_i}_G(t_i) \right) \cdot \prod_{\substack{i \in [k] \\ q_i = \perp}} 1 \ .$$

Recall that $g_p \colon [k] \to [k]$ assigns to each index its governing index. For better readability, we write just g'. Note that due to the special form of substitution we automatically fulfill $u \models E$ and can thus drop it.

$$\overset{(3)}{=} \sum_{\substack{p'=c'[q_1,\ldots,q_k] \xrightarrow{E,\emptyset} q \in P' \\ \forall i \in \mathrm{ran}(g') \colon u_i \in T_\Delta, t_i \in \pi^{-1}(u_i) \\ u=c'[u_{g'(1)},\ldots,u_{g'(k)}]}} \left(\sum_{\substack{p=c[q_1,\ldots,q_k] \xrightarrow{E,\emptyset} q \in P \\ c \in (\pi')^{-1}(c')}} \mathrm{wt}_p \right) \cdot \prod_{i \in \mathrm{ran}(g')} \mathrm{wt}_G^{q_i}(t_i)$$

We note that $g_{p'} = g_p$ for all used productions p, so we just write g. Additionally, for every q_i with $i \in [k] \setminus \mathrm{ran}(g)$ we have $q_i = \bot$ and thus $\mathrm{wt}_G^{q_i}(t_{g(i)}) = 1$ because there is exactly one such derivation with weight 1.

$$= \sum_{\substack{p=c[q_1,\ldots,q_k] \xrightarrow{E,\emptyset} q \in P \\ \forall i \in \mathrm{ran}(g) \colon t_i \in T_\Sigma \\ u=\pi(c[t_{g(1)},\ldots,t_{g(k)}])}} \mathrm{wt}_p \cdot \prod_{i=1}^{k} \mathrm{wt}_G^{q_i}(t_{g(i)})$$

$$= \sum_{t \in \pi^{-1}(u)} \left(\sum_{\substack{p=c[q_1,\ldots,q_k] \xrightarrow{E,\emptyset} q \in P \\ t_1,\ldots,t_k \in T_\Sigma \\ t=c[t_1,\ldots,t_k] \\ t \models E}} \mathrm{wt}_p \cdot \prod_{i=1}^{k} \mathrm{wt}_G^{q_i}(t_i) \right) \overset{(1)}{=} \sum_{t \in \pi^{-1}(u)} \mathrm{wt}_G^{q}(t)$$

We complete the proof for every $u \in T_\Delta$ as follows.

$$G'_u = \sum_{q \in Q} F_q \cdot \mathrm{wt}_{G'}^{q}(u) \overset{(4)}{=} \sum_{q \in Q \setminus \{\bot\}} F_q \cdot \left(\sum_{t \in \pi^{-1}(u)} \mathrm{wt}_G^{q}(t) \right) = \sum_{t \in \pi^{-1}(u)} \left(\sum_{q \in Q} F_q \cdot \mathrm{wt}_G^{q}(t) \right)$$

$$= \sum_{t \in \pi^{-1}(u)} G_t \hspace{6cm} \square$$

5 Towards the HOM Problem

The strategy of [15] for deciding the HOM problem first represents the homomorphic image $L' = h(L)$ of the regular tree language L with the help of an WTGc G'. For deciding whether L' is regular, a tree automaton G'' simulating the behavior of G' up to a certain bounded height is constructed. If the automata G' and G'' are equivalent, i.e., $G'' = G'$, then L' is regular. In the remaining case, pumping arguments are used to prove that it is impossible to find any TA for L'. Overall, this reduces the HOM problem to an equivalence problem.

Towards solving the HOM problem in the weighted case we now proceed similarly. First, we show that WTGc can encode each (well-defined) homomorphic image of a regular weighted tree language. This ability motivated their definition in the unweighted case [15, Proposition 4.6], and it also applies in the weighted case with minor restrictions that just enforce that all obtained sums are finite.

Theorem 5 *Let $G = (Q, \Sigma, F, P, \mathrm{wt})$ be a WTA and $h \in T_\Delta^{T_\Sigma}$ be a nondeleting and nonerasing tree homomorphism. There exists an eq" restricted positive WTGc G' with $G' = h(G)$.*

Proof We recall that we also use h for the mapping $h \in T_\Delta(X)^\Sigma$ inducing the tree homomorphism h. We construct a WTGc G' for $h(G)$ in two stages. First, let

$$G'' = \left(Q \cup \{\bot\}, \Delta \cup \Delta \times P, F'', P'', \mathrm{wt}''\right)$$

such that for every $p = \sigma(q_1, \ldots, q_k) \to q \in P$ and $h_\sigma = u = \delta(u_1, \ldots, u_n)$,

$$p'' = \left(\langle \delta, p \rangle(u_1, \ldots, u_n)[\![q_1, \ldots, q_k]\!] \xrightarrow{E, \emptyset} q\right) \in P''$$

with $E = \bigcup_{i \in [k]} \mathrm{pos}_{x_i}(u)^2$, in which the substitution $\langle \delta, p \rangle(u_1, \ldots, u_n)[\![q_1, \ldots, q_k]\!]$ replaces for every $i \in [k]$ only the left-most occurrence of x_i in $\langle \delta, p \rangle(u_1, \ldots, u_n)$ by q_i and all other occurrences by $\bot$. Moreover $\mathrm{wt}''_{p''} = \mathrm{wt}_p$. Additionally, we let

$$p''_\delta = \delta(\bot, \ldots, \bot) \to \bot \in P''$$

with weight $\mathrm{wt}''_{p''_\delta} = 1$ for every $k \in \mathbb{N}$ and $\delta \in \Delta_k \cup \Delta_k \times P$. No other productions are in P''. Finally, we let $F''_q = F_q$ for all $q \in Q$ and $F''_\bot = 0$. Obviously, G'' is eq" restricted and positive.

