



Weighted Tree Automata with Constraints

Andreas Maletti and Andreea-Teodora Nász^(✉)

Faculty of Mathematics and Computer Science, Universität Leipzig,
PO box 100 920, 04009 Leipzig, Germany
{maletti,nasz}@informatik.uni-leipzig.de

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, 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 [23]. This development led to the fruitful study of recognizable formal power series [22], 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 [21] and trees [4]. Finite-state tree automata were introduced independently in [7, 24, 25] 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 [28]. Indeed several applications require the combination of the two mentioned generalizations and a broad range of weighted tree automaton (WTA) models has been studied (cf. [13, Chapter 9] for an overview). It is

Research financially supported by a scholarship awarded to T. Nasz by the Free State of Saxony (Funding no. LAU-R-I-9-2-1021).

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 [20] aims to remedy this problem and introduces a tree automaton model that explicitly can require certain subtrees to be equal or unequal. Such models are very useful when investigating transduction models (see [13] for an overview) that can copy subtrees (thus resulting in equal subtrees) 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.

In the weighted setting, the HOM problem is also interesting as it once again provides an answer whether a given homomorphic image of a regular weighted tree language can efficiently be represented. 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 automata with constraints (WTAc for short) in this contribution. We demonstrate that those WTAc can again represent all homomorphic images of the regular weighted tree languages. Thus, in principle, it only remains to provide a decision procedure for determining whether a given WTAc 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].

2 Preliminaries

We denote the set of nonnegative integers by \mathbb{N} , and for every $k \in \mathbb{N}$, we let $[k] = \{i \in \mathbb{N} \mid 1 \leq i \leq k\}$. For all sets T and Z let T^Z be the set of all mappings $\varphi: Z \rightarrow T$, and correspondingly we sometimes write φ_z instead of $\varphi(z)$ for every $\varphi \in T^Z$. The cardinality of Z is denoted by $|Z|$.

A *ranked alphabet* (Σ, rk) is a pair consisting of a finite set Σ and a mapping $\text{rk} \in \mathbb{N}^\Sigma$ 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 $X = \{x_i \mid i \in \mathbb{N}\}$ be a countable set of (formal) variables. For each $n \in \mathbb{N}$ we let $X_n = \{x_i \mid i \in [n]\}$. Given a ranked alphabet Σ and a set Z , the set $T_\Sigma(Z)$ of Σ^n *trees indexed by Z* is the smallest set such that $Z \subseteq T_\Sigma(Z)$