In order to better describe the behaviour of G'', let us introduce the following notation. Given a tree $t = \sigma(t_1, \ldots, t_k) \in T_\Sigma$ and a complete left-most derivation $d = (p_1, w_1) \cdots (p_m, w_m)$ of G for t, let $d_1, \ldots, d_k$ be the derivations for $t_1, \ldots, t_k$, respectively that are incorporated in d and $h_\sigma = \delta(u_1, \ldots, u_n)$. Then we define the tree $h(t, d) \in T_{\Delta \cup \Delta \times P}$ inductively by

$$h(t, d) = \langle \delta, p_m \rangle(u_1, \ldots, u_n)\big[h(t_1, d_1), \ldots, h(t_k, d_k)\big].$$

Using this notation, let us now prove that for each $q \in Q$ we have

$$\left\{s \in T_{\Delta \cup \Delta \times P} \mid D^q_{G''}(s) \neq \emptyset\right\} = \left\{h(t, d) \mid t \in T_\Sigma, d \in D^q_G(t)\right\} \tag{5}$$

and, in turn, every such $D^q_{G''}(s)$ is a singleton set with $\mathrm{wt}_{G''}(d'') = \mathrm{wt}_G(d)$ for the unique $d'' \in D^q_{G''}\big(h(t, d)\big)$.

We start with the inclusion from right to left. To this end, let $t \in T_\Sigma$ be a tree and $d = (p_1, w_1) \cdots (p_m, w_m)$ be a complete left-most derivation of G for t to some nonterminal $q \in Q$. Let $t = \sigma(t_1, \ldots, t_k)$ be the input tree with $h_\sigma = \delta(u_1, \ldots, u_n)$, let $p_m = \sigma(q_1, \ldots, q_k) \to q$ be the production utilized last in d, and let d_i be the complete left-most derivation for t_i to q_i incorporated in d for every $i \in [k]$. For every $i \in [k]$, we utilize the induction hypothesis to conclude that $D^{q_i}_{G''}\big(h(t_i, d_i)\big)$ is a

singleton set, so let $d_i'' \in D_{G''}^{q_i}\big(h(t_i, d_i)\big)$ be the unique element, for which we additionally have $\mathrm{wt}_{G''}(d_i'') = \mathrm{wt}_G(d_i)$. Moreover, for every $i \in [k]$ there is a derivation $d_i^\perp$ for $h(t_i, d_i)$ with weight 1 that exclusively utilizes the nonterminal $\perp$. We define

$$ s = \langle \delta, p_m \rangle(u_1, \ldots, u_n)\big[h(t_1, d_1), \ldots, h(t_k, d_k)\big] \ . $$

For every $i \in [k]$, let v_i be the left-most position labeled by x_i in h_σ. We consider the derivations $v_1 d_1'', \ldots, v_k d_k''$, and for every other occurrence v of x_i in h_σ we consider the derivation $v d_i^\perp$. Let d'' be the derivation assembled from the considered subderivations followed by (p_m'', ε), where $p_m'' = \langle \delta, p_m \rangle(u_1, \ldots, u_n)[\![q_1, \ldots, q_k]\!] \xrightarrow{E, \emptyset} q$ with the constraints $E = \bigcup_{i=1}^k \mathrm{pos}_{x_i}(h_\sigma)^2$. Clearly, the production p_m'' is the only applicable one since the only other production whose left-hand side is labeled by $\langle \delta, p_m \rangle$ at the root reaches $\perp \neq q$. Reordering the derivation d'' to be left-most, we obtain the desired complete left-most derivation $\underline{d}''$ for s, for which we also have $\mathrm{wt}_{G''}(\underline{d}'') = \mathrm{wt}_G(d)$. This proves that $\underline{d}''$ is the required single element of $D_{G''}^q(s) = D_{G''}^q\big(h(t, d)\big) \neq \emptyset$.

On the other hand, consider $s \in T_{\Delta \cup \Delta \times P}$ such that there exists a complete left-most derivation $d'' = (p_1'', w_1'') \cdots (p_m'', w_m'')$ for s to q; i.e. $d'' \in D_{G''}^q(s) \neq \emptyset$. The final rule p_m'' that is applied must be of the form

$$ p_m'' = \langle \delta, p \rangle(u_1, \ldots, u_n)[\![q_1, \ldots, q_k]\!] \xrightarrow{E, \emptyset} q $$

with $\delta(u_1, \ldots, u_n)[\![q_1, \ldots, q_k]\!] = h_\sigma[\![q_1, \ldots, q_k]\!]$ for some symbol $\sigma \in \Sigma_k$ and production $p = \sigma(q_1, \ldots, q_k) \to q$. For every $i \in [k]$, we denote by w_i the unique position among $\mathrm{pos}_{x_i}(h_\sigma)$ that is labeled by q_i in $h_\sigma[\![q_1, \ldots, q_k]\!]$ (recall that every other position related to w_i via E will be labeled by $\perp$). By the induction hypothesis applied to $s|_{w_i}$, for which the complete left-most derivation d_i'' for $s|_{w_i}$ to q_i incorporated in d'' exists, there exists a tree $t_i \in T_\Sigma$ and a complete left-most derivation d_i of G for t_i to q_i such that $s|_{w_i} = h(t_i, d_i)$ and $\mathrm{wt}_G(d_i) = \mathrm{wt}_{G''}(d_i'')$. For the tree $t = \sigma(t_1, \ldots, t_k)$ we obtain that $s = h(t, d)$ for the complete left-most derivation $d \in D_G^q(t)$ given by

$$ d = (1 d_1) \cdots (k d_k)(p, \varepsilon) \ , $$

for which we also have $\mathrm{wt}_G(d) = \mathrm{wt}_{G''}(d'')$, which completes this proof.

So far, Q'' and P'' are larger than Q and P only by a constant (assuming a fixed alphabet Σ) caused by the additional sink nonterminal $\perp$ and its productions, but the alphabet size increases by the summand $|\Delta| \cdot |P|$. Computing a single production only takes linear time in the size of h (assuming constant-time access to the tree in the range of h), so G'' is constructed in linear time in $|G| \cdot |h|$.

We now delete the annotation with the help of the relabeling $\pi \in \Delta^{\Delta \cup \Delta \times P}$ given for every $\delta \in \Delta$ and $p \in P$ by $\pi_\delta = \pi_{\langle \delta, p \rangle} = \delta$ following the construction in Theorem 4.

$$ \pi(G'')_u = \sum_{s \in \pi^{-1}(u)} G_s'' = \sum_{s \in \pi^{-1}(u)} \left(\sum_{q \in Q} F_q'' \cdot \mathrm{wt}_{G''}^q(s) \right) = \sum_{\substack{q \in Q, s \in \pi^{-1}(u) \\ d'' \in D_{G''}^q(s)}} F_q'' \cdot \mathrm{wt}_{G''}(d'') $$

$$(5) = \sum_{\substack{q \in Q, \, s \in \pi^{-1}(u) \\ t \in T_\Sigma, \, d \in D_G^q(t) \\ s = h(t,d)}} F_q \cdot \mathrm{wt}_G(d) = \sum_{\substack{q \in Q \\ t \in h^{-1}(u)}} F_q \cdot \mathrm{wt}_G^q(t) = \sum_{t \in h^{-1}(u)} G_t = h(G)_u$$

for every $u \in T_\Delta$. The overall time complexity is thus still in $\mathcal{O}(|G| \cdot |h|)$. The construction of Theorem 4 is applicable because $\bot$ is clearly a sink nonterminal in G'' and G'' is an eq" restricted positive WTGc. $\square$

Let us illustrate the construction on a simple example.

Example 5 Consider the WTA $G = (\{q, q'\}, \Sigma, F, P, \mathrm{wt})$ over the semiring $\mathbb{N}$ of nonnegative integers with $\Sigma = \{\alpha^{(0)}, \phi^{(1)}, \gamma^{(1)}, \varepsilon^{(1)}\}$, $F_q = 0$, $F_{q'} = 1$, and the set of productions and their weights given by

$$p_1 = \alpha \to_1 q \qquad p_2 = \gamma(q) \to_2 q \qquad p_3 = \varepsilon(q) \to_1 q \quad \text{and} \quad p_4 = \phi(q) \to_1 q' .$$

Then $\mathrm{supp}(G) = \{\phi(t) \mid t \in T_{\Sigma \setminus \{\phi\}}\}$ and $G_t = 2^{|\mathrm{pos}_\gamma(t)|}$ for every $t \in \mathrm{supp}(G)$. Consider the ranked alphabet $\Delta = \{\alpha^{(0)}, \gamma^{(1)}, \sigma^{(2)}\}$ and the tree homomorphism h induced by $h_\alpha = \alpha$, $h_\gamma = h_\epsilon = \gamma(x_1)$, and $h_\phi = \sigma(\gamma(x_1), x_1)$. Consequently,

$$\mathrm{supp}(h(G)) = \{\sigma(\gamma^{n+1}(\alpha), \gamma^n(\alpha)) \mid n \in \mathbb{N}\}$$

and $h(G)_t = \sum_{k=0}^n \binom{n}{k} 2^k = 3^n$ for every $t = \sigma(\gamma^{n+1}(\alpha), \gamma^n(\alpha)) \in \mathrm{supp}(h(G))$. A WTGc for $h(G)$ is constructed as follows. First, we let

$$G'' = (\{q, q', \bot\}, \Delta \cup \Delta \times P, F'', P'', \mathrm{wt}'')$$

with $F''_{q'} = 1$, $F''_q = F''_\bot = 0$ and the productions and their weights are given by

$$\langle \alpha, p_1 \rangle \to_1 q \quad \langle \gamma, p_2 \rangle(q) \to_2 q \quad \langle \gamma, p_3 \rangle(q) \to_1 q \quad \langle \sigma, p_4 \rangle(\gamma(q), \bot) \xrightarrow{1 1 = 2}_1 q'$$

and $\delta(\bot, \ldots, \bot) \to_1 \bot$ for all $\delta \in \Delta \cup \Delta \times P$. Next we remove the second component of the symbols of $\Delta \times P$ and add the weights of all productions that yield the same production once the second components are removed. In our example, this applies to the production $\gamma(q) \to q$, which is the result of the two productions $\langle \gamma, p_2 \rangle(q) \to_2 q$ and $\langle \gamma, p_3 \rangle(q) \to_1 q$, so its weight is $2 + 1 = 3$. Overall, we obtain the WTGc $G' = (\{q, q', \bot\}, \Delta, F'', P', \mathrm{wt}')$ with the following productions for all $\delta \in \Delta$:

$$\alpha \to_1 q \qquad \gamma(q) \to_3 q \qquad \sigma(\gamma(q), \bot) \xrightarrow{1 1 = 2}_1 q' \qquad \delta(\bot, \ldots, \bot) \to_1 \bot . \qquad \square$$

Trees generated by a WTGc must satisfy certain equality constraints on their subtrees. Therefore, if we naively swap subtrees of generated trees, then we might violate such an equality constraint and obtain a tree that is no longer generated by the WTGc. Luckily, the particular kind of WTGc constructed in Theorem 5, namely eq" restricted

positive WTGc, allows us to refine the subtree substitution such that it takes into consideration the equality constraints in force. The following definition is the natural adaptation of [15, Definition 5.1] for (Boolean) tree automata with constraints.