and $\sigma(t_1, \dots, t_k) \in T_\Sigma(Z)$ for every $k \in \mathbb{N}$, $\sigma \in \Sigma_k$, and $t_1, \dots, t_k \in T_\Sigma(Z)$. 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 $\text{pos}(\sigma(t_1, \dots, t_k)) = \{\varepsilon\} \cup \bigcup_{i \in [k]} \{iw \mid w \in \text{pos}(t_i)\}$ for all $k \in \mathbb{N}$, $\sigma \in \Sigma_k$, and $t_1, \dots, t_k \in T_\Sigma(Z)$. The size $|t|$ of t is defined as $|t| = |\text{pos}(t)|$. 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 $t = \sigma(t_1, \dots, t_k)$ by $t(\varepsilon) = \sigma$, $t(iw') = t_i(w')$, $t|_\varepsilon = t$, $t|_{iw'} = t_i|_{w'}$, $t[t']_\varepsilon = t'$, and $t[t']_{iw'} = \sigma(t_1, \dots, t_{i-1}, t_i[t']_{w'}, t_{i+1}, \dots, t_k)$ for all $k \in \mathbb{N}$, $\sigma \in \Sigma_k$, $t_1, \dots, t_k \in T_\Sigma(Z)$, $i \in [k]$, and $w' \in \text{pos}(t_i)$. For all $\sigma \in \Sigma \cup Z$, we let $\text{pos}_\sigma(t) = \{w \in \text{pos}(t) \mid t(w) = \sigma\}$ and $\text{var}(t) = \{x \in X \mid \text{pos}_x(t) \neq \emptyset\}$. 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 $\sigma(t_1, \dots, t_k)\theta = \sigma(t_1\theta, \dots, t_k\theta)$ for all $k \in \mathbb{N}$, $\sigma \in \Sigma_k$, and $t_1, \dots, t_k \in T_\Sigma(Z)$. We also write the substitution $\theta \in T_\Sigma(Z)^V$ as (i) $[v_1 \leftarrow \theta_{v_1}, \dots, v_n \leftarrow \theta_{v_n}]$ if $V = \{v_1, \dots, v_n\}$ or (ii) $[\theta_{x_1}, \dots, \theta_{x_n}]$ 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 $\mathbb{B} = (\{0, 1\}, \vee, \wedge, 0, 1)$, (ii) the semiring $\mathbb{N} = (\mathbb{N}, +, \cdot, 0, 1)$, (iii) the tropical semiring $\mathbb{T} = (\mathbb{N} \cup \{\infty\}, \min, +, \infty, 0)$, and (iv) the arctic semiring $\mathbb{A} = (\mathbb{N} \cup \{-\infty\}, \max, +, -\infty, 0)$. Given semirings $(\mathbb{S}, +, \cdot, 0, 1)$ and $(\mathbb{T}, \oplus, \odot, \perp, \top)$, a *semiring homomorphism* is a mapping $h \in \mathbb{T}^\mathbb{S}$ such that $h(0) = \perp$, $h(1) = \top$, 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 let $h' \in T_\Delta(X)^\Sigma$ be a mapping 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) $h_{\sigma(t_1, \dots, t_k)} = h'_\sigma[h_{t_1}, \dots, h_{t_k}]$ for all $k \in \mathbb{N}$, $\sigma \in \Sigma_k$, and $t_1, \dots, t_k \in T_\Sigma$. The mapping h is called the *tree homomorphism induced by h'* , and we identify h' and its induced tree homomorphism h . It 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 $\sigma \in \Sigma_k$. 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] (with constraints on direct subtrees). An overview of further developments for these automata can be found in [26]. We essentially use the version recently utilized to solve the HOM problem [15, Definition 4.1].

Definition 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 $\ell \in T_\Sigma(Q) \setminus Q$, $q \in Q$, 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 governing 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 $\ell = \sigma(q_1, \dots, q_k)$ for some $k \in \mathbb{N}$, $\sigma \in \Sigma_k$, and $q_1, \dots, q_k \in Q$, and it is *unconstrained* if $E = \emptyset = I$; in this case we also simply write $\ell \rightarrow q$. 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 $F \in \{0, 1\}^Q$. Finally, if we utilize the Boolean semiring \mathbb{B} , then we reobtain the unweighted versions and omit the ‘wt’ 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 and 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 $t \not\models C$ 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 $w \in \text{pos}(\xi)$, we write $\xi \Rightarrow_{G, t}^{p, w} \zeta$ 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 $t|_w \not\models I$. 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 $\xi_1, \dots, \xi_n \in T_\Sigma(Q)$ such that $t \Rightarrow_{G, t}^{p_1, w_1} \xi_1 \Rightarrow_{G, t}^{p_2, w_2} \cdots \Rightarrow_{G, t}^{p_n, w_n} \xi_n$. 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 ξ_1, \dots, ξ_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 $D_G^q(t)$. The WTGc G is *unambiguous* if $|\sum_{q \in \text{supp}(F)} |D_G^q(t)| \leq 1$ for every $t \in T_\Sigma$.

Definition 3. *The weight of a derivation $d = (p_1, w_1) \cdots (p_n, w_n)$ is defined to be $\text{wt}_G(d) = \prod_{i=1}^n \text{wt}(p_i)$. The weighted tree language generated by G , written simply $G \in \mathbb{S}^{T_\Sigma}$, is defined for every $t \in T_\Sigma$ by*

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

□

Two WTGc are *equivalent* if they generate the same weighted tree language. Finally, a weighted tree language is *regular* if it is generated by a WTG, and it is *constraint-regular* if it is generated by a WTGc. Since the weights of productions are multiplied, we can assume that $\text{wt}_p \neq 0$ for all $p \in P$.