Definition 6 Let $G = (Q, \Sigma, F, P, \text{wt})$ be an eq" restricted positive WTGc with sink nonterminal $\perp$. Moreover, let $q, q' \in Q$, $t, t' \in T_\Sigma$, and $d \in D_G^q(t)$ as well as $d' \in D_G^{q'}(t')$ such that $q \neq \perp \neq q'$ and $d = \underline{d}(p, \varepsilon)$ uses $p = c[q_1, \ldots, q_k] \xrightarrow{E, \emptyset} q \in P$ as its final production. For every $i \in [k]$ let $w_i = \text{pos}_{x_i}(c)$ and d_i be the unique left-most derivation for $t_i = t|_{\text{pos}_{x_i}(c)}$ incorporated in d. Finally, for every tree $u \in T_\Sigma$ let $d_u^\perp$ be the unique left-most derivation for u to $\perp$. For every $w \in \text{pos}(t)$, for which the derivation for $t|_w$ incorporated in d yields q', we recursively define the derivation substitution $d[\![d']\!]_w$ of d' into d at w and the resulting tree $t[\![t']\!]_w^d$ as follows. If $w = \varepsilon$, then $d[\![d']\!]_\varepsilon = d'$ and $t[\![t']\!]_\varepsilon^d = t'$. Otherwise $w = w_j \underline{w}$ for some $j \in [k]$ and we have

$$d[\![d']\!]_w = d_1' \cdots d_k'(p, \varepsilon) \qquad \text{and} \qquad t[\![t']\!]_w^d = c[t_1', \ldots, t_k'] \ ,$$

where for each $i \in [k]$ we have

- if $i = j$ (i.e., w_i is a prefix of w), then $d_i' = w_i(d_i[\![d']\!]_{\underline{w}})$ and $t_i' = t_i[\![t']\!]_{\underline{w}}^{d_i'}$,
- if $q_i = \perp$ and $w_i \in [w_j]_{\equiv_E}$ (i.e., it is a position that is equality restricted to w_j), then $d_i' = w_i d_u^\perp$ and $t_i' = u$ with $u = t_j[\![t']\!]_{\underline{w}}^{d_j'}$, and
- otherwise $d_i' = w_i d_i$ and $t_i' = t_i$ (i.e., derivation and tree remain unchanged).

It is straightforward to verify that $d[\![d']\!]_w$ is a complete left-most derivation of G for $t[\![t']\!]_w^d$ to q. □

Example 6 We consider the WTGc $G = (\{q, \perp\}, \Sigma, F, P, \text{wt})$ with input ranked alphabet $\Sigma = \{a^{(0)}, g^{(2)}, f^{(2)}\}$, final weights $F_q = 1$ and $F_\perp = 0$ as well as productions

$$p_a = a \rightarrow_1 q \qquad p_g = g(q, \perp) \xrightarrow{1=2}_1 q \quad \text{and} \quad p_f = f(q, f(q, \perp)) \xrightarrow{1=22}_1 q$$

besides the sink nonterminal productions $p_\sigma^\perp = \sigma(\perp, \ldots, \perp) \rightarrow_1 \perp$ for all $\sigma \in \Sigma$. As before, for every $u \in T_\Sigma$ we let $d_u^\perp \in D_G^\perp(u)$ be the unique derivation of G for u to $\perp$, which utilizes only the nonterminal $\perp$. According to Definition 6 we choose the states $q = q'$ and the trees t and t' and derivations d and d' as given in Fig. 2 and below.

$$d = (p_a, 11) (p_a^\perp, 12) (p_g, 1) (p_a, 21) (p_a^\perp, 221) (p_a^\perp, 222) (p_g^\perp, 22) (p_f, \varepsilon)$$
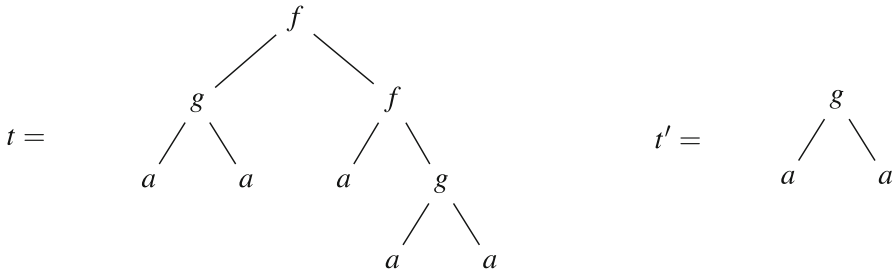$$d' = (p_a, 1) (p_a^\perp, 2) (p_g, \varepsilon)$$

Fig. 2 Input trees t and t' from Example 6

We select that position $w = 11$ and observe that that the derivation for $t|_{11}$ is (p_a, ε), which yields $q = q'$. We compute $d[\![d']\!]_w$ as follows

$$
\begin{aligned}
d[\![d']\!]_{11} &= \Big(1(d'_1[\![d']\!]_1)\Big)\Big(21(p_a, \varepsilon)\Big)\Big(22d_u^\perp\Big)(p_f, \varepsilon) \\
&= \Big(1\Big(1d'\Big)\Big(2d_{g(a,a)}^\perp\Big)(p_g, \varepsilon)\Big)(p_a, 21)(22d_u^\perp)(p_f, \varepsilon) \\
&= (p_a, 111)(p_a^\perp, 112)(p_g, 11)(12d_{g(a,a)}^\perp)(p_g, 1)(p_a, 21)(22d_u^\perp)(p_f, \varepsilon) ,
\end{aligned}
$$

where $d'_1 = (p_a, 1)(p_a^\perp, 2)(p_g, \varepsilon)$ and $u = g\big(g(a, a), g(a, a)\big)$. We note that $w = 11$ is explicitly equality constrained to position 12 in d via the constraint $1 = 2$ at position 1 and implicitly equality constrained to positions 221 and 222 via the constraint $1 = 22$ at the root ε. Thus, we obtain $d[\![d']\!]_{11}$ by substituting d' into d at position 11 as well as substituting $d_{t'}^\perp$ into d at positions 12, 221, and 222. The obtained tree $t[\![t']\!]_w^d$ is displayed in Fig. 3. $\qquad\square$

As our example illustrates, the tree $t[\![t']\!]_w^d$ is obtained from t by (i) identifying the set of all positions of t that are explicitly or implicitly equality constrained to w by the productions in the derivation d and (ii) substituting t' into t at every such position. If $w' \in \mathrm{pos}(t)$ is parallel to all positions constrained to w, like position 21 in Example 6, then $t[\![t']\!]_w|_{w'} = t|_{w'}$. Note that $t|_{21}$ is equal to the replaced subtree $t|_{11}$, but we only replace constrained subtrees and not all equal subtrees.
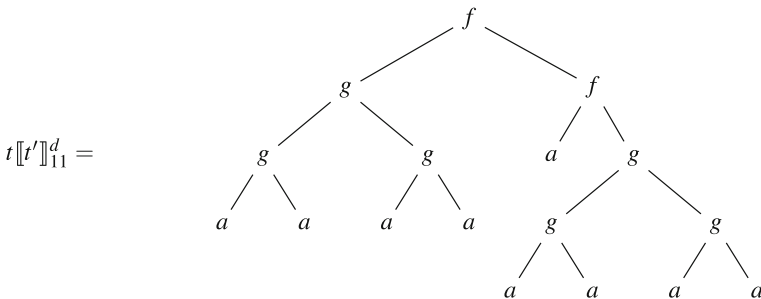

Fig. 3 Obtained pumped tree $t[\![t']\!]_{11}^d$ from Example 6

This substitution allows us to prove a pumping lemma for eq" restricted positive WTGc, which can generate all (nondeleting and nonerasing) homomorphic images of regular weighted tree languages by Theorem 5. To this end, we need some final notions. Let $G = (Q, \Sigma, F, P, \mathrm{wt})$ be a WTGc. Moreover, let $p = \ell \xrightarrow{E,I} q \in P$ be a production. We define the *height* $\mathrm{ht}(p)$ of p by $\mathrm{ht}(p) = \mathrm{ht}(\ell)$ (i.e., the height of its left-hand side). Moreover, we let

$$\mathrm{ht}(P) = \max\big\{\mathrm{ht}(p) \mid p \in P\big\} \quad \text{and} \quad \mathrm{ht}(G) = (|Q| + 1) \cdot \mathrm{ht}(P) \;.$$

Lemma 4 *Let $G = (Q, \Sigma, F, P, \mathrm{wt})$ be an eq" restricted positive WTGc with sink nonterminal $\bot$. There exists $n \in \mathbb{N}$ s.t. for every tree $t_0 \in T_\Sigma$, nonterminal $q \in Q \setminus \{\bot\}$, and derivation $d \in D_G^q(t_0)$ such that $\mathrm{ht}(t_0) > n$ and $\mathrm{wt}_G(d) \neq 0$ there are infinitely many trees $t_1, t_2, \ldots$ and derivations $d_1, d_2, \ldots$ such that $d_i \in D_G^q(t_i)$ and $\mathrm{wt}_G(d_i) \neq 0$ for all $i \in \mathbb{N}$.*

Proof Without loss of generality, suppose that for every $c[q_1, \ldots, q_k] \xrightarrow{E,\emptyset} q' \in P$ with $q' \neq \bot$ and $k \neq 0$ there exists $i \in [k]$ such that $q_i \neq \bot$. This can easily be achieved by introducing a copy $\top$ of nonterminal $\bot$ and replacing one instance of $\bot$ by $\top$ in offending productions. Similarly, we can assume without loss of generality that the construction in the proof of Lemma 3 has been applied to G. If this is the case, then we can select $n = \mathrm{ht}(G)$. Let $t_0 \in T_\Sigma$ be such that $\mathrm{ht}(t_0) > n$. Let $Q' = Q \setminus \{\bot\}$, $d \in D_G^q(t_0)$ be a derivation with $\mathrm{wt}_G(d) \neq 0$, and select a position $w \in \mathrm{pos}(t_0)$ of maximal length such that d incorporates a derivation for $t_0|_w$ to some $q' \in Q'$. Then

$$|w| \geq \mathrm{ht}(t_0) - \mathrm{ht}(P) \geq \mathrm{ht}(G) - \mathrm{ht}(P) = |Q| \cdot \mathrm{ht}(P) \;,$$

which yields that at least $|Q|$ proper prefixes w' of w exist such that d incorporates a derivation for $t_0|_{w'}$ to some $q' \in Q'$. Hence there exist prefixes w', w'' of w such that d incorporates a derivation d' for $t' = t_0|_{w'}$ to $q' \in Q'$ as well as a derivation for $t_0|_{w''}$ to the same nonterminal q', and, without loss of generality, w' is a proper prefix of w''. Then $d[\![d']\!]_{w''}$ is a derivation of G for $t_1 = t[\![t']\!]_{w''}^d$ to q with $\mathrm{ht}(t_1) > \mathrm{ht}(t_0)$. Since we achieve the same state q, the annotation of the proof of Lemma 3 guarantees that $\mathrm{wt}_G(d_1) \neq 0$. Iterating this substitution yields the desired trees $t_1, t_2, \ldots$ and derivations $d_1, d_2, \ldots$. $\qquad\square$

A WTGc generating a (nondeleting and nonerasing) homomorphic image of a regular weighted tree language, if constructed as described in Theorem 5, will never have overlapping constraints since constraints always point to leaves of the left-hand sides of productions as required by classic WTGc. It is intuitive that this limitation to the operating range of constraints leads to an actual restriction in the expressive power of WTGc, but we will only prove it for eq-restricted positive WTGc.