Example 4. Consider the WTGc $G = (Q, \Sigma, F, P, \text{wt})$ over \mathbb{A} with $Q = \{q, q'\}$, $\Sigma = \{\alpha^{(0)}, \gamma^{(1)}, \sigma^{(2)}\}$, $F_q = -\infty$, $F_{q'} = 0$, and P and ‘wt’ given by the productions $p_1 = \alpha \rightarrow_0 q$, $p_2 = \gamma(q) \rightarrow_1 q$, and $p_3 = \sigma(\gamma(q), q) \xrightarrow{11=2}_1 q'$. 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' . Overall, we have $\text{supp}(G) = \{\sigma(\gamma^{i+1}(\alpha), \gamma^i(\alpha)) \mid i \in \mathbb{N}\}$ and $G_t = |\text{pos}_\gamma(t)|$ for every $t \in \text{supp}(G)$. □

For the restricted model of WTAc we introduce another semantics, called initial algebra semantics, which is often more convenient in proofs.

Definition 5. *If G is a WTAc, then for each $q \in Q$ we define $\text{wt}_G^q \in \mathbb{S}^{T_\Sigma}$ for every $t = \sigma(t_1, \dots, t_k)$ with $k \in \mathbb{N}$, $\sigma \in \Sigma_k$, and $t_1, \dots, t_k \in T_\Sigma$ by*

$$\text{wt}_G^q(t) = \sum_{\substack{p = \sigma(q_1, \dots, q_k) \xrightarrow{E, I} q \in P \\ t \models E, t \not\models I}} \text{wt}_p \cdot \prod_{i=1}^k \text{wt}_G^{q_i}(t_i) .$$

□

It is a routine matter to verify $\text{wt}_G^q(t) = \sum_{d \in D_G^q(t)} \text{wt}_G(d)$ for every $q \in Q$ and $t \in T_\Sigma$. Indeed as for WTG and WTA [13] also every WTGc can be turned into an equivalent WTAc at the expense of additional nonterminals by decomposing the left-hand sides.

Lemma 6 (cf. [15, Lemma 4.8]). *WTGc and WTAc are equally expressive.*

Proof. Let $G = (Q, \Sigma, F, P, \text{wt})$ be a WTGc with a non-normalized production $p = \sigma(\ell_1, \dots, \ell_k) \xrightarrow{E,I} q \in P$, let U be an infinite set with $Q \subseteq U$ 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 $G' = (Q', \Sigma, F', P', \text{wt}')$ such that $Q' = Q \cup \{\varphi_{\ell_1}, \dots, \varphi_{\ell_k}\}$, $F'_q = F_q$ for all $q \in Q$ and $F'_{q'} = 0$ for all $q' \in Q' \setminus Q$, and

$$P' = (P \setminus \{p\}) \cup \{\sigma(\varphi_{\ell_1}, \dots, \varphi_{\ell_k}) \xrightarrow{E,I} q\} \cup \{\ell_i \rightarrow \varphi_{\ell_i} \mid i \in [k], \ell_i \notin Q\},$$

and for every $p' \in P'$

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

□

Example 7. Consider the WTGc G of Example 4 and its non-normalized production $p = \sigma(\gamma(q), q) \xrightarrow{11=2}_1 q'$. Applying the construction in the proof of Lemma 6 we replace p by the productions $\sigma(q'', q) \xrightarrow{11=2}_1 q$ and $\gamma(q) \rightarrow_0 q''$, where q'' is some new nonterminal. The such obtained WTGc is already a WTAc. □

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

Lemma 8. *WTAc and WTAc with Boolean final weights are equally expressive.*

Let $d \in D_G^q(t)$ 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] based on Dickson’s Lemma [6], which has been lifted to tree automata in [9]. We note that the construction preserves Boolean final weights.