Proposition 2 *Let $\mathbb{S}$ be a zero-sum free semiring. The class of positive constraint-regular weighted tree languages is strictly more expressive than the class of weighted tree languages generated by eq" restricted positive WTGc.*

Proof Let us consider the positive WTGc $G = (\{q, q'\}, \Sigma, F, P, \mathrm{wt})$ with input ranked alphabet $\Sigma = \{f^{(2)}, \underline{f}^{(2)}, g^{(2)}, a^{(0)}\}$, final weights $F_q = 1$ and $F_{q'} = 0$, and the following productions, of which each has weight 1.

$$a \to_1 q' \qquad g(q', q') \to_1 q \qquad f(q, q) \xrightarrow{12=21}_1 q \qquad \underline{f}(q, q) \xrightarrow{12=21}_1 q$$

The first two productions are only used on leaves and on subtrees of the form $g(a, a)$. Every other position w (i.e., neither leaf nor position with two leaves as children) is labeled either f or $\underline{f}$ and additionally every derivation enforces the constraint $12 = 21$, so the subtrees $t|_{w12}$ and $t|_{w21}$ of the input tree t need to be equal for a complete derivation of G to exist.

For the sake of a contradiction, suppose that there exists an eq" restricted positive WTGc $G' = (Q', \Sigma, F', P', \mathrm{wt}')$ that is equivalent to G. We recursively define the trees $t_n \in T_\Sigma$ and $t'_n \in T_\Sigma$ for every $n \in \mathbb{N}$ with $n \geq 1$ by

$$
\begin{array}{lll}
t_0 = a & t_1 = g(t_0, t_0) & t_{n+1} = f(t_n, t_n) \\
t'_0 = a & t'_1 = g(t'_0, t_0) & t'_{n+1} = \underline{f}(t'_n, t_n)
\end{array}
$$

Clearly, t_n and t'_n are both complete binary trees of height n. Naturally, the leaves are labeled a, and the penultimate level in both trees is always labeled g. In t_n the remaining levels are universally labeled f, whereas in t'_n the left-most spine on those levels is labeled $\underline{f}$. We illustrate an example tree t'_n in Fig. 4. Obviously $G(t_n) = 1$ as well as $G(t'_n) = 1$ for every $n \in \mathbb{N}$ with $n \geq 1$. Furthermore we note that the derivations of G only enforce equality constraints on positions of the form $w12$ or $w21$, but since $\mathrm{pos}_{\underline{f}}(t'_n) \subseteq \{1\}^*$, the positions, in which the labels in t_n and t'_n differ, are not affected by any equality constraint. This can be used to verify that $G(t'_n) = 1$ for each $n \geq 1$.

In the following, let $n = 3\,\mathrm{ht}(G') + 2$. Since G' is equivalent to G, we need to have $G'(t'_n) = 1$ as well, which requires a complete derivation of G' for t'_n to some final nonterminal $q_0 \in Q'$. Let $d \in D_{G'}^{q_0}(t'_n)$ be such a derivation. Moreover, let $d = \underline{d}(p, \varepsilon)$
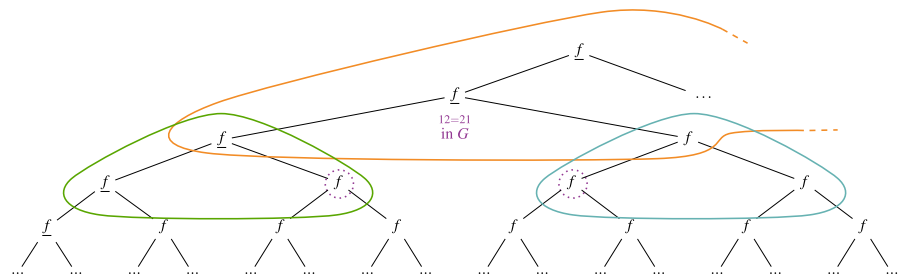

Fig. 4 A snippet of the tree t'_n and the productions used by G'

for some production $p = c[q_1, \ldots, q_k] \xrightarrow{E, \emptyset} q_0 \in P'$. Since the input tree t'_n contains positions

$$\left\{ 1^i = \underbrace{11 \cdots 1}_{i \text{ times}} \mid 0 \le i \le n \right\} \subseteq \mathrm{pos}(t'_n) \ ,$$

there must exist $j \in \mathbb{N}$ such that $c(1^j) = x_1$; i.e., position 1^j is labeled x_1 in c. Obviously, $j \le \mathrm{ht}(G')$, so the height of the subtree $t'' = t'_n|_{1^j}$, which is still a complete binary tree, is at least $2\,\mathrm{ht}(G') + 2$. We can thus apply Lemma 4 to the tree t'' in such a way that it modifies its second direct subtree (starting from $1^j \in \mathrm{pos}(t'_n)$, we descend to $1^j 2$; from there, we either find a subderivation to some nonterminal different from $\bot$, or all subtrees below $1^j 2$ are copies of subtrees below $1^j 1$, and in that case, we apply the pumping to an equality constrained subtree below $1^j 1$, which then also modifies the corresponding subtree below $1^j 2$). Let u be the such obtained pumped tree, which according to zero-sum freeness and Lemma 4 is also in the support of G'; i.e., $u \in \mathrm{supp}(G')$. Let d' be the derivation constructed in Lemma 4 corresponding to u. We have $u(1^{j-1}) = f$, so the position 1^{j-1} is labeled f. Since G and G' are equivalent, there must be a derivation of G for u as well, which enforces the equality constraint $u|_{1^{j-1}12} = u|_{1^{j-1}21}$. By construction we have $t'_n|_{1^{j-1}12} \ne u|_{1^{j-1}12}$. Since the positions $1^{j-1}12$ and $1^{j-1}21$ have no common suffix, this equality can only be guaranteed by G' if $1^{j-1}12$ and $1^{j-1}21$ are themselves (explicitly or implicitly) equality constrained in d'. The potentially several constraints that achieve this must of course be located at prefixes of $1^{j-1}12$ and $1^{j-1}21$, and since the production used in d' at the root is still p and stretches all the way to 1^j, this can only be achieved if d' enforces $1^{j-1}1 = 1^{j-1}2$ via p at the root as well as $1 = 2$ at $1^{j-1}1$ or at $1^{j-1}2$. However, this is a contradiction as $u(1^{j-1}1) = f \ne f = u(1^{j-1}2)$, so we cannot have an explicit or implicit equality constraint between $1^{j-1}12$ and $1^{j-1}21$, so $u|_{1^{j-1}21} = t'_n|_{1^{j-1}21}$, but contradicts that G has a complete derivation for u. $\square$

Although for zero-sum free semirings, the support of a regular weighted tree language is again regular, in general, the converse is not true, so we cannot apply the decision procedure of [15] to the support of a homomorphic image in order to decide its regularity. Instead, we hope to extend the unweighted argument in a way that tracks the weights sufficiently close. For this, we prepare two decidability results, which rely mostly on the corresponding results in the unweighted case. To this end, we need to relate our WTGc constructed in Theorem 5 to the classic TGc used in [15]. At this point we mention that their classic TGc additionally require that equality constrained positions have the same nonterminal label. Compared to our eq-restriction this change is entirely immaterial in the unweighted case.

Theorem 6 *Let $\mathbb{S}$ be a zero-sum free semiring. Moreover, let $G = (Q, \Sigma, F, P, \mathrm{wt})$ be a WTA and $h \in T_\Delta^{T_\Sigma}$ be a nondeleting and nonerasing tree homomorphism. Finally, let $G' = h(G)$. Emptiness and finiteness of $\mathrm{supp}(G')$ are decidable.*

Proof We apply the construction in the proof of Lemma 3 to the eq'' restricted positive WTGc $G' = (Q', \Sigma, F', P', \mathrm{wt}')$ constructed according to Theorem 5. Thus we ensure that all derivations have non-zero weight. Due to zero-sum freeness, we can simply drop the weights and obtain an eq'' restricted positive TGc $G'' =$

(Q'', Σ, F'', P'') generating supp(G'). Emptiness and finiteness are decidable for the tree language supp(G') generated by G'' according to [15, Corollaries 5.11 & 5.20].

□

Conclusion

The purpose of this contribution is to lay out the groundwork for investigating and eventually ideally deciding the weighted HOM problem. For this, we have introduced the model of eq-restricted WTGh and showed that they are well-suited to efficiently represent homomorphic images of regular weighted tree languages (Theorem 5). Apart from classical closure properties, we have proved a pumping lemma for these devices (Lemma 4).

Recently, significant progress has been made on the topic of the weighted HOM problem. Most notably, the $\mathbb{N}$-weighted version of this problem was proved to be decidable [21]. There, the HOM problem is reduced to a specific property of the WTGh constructed in our Theorem 5 for the homomorphic image, and this specific property is shown to be decidable. The proof of the latter part is based on our pumping Lemma 4. Additionally [23] shows that if the input of the HOM problem satisfies particular conditions (intuitively, the tree homomorphism must satisfy a condition generalizing injectivity and the input WTA must satisfy an ambiguity restriction), then the WTGh for the homomorphic image constructed in Theorem 5 is unambiguous. In that case, the (thus restricted) weighted HOM problem over any zero-sum free semiring can be reduced to the unweighted HOM problem [15] and is therefore decidable. Our current efforts are centered around the weighted HOM problem over fields, for which we hope to prove decidability with the same strategy that was used in [21] for $\mathbb{N}$-weights and a pumping lemma similar to Lemma 4 for zero-sum free semirings.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

References

1. Bogaert, B., Tison, S.: Equality and disequality constraints on direct subterms in tree automata. In: A. Finkel, M. Jantzen (eds.) Proc. 9th Ann. Symp. Theoretical Aspects of Computer Science, *LNCS*, vol. 577, pp. 161–171. Springer (1992). https://doi.org/10.1007/3-540-55210-3_181

2. Borchardt, B.: The theory of recognizable tree series. Ph.D. thesis, Technische Universität Dresden (2005)

3. Bozapalidis, S., Rahonis, G.: On the closure of recognizable tree series under tree homomorphisms. J. Autom. Lang. Comb. **10**(2–3), 185–202 (2005). https://doi.org/10.25596/jalc-2005-185

4. Comon, H., Dauchet, M., Gilleron, R., Löding, C., Jacquemard, F., Lugiez, D., Tison, S., Tommasi, M.: Tree automata —Techniques and applications. (2007) http://www.grappa.univ-lille3.fr/tata