Lemma 9. *For every WTAc G there exists a WTAc $G' = (Q', \Sigma, F', P', \text{wt}')$ that is equivalent and $\text{wt}'_{G'}(d') \neq 0$ for all $q' \in Q'$, $t' \in T_\Sigma$, and $d' \in D_{G'}^{q'}(t')$.*

For zero-sum free semirings [16,17] we obtain that the support $\text{supp}(G)$ of an WTAc can be generated by a TAc. 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 10 (of Lemmata 6 and 9). *If \mathbb{S} is zero-sum free, then $\text{supp}(G)$ is constraint-regular for every WTGc G .*

Proof. We apply Lemmata 6 and 8 to obtain an equivalent WTAc with Boolean final weights and then Lemma 9 to obtain the WTAc $G' = (Q', \Sigma, F', P', \text{wt}')$ 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 $d' \in D_{G'}^{q'}(t')$ 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 TAc $(Q', \Sigma, \text{supp}(F'), P')$ generates the tree language $\text{supp}(G')$, which is thus constraint-regular. \square

4 Closure Properties

In this section 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 11. *The constraint-regular weighted tree languages (over the same ranked-alphabet) are closed under sums.* \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 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 $\sigma(q_1, \dots, q_k) \xrightarrow{E, I} q \in P$ and $\sigma(q_1, \dots, q_k) \xrightarrow{E', I'} q \in P$. In other words, two productions cannot differ only in the sets of constraints. It is straightforward to turn any WTAc into an equivalent constraint-determined WTAc by introducing additional nonterminals (e.g. annotate the constraints to the state on the right-hand side).

Theorem 12. *The constraint-regular weighted tree languages (over the same ranked alphabet) are closed under HADAMARD product.*

Proof. Let $A, A' \in \mathbb{S}^{T_\Sigma}$ be constraint-regular. Without loss of generality (see Lemma 6) we can assume constraint-determined WTAc $G = (Q, \Sigma, F, P, \text{wt})$ and $G' = (Q', \Sigma, F', P', \text{wt}')$ 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 $p = \sigma(q_1, \dots, q_k) \xrightarrow{E, I} q \in P$ and production $p' = \sigma(q'_1, \dots, q'_k) \xrightarrow{E', I'} q' \in P'$ the production

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

belongs to P'' and its weight is $\text{wt}_{P''}'' = \text{wt}_p \cdot \text{wt}_{P'}'$. No other productions belong to P'' . The 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 productions $p \in P$ and $p' \in P'$ that construct a production $p'' \in P''$. \square

Example 13. Let $G = (\{q\}, \Sigma, F, P, \text{wt})$ and $G' = (\{z\}, \Sigma, F', P', \text{wt}')$ be WTAc over \mathbb{A} and $\Sigma = \{\alpha^{(0)}, \gamma^{(1)}, \sigma^{(2)}\}$, $F_q = F'_z = 0$, and the productions

$$\alpha \rightarrow_0 q \qquad \gamma(q) \rightarrow_2 q \qquad \sigma(q, q) \xrightarrow{1=2}_0 q \qquad (P)$$

$$\alpha \rightarrow_0 z \qquad \gamma(z) \xrightarrow{11 \neq 12}_1 z \qquad \sigma(z, z) \rightarrow_1 z. \qquad (P')$$

We observe that

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

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

and $G_t = 2|\text{pos}_\gamma(t)|$ as well as $G'_t = |\text{pos}_\gamma(t)| + |\text{pos}_\sigma(t)|$ for all $t \in \text{supp}(G)$ and $t' \in \text{supp}(G')$. 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.

$$\alpha \rightarrow_0 \langle q, z \rangle \qquad \gamma(\langle q, z \rangle) \xrightarrow{11 \neq 12}_3 \langle q, z \rangle \qquad \sigma(\langle q, z \rangle, \langle q, z \rangle) \xrightarrow{1=2}_1 \langle q, z \rangle$$