5. Comon, H., Jacquemard, F.: Ground reducibility and automata with disequality constraints. In: P. Enjalbert, E.W. Mayr, K.W. Wagner (eds.) Proc. 11th Ann. Symp. Theoretical Aspects of Computer Science, *LNCS*, vol. 775, pp. 151–162, Springer (1994). https://doi.org/10.1007/3-540-57785-8_138

6. Dickson, L.E.: Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors. Amer. J. Math. **35**(4), 413–422 (1913). https://doi.org/10.2307/2370405

7. Doner, J.: Tree acceptors and some of their applications. J. Comput. System Sci. **4**(5), 406–451 (1970). https://doi.org/10.1016/S0022-0000(70)80041-1

8. Drewes, F.: Grammatical picture generation: A tree-based approach. In: Theoretical Computer Science — An EATCS Series, vol. 201, Springer (2006). https://doi.org/10.1007/3-540-32507-7

9. Droste, M., Heusel, D.: The supports of weighted unranked tree automata. Funda. Inform. **136**(1–2), 37–58 (2015). https://doi.org/10.3233/FI-2015-1143

10. Ésik, Z., Kuich, W.: Formal tree series. J. Autom. Lang. Comb. **8**(2), 219–285 (2003). https://doi.org/10.25596/jalc-2003-219

11. Fülöp, Z., Maletti, A., Vogler, H.: Preservation of recognizability for synchronous tree substitution grammars. In: F. Drewes, M. Kuhlmann (eds.) Proc. Workshop Applications of Tree Automata in Natural Language Processing, pp. 1–9, ACL(2010)

12. Fülöp, Z., Maletti, A., Vogler, H.: Weighted extended tree transducers. Fundam. Inform. **111**(2), 163–202 (2011). https://doi.org/10.3233/FI-2011-559

13. Fülöp, Z., Vogler, H.: Weighted tree automata and tree transducers. In: M. Droste, W. Kuich, H. Vogler (eds.) Handbook of Weighted Automata, Monographs in Theoretical Computer Science — An EATCS Series, chap. 9, pp. 313–403. Springer, (2009). https://doi.org/10.1007/978-3-642-01492-5_9

14. Gécseg, F., Steinby, M.: Tree automata. Tech. Rep. (2015) arXiv:1509.06233

15. Godoy, G., Giménez, O.: The HOM problem is decidable. J. ACM **60**(4), 1–44 (2013). https://doi.org/10.1145/2508028.2501600

16. Golan, J.S.: Semirings and their Applications. Kluwer Academic, Dordrecht (1999) https://doi.org/10.1007/978-94-015-9333-5

17. Hebisch, U., Weinert, H.J.: Semirings–Algebraic Theory and Applications in Computer Science, *Series in Algebra*, vol. 5. World Sci (1998) https://doi.org/10.1142/3903

18. Jurafsky, D., Martin, J.H.: Speech and language processing, 2nd edn. Prentice Hall (2008)

19. Kirsten, D.: The support of a recognizable series over a zero-sum free, commutative semiring is recognizable. Acta Cybernet. **20**(2), 211–221 (2011). https://doi.org/10.14232/actacyb.20.2.2011.1

20. Maletti, A., Nász, A.T.: Weighted tree automata with constraints. In: V. Diekert, M. Volkov (eds.) Proc. 26th Int. Conf. Developments in Language Theory, pp. 226–238. Springer (2022). https://doi.org/10.1007/978-3-031-05578-2_18

21. Maletti, A., Nász, A.T., Paul, E.: Weighted HOM-problem for nonnegative integers. Tech. Rep. (2023) arXiv:2305.04117

22. Mongy-Steen, J.: Transformation de noyaux reconnaissables d'arbres–forêts RATEG. Ph.D. thesis, Université de Lille (1981)

23. Nász, A.T.: Solving the weighted hom-problem with the help of unambiguity. In: Z. Gazdag, S. Iván (eds.) Proc. 16th Int. Conf. Automata and Formal Languages, EPTCS. Open Publishing Association (2023). To appear

24. Perrin, D.: Recent results on automata and infinite words. In: M.P. Chytil, V. Koubek (eds.) Proc. 11th Int. Symp. Mathematical Foundations of Computer Science, *LNCS*, vol. 176, pp. 134–148. Springer (1984). https://doi.org/10.1007/BFb0030294

25. Salomaa, A., Soittola, M.: Automata-theoretic aspects of formal power series. Springer (1978). https://doi.org/10.1007/978-1-4612-6264-0

26. Schützenberger, M.P.: On the definition of a family of automata. Inform. Control **4**(2–3), 245–270 (1961). https://doi.org/10.1016/S0019-9958(61)80020-X

27. Thatcher, J.W.: Characterizing derivation trees of context-free grammars through a generalization of finite automata theory. J. Comput. Syst. Sci. **1**(4), 317–322 (1967). https://doi.org/10.1016/S0022-0000(67)80022-9

28. Thatcher, J.W., Wright, J.B.: Generalized finite automata theory with an application to a decision problem of second-order logic. Math. Syst Theory **2**(1), 57–81 (1968). https://doi.org/10.1007/BF01691346

29. Tison, S.: Tree automata, (dis-)equality constraints and term rewriting: What's new? In: M. Schmidt-Schauß (ed.) Proc. 22nd Int. Conf. Rewriting Techniques and Applications, *LIPIcs*, vol. 10, pp. 1–2. Schloss Dagstuhl–Leibniz-Zentrum für Informatik (2011). https://doi.org/10.4230/LIPIcs.RTA.2011.1

30. Wang, H.: On characters of semirings. Houston J. Math. **23**(3), 391–405 (1997)

31. Wilhelm, R., Seidl, H., Hack, S.: Compiler Design-Syntactic and Semantic Analysis. Springer (2012). https://doi.org/10.1007/978-3-642-17548-0