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 $t \in \text{supp}(G) \cap \text{supp}(G')$. \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 14. *Let $h \in \mathbb{T}^{\mathbb{S}}$ be a semiring homomorphism into a finite semiring \mathbb{T} . For every WTAc $G = (Q, \Sigma, F, P, \text{wt})$ over \mathbb{S} there exists an unambiguous WTAc $G' = (\mathbb{T}^Q, \Sigma, F', P', \text{wt}')$ such that for every $t \in T_\Sigma$ and $\varphi \in \mathbb{T}^Q$*

$$\text{wt}_{G'}^\varphi(t) = \begin{cases} 1 & \text{if } \varphi_q = h_{\text{wt}_G^q(t)} \text{ for all } q \in Q \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, $G'_t = h_{G(t)}$ for every $t \in T_\Sigma$.

Proof. Let $\mathcal{C} = \{E \mid \sigma(q_1, \dots, q_k) \xrightarrow{E, I} q \in P\} \cup \{I \mid \sigma(q_1, \dots, q_k) \xrightarrow{E, I} q \in P\}$ be the constraints that occur in G . 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 $\varphi^1, \dots, \varphi^k \in \mathbb{T}^Q$, and constraints $\mathcal{E} \subseteq \mathcal{C}$ we let $p' = \sigma(\varphi^1, \dots, \varphi^k) \xrightarrow{\mathcal{E}, \mathcal{I}} \varphi \in P'$, where $\mathcal{I} = \mathcal{C} \setminus \mathcal{E}$ and for every $q \in Q$

$$\varphi_q = \sum_{\substack{p=\sigma(q_1, \dots, q_k) \\ E \subseteq \mathcal{E}, I \subseteq \mathcal{I}}} h_{\text{wt}(p)} \cdot \varphi_{q_1}^1 \cdot \dots \cdot \varphi_{q_k}^k \quad (1)$$

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 $2^{|\mathcal{C}|}$ 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 $\varphi_q = h_{\text{wt}'_{G'}(t)}$ for every $q \in Q$. The weight of that derivation is 1. These statements are proven inductively. The final statement $G'_t = h_{G'(t)}$ is an easy consequence of the previous statements. \square

Example 15. Reconsider the WTAc obtained as the disjoint union of the WTAc G and G' of Example 13 as well as the semiring homomorphism $h \in \mathbb{B}^{\mathbb{A}}$ given by $h_a = 1$ for all $a \in \mathbb{A} \setminus \{-\infty\}$ and $h_{-\infty} = 0$. The set \mathcal{C} of utilized constraints is $\{(1, 2), (11, 12)\}$ 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.

$$\begin{array}{ll} \alpha \xrightarrow{1 \neq 2, 11 \neq 12} \{q, z\} & \\ \gamma(Q') \xrightarrow{1=2, 11=12} Q' \cap \{q\} & \gamma(Q') \xrightarrow{1 \neq 2, 11=12} Q' \cap \{q\} \\ \gamma(Q') \xrightarrow{1=2, 11 \neq 12} Q' & \gamma(Q') \xrightarrow{1 \neq 2, 11 \neq 12} Q' \\ \sigma(Q', Q'') \xrightarrow{1=2, 11=12} Q' \cap Q'' & \sigma(Q', Q'') \xrightarrow{1 \neq 2, 11=12} Q' \cap Q'' \cap \{z\} \\ \sigma(Q', Q'') \xrightarrow{1=2, 11 \neq 12} Q' \cap Q'' & \sigma(Q', Q'') \xrightarrow{1 \neq 2, 11 \neq 12} Q' \cap Q'' \cap \{z\} \end{array}$$

Each $t \in T_\Sigma$ has exactly one left-most complete derivation in G'' ; it derives to Q' , where (i) $q \in Q'$ iff $t \in \text{supp}(G)$ and (ii) $z \in Q'$ iff $t \in \text{supp}(G')$. \square

Corollary 16 (of Theorem 14). *Let \mathbb{S} be finite. For every WTAc over \mathbb{S} there exists an equivalent unambiguous WTAc.* \square

Corollary 17 (of Theorem 14). *Let \mathbb{S} be zero-sum free. For every WTAc G over \mathbb{S} there exists an unambiguous TAc generating $\text{supp}(G)$.*

Proof. Utilizing Lemma 8 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 [27]. By Lemma 9 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 $h_{\text{wt}(p)}$ by 1 in (1). The such obtained TAc generates $\text{supp}(G)$. \square

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 $A|_{\text{supp}(A')} \in \mathbb{S}^{T_\Sigma}$ for every $t \in T_\Sigma$ by $A|_{\text{supp}(A')}(t) = A(t)$ if $t \in \text{supp}(A')$ and $A|_{\text{supp}(A')}(t) = 0$ otherwise. Utilizing the unambiguous WTAc and the HADAMARD product, we can show that $A|_{\text{supp}(A')}$ is constraint-regular if A and A' are constraint-regular and the semiring \mathbb{S} is zero-sum free.

Theorem 18. *Let \mathbb{S} be zero-sum free. For all WTAc G and G' there exists a WTAc H such that $H = G|_{\text{supp}(G')}$.*

Proof. By Corollary 10 the support $\text{supp}(G')$ is constraint-regular. Hence we can obtain an unambiguous WTAc G'' for $\text{supp}(G')$ using Theorem 14. 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 12 generates exactly $G|_{\text{supp}(G')}$ as required. \square

5 Towards HOM Problem

The strategy of [15] for deciding the HOM problem first represents the homomorphic image L of the regular tree language 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 $G' = G$, then L is regular. If not, pumping arguments are used to prove that it is impossible to find any TA for L . Overall, they reduce 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 19. *Let $G = (Q, \Sigma, F, P, \text{wt})$ be a WTA and $h \in T_{\Delta}^{T_\Sigma}$ be a nondeleting and nonerasing tree homomorphism. There exists a WTGc G' with $G' = h(G)$.*

Proof. We construct a WTGc G' for $h(G)$ in two stages. First, we construct the WTGc $G'' = (Q \cup \{\perp\}, \Delta \cup \Delta \times P, F'', P'', \text{wt}'')$ such that for every production $p = \sigma(q_1, \dots, q_k) \rightarrow q \in P$ and $h_\sigma = u = \delta(u_1, \dots, u_n)$,

$$p'' = \left(\langle \delta, p \rangle(u_1, \dots, u_n) \llbracket q_1, \dots, q_k \rrbracket \xrightarrow{E, \emptyset} q \right) \in P'' \quad \text{with} \quad E = \bigcup_{i \in [k]} \text{pos}_{x_i}(u)^2$$

where the substitution $\langle \delta, p \rangle(u_1, \dots, u_n) \llbracket q_1, \dots, q_k \rrbracket$ replaces for every $i \in [k]$ only the left-most occurrence of x_i in $\langle \delta, p \rangle(u_1, \dots, u_n)$ by q_i and all other occurrences by \perp . Moreover $\text{wt}''_{p''} = \text{wt}_p$. Additionally, we let $p''_\delta = \delta(\perp, \dots, \perp) \rightarrow \perp \in P''$ with $\text{wt}''_{p''_\delta} = 1$ 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 $F''_\perp = 0$.

We can now delete the annotation. First we remove all productions to \perp that are labeled with symbols from $\Delta \times P$. Second, we use a deterministic relabeling

to remove the second components of labels of $\Delta \times P$. Thus, we overall obtain a WTGc G' (using only equality constraints) such that $G' = h(G)$.

The sole purpose of the annotations is to establish a one-to-one correspondence between the valid runs of G and those of G'' , before evaluating the sums to compute $h(G)$. This simplifies the understanding of the correctness of the construction, but is otherwise superfluous and may be omitted for efficiency. \square

Let us illustrate the construction on a simple example.

Example 20. Consider the WTA $G = (\{q, q'\}, \Sigma, F, P, \text{wt})$ over the semiring \mathbb{N} with $\Sigma = \{\alpha^{(0)}, \phi^{(1)}, \gamma^{(1)}, \epsilon^{(1)}\}$, $F_q = 0$, $F_{q'} = 1$, and the set of productions and their weights given by $p_1 = \alpha \rightarrow_1 q$, $p_2 = \gamma(q) \rightarrow_2 q$, $p_3 = \epsilon(q) \rightarrow_1 q$ and $p_4 = \phi(q) \rightarrow_1 q'$. We have $\text{supp}(G) = \{\phi(t) \mid t \in T_{\Sigma \setminus \{\phi\}}\}$ and $G_t = 2^{|\text{pos}_\gamma(t)|}$ for $t \in \text{supp}(G)$. Consider the ranked alphabet $\Delta = \{\alpha^{(0)}, \gamma^{(1)}, \sigma^{(2)}\}$ and the homomorphism h induced by $h_\alpha = \alpha$, $h_\gamma = h_\epsilon = \gamma(x_1)$, and $h_\phi = \sigma(\gamma(x_1), x_1)$. So $\text{supp}(h(G)) = \{\sigma(\gamma^{n+1}(\alpha), \gamma^n(\alpha)) \mid n \in \mathbb{N}\}$ and $h(G)_t = \sum_{k=0}^n \binom{n}{k} 2^k = 3^n$ for every $t = \sigma(\gamma^{n+1}(\alpha), \gamma^n(\alpha)) \in \text{supp}(h(G))$. A WTGc for $h(G)$ is constructed as follows. First, we let $G'' = (\{q, q', \perp\}, \Delta \cup \Delta \times P, F'', P'', \text{wt}'')$ with $F''_q = 1$, $F''_{q'} = F''_\perp = 0$ and the productions and their weights are given by

$$\langle \alpha, p_1 \rangle \rightarrow_1 q \quad \langle \gamma, p_2 \rangle(q) \rightarrow_2 q \quad \langle \gamma, p_3 \rangle(q) \rightarrow_1 q \quad \langle \sigma, p_4 \rangle(\gamma(q), \perp) \xrightarrow{11=2}_1 q'$$

and $\delta(\perp, \dots, \perp) \rightarrow_1 \perp$ for all $\delta \in \Delta \cup \Delta \times P$. Next we remove the second component of the labels and add weights of productions that become equal. This applies to the production $\gamma(q) \rightarrow q$, which obtains the sum of the two productions (with annotations p_2 and p_3). So we obtain the WTGc $G' = (\{q, q', \perp\}, \Delta, F'', P', \text{wt}'')$ with the following productions for all $\delta \in \Delta$.

$$\alpha \rightarrow_1 q \quad \gamma(q) \rightarrow_3 q \quad \sigma(\gamma(q), \perp) \xrightarrow{11=2}_1 q' \quad \delta(\perp, \dots, \perp) \rightarrow_1 \perp$$

\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 from [15] to the support of $h(G)$ 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 19 to those used in [15]. This requires that the equality constraints in every production refer to positions that occur in its left-hand side and are labeled by the same nonterminal.

Definition 21. A WTGc $G = (Q, \Sigma, F, P, \text{wt})$ is classic if $\{v, v'\} \subseteq \text{pos}(\ell)$ and $\ell(v) = \ell(v') \in Q$ for every production $\ell \xrightarrow{E, I} q \in P$ and $(v, v') \in E$. \square

Theorem 22. Let \mathbb{S} be a zero-sum free semiring, $G = (Q, \Sigma, F, P, \text{wt})$ be a WTA and $h \in T_{\Delta}^{T\Sigma}$ be a nondeleting and nonerasing tree homomorphism. Finally, let $A = h(G)$. Emptiness and finiteness of $\text{supp}(A)$ are decidable.

The proof of Theorem 22 applies the corresponding result for the unweighted case. In short, we use Theorem 19 to represent A by a WTGc for which we drop the weights. The resulting TGc representing $\text{supp}(A)$ is then modified into an equivalent, classic one. For this, emptiness and finiteness are decidable by [15].

References

1. Bogaert, B., Tison, S.: Equality and disequality constraints on direct subterms in tree automata. In: Finkel, A., Jantzen, M. (eds.) STACS 1992. LNCS, vol. 577, pp. 159–171. Springer, Heidelberg (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)
4. Comon, H., et al.: Tree automata – Techniques and applications (2007)
5. Comon, H., Jacquemard, F.: Ground reducibility and automata with disequality constraints. In: Enjalbert, P., Mayr, E.W., Wagner, K.W. (eds.) STACS 1994. LNCS, vol. 775, pp. 149–162. Springer, Heidelberg (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)
7. Doner, J.: Tree acceptors and some of their applications. *J. Comput. System Sci.* **4**(5), 406–451 (1970)
8. Drewes, F.: Grammatical Picture Generation: A Tree-Based Approach. Springer, Heidelberg (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)
10. Ésik, Z., Kuich, W.: Formal tree series. *J. Autom. Lang. Comb.* **8**(2), 219–285 (2003)
11. Fülöp, Z., Maletti, A., Vogler, H.: Preservation of recognizability for synchronous tree substitution grammars. In: Proceedings of the 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. *Fundamenta Informaticae* **111**(2), 163–202 (2011)
13. Fülöp, Z., Vogler, H.: Weighted tree automata and tree transducers. In: Droste, M., Kuich, W., Vogler, H. (eds.) Handbook of Weighted Automata. Monographs in Theoretical Computer Science. An EATCS Series, pp. 313–403. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-01492-5_9
14. Gécseg, F., Steinby, M.: Tree automata. Technical report 1509.06233, arXiv (2015)
15. Godoy, G., Giménez, O.: The HOM problem is decidable. *J. ACM* **60**(4), 1–44 (2013)
16. Golan, J.S.: Semirings and Their Applications. Kluwer Academic, Dordrecht (1999)
17. Hebisch, U., Weinert, H.J.: Semirings - Algebraic Theory and Applications in Computer Science. World Scientific, Singapore (1998)
18. Jurafsky, D., Martin, J.H.: Speech and Language Processing, 2nd edn. Prentice Hall, Hoboken (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)

20. Mongy-Steen, J.: Transformation de noyaux reconnaissables d'arbres. Forêts RATEG. Ph.D. thesis, Université de Lille (1981)
21. Perrin, D.: Recent results on automata and infinite words. In: Chytil, M.P., Koubek, V. (eds.) MFCS 1984. LNCS, vol. 176, pp. 134–148. Springer, Heidelberg (1984). <https://doi.org/10.1007/BFb0030294>
22. Salomaa, A., Soittola, M.: Automata-Theoretic Aspects of Formal Power Series. Springer, New York (1978). <https://doi.org/10.1007/978-1-4612-6264-0>
23. Schützenberger, M.P.: On the definition of a family of automata. *Inf. Control* **4**(2–3), 245–270 (1961)
24. 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)
25. 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>
26. Tison, S.: Tree automata, (dis-)equality constraints and term rewriting: what's new? In: Proceedings of the 22nd International Conference on Rewriting Techniques and Applications. LIPIcs, vol. 10, pp. 1–2. Schloss Dagstuhl – Leibniz-Zentrum für Informatik (2011)
27. Wang, H.: On characters of semirings. *Houston J. Math.* **23**(3), 391–405 (1997)
28. Wilhelm, R., Seidl, H., Hack, S.: *Compiler Design*. Springer, Heidelberg (2013). <https://doi.org/10.1007/978-3-642-17540-4>