Solving the Weighted HOM-Problem With the Help of Unambiguity

Andreea-Teodora Nász
Leipzig University
Leipzig, Germany
Faculty of Mathematics and Computer Science
PO box 100 920, 04009 Leipzig, Germany
nasz@informatik.uni-leipzig.de

The *HOM-problem*, which asks whether the image of a regular tree language under a tree homomorphism is again regular, is known to be decidable by [Godoy, Giménez, Ramos, Àlvarez: The HOM problem is decidable. STOC (2010)]. Research on the weighted version of this problem, however, is still in its infancy since it requires customized investigations. In this paper we address the weighted HOM-problem and strive to keep the underlying semiring as general as possible. In return, we restrict the input: We require the tree homomorphism h to be *tetris-free*, a condition weaker than injectivity, and for the given weighted tree automaton, we propose an ambiguity notion with respect to h. These assumptions suffice to ensure decidability of the thus restricted HOM-problem for all zero-sum free semirings by allowing us to reduce it to the (decidable) unweighted case.

1 Introduction

Over the past decades, various extensions to the well-known model of finite-state automata have been proposed. These acceptors were taken to the next level when their qualitative evaluation was generalized to a quantitative one, which led to the concept of *weighted automata* [29]. Such devices assign a weight to each input word, thus computing so-called *formal power series*. Weighted automata are commonly used to model numerical factors related to the input, such as costs, probabilities and consumption of resources or time, and enjoy consistent attention from the research community focused on automata theory [8, 28]. The favored algebraic structure for performing weight calculations are semirings [14, 16], as they are quite general while still being computationally efficient due to their distributivity.

Another dimension for generalizing finite-state automata lifts their input to more complex data structures such as infinite words [20, 26], trees [2], graphs [1] and pictures [11, 27]. Particularly, *finite-state tree automata* were introduced independently in [5, 31, 32]. The so-called *regular tree languages* they recognize have been studied extensively [2], and find applications in a variety of areas like natural language processing [17], picture generation [6] and compiler construction [33]. In many cases, applications require both types of generalizations, and so several models of *weighted tree automata* (WTA) and the *regular tree series* they recognize continue to be studied [9].

The price to pay for the simplicity of tree automata lies in their significant limitations. For instance, they cannot ensure that certain subtrees of input trees are equal [10], much like the classical (string) automata cannot ensure that the number of a's and b's in a word is equal. This defect was tackled with extensions proposed in [25] and [3, 12, 13] where *tree automata with constraints* can explicitly require or forbid certain subtrees to be equal. Such devices have played a crucial part in deciding the *HOM-problem*: This long-standing open question [2] asks, given a regular tree language and a tree homomorphism, whether the image is again regular. A tree homomorphism performs a transformation on trees

Zs. Gazdag, Sz. Iván, G. Kovásznai (Eds.): 16th International Conference on Automata and Formal Languages (AFL 2023)
EPTCS 386, 2023, pp. 200–214, doi:10.4204/EPTCS.386.16

and can duplicate subtrees, therefore the trees in the homomorphic image might have certain identical subtrees, which calls for the constraints mentioned above. In [12], the authors first represent this homomorphic image of a regular tree language by a tree automaton with explicit constraints, and then decide algorithmically if the language it recognizes is regular despite the constraints it imposes.

The nature of the *weighted* HOM-problem, where a regular tree series and a tree homomorphism are given as input, requires an individual investigation for different semirings. Recently, the approach from [12] was adjusted to the special case of nonnegative integers [23], but so far, the question remains open for other semirings. In this paper, we reverse the strategy and impose conditions on the input in order to decide the thus restricted HOM-problem for a larger class of semirings. More precisely, we require our protagonist – the weighted tree automaton with constraints – to be unambiguous, and reduce the question of its regularity to the unweighted case from [12] for any zero-sum free (commutative) semiring. Afterwards, we phrase a condition on the input of the HOM-problem which ensures that our reduction is applicable.

This article consists of five sections including its introduction. Our main contributions can be summarized as follows:

- In Section 2 we establish notations and recall the main objects that will play a role throughout the paper, primarily the *weighted tree automata with hom-constraints (WTAh)* which are used to represent homomorphic images of regular tree series.
- In Section 3 we prove that regularity is decidable for the unambiguous devices of this type over zero-sum free semirings. We achieve this by reducing the question to the unweighted case where regularity is known to be decidable [12].
- In Section 4 we integrate this decidability result into the HOM-problem. To this end, we phrase a condition on the input of the HOM-problem which guarantees that the WTAh constructed for this instance is unambiguous. Thus, the HOM-problem with input restricted accordingly is decidable for any zero-sum free semiring.
- Finally, in Section 5 we briefly summarize our results and discuss further research that will extend the present work.

2 Preliminaries and Technical Background

We begin as usual with the necessary background for this paper.

General Notation

We denote the set $\{0,1,2,\ldots\}$ of nonnegative integers by $\mathbb{N}$, and we let $[k] = \{1,\ldots,k\}$ for every $k \in \mathbb{N}$. Let A and B be sets. We write $|A|$ for the cardinality of A, and A^* for the set of finite strings over A. The empty string is ε and the length of a string w is $|w|$. For a mapping $f\colon A \to B$ and $S \subseteq B$ we denote the inverse image of S under f by $f^{-1}(S)$, and we write $f^{-1}(b)$ instead of $f^{-1}(\{b\})$ for every $b \in B$.

Trees

A *ranked alphabet* is a pair (Σ,rk) that consists of a finite set Σ and a rank mapping $\mathrm{rk}\colon \Sigma \to \mathbb{N}$. For every $k \geq 0$, we define $\Sigma_k = \mathrm{rk}^{-1}(k)$, and we sometimes write $\sigma^{(k)}$ to indicate that $\sigma \in \Sigma_k$. We often abbreviate (Σ,rk) by Σ leaving rk implicit. Let Z be a set disjoint with Σ. The set of Σ-*trees over* Z, denoted $T_\Sigma(Z)$, is the smallest set T such that (i) $\Sigma_0 \cup Z \subseteq T$ and (ii) $\sigma(t_1,\dots,t_k) \in T$ for every $k \in \mathbb{N}$, $\sigma \in \Sigma_k$ and $t_1,\dots,t_k \in T$. We abbreviate $T_\Sigma(\emptyset)$ simply to T_Σ, and call any subset $L \subseteq T_\Sigma$ a *tree language*. Consider $t \in T_\Sigma(Z)$. The set $\text{pos}(t) \subseteq \mathbb{N}^*$ of *positions of t* is defined inductively by $\text{pos}(t) = \varepsilon$ for every $t \in \Sigma_0 \cup Z$, and by

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

for all $k \in \mathbb{N}$, $\sigma \in \Sigma_k$ and $t_1,\dots,t_k \in T_\Sigma(Z)$. The set of positions of t inherits the lexicographic order $\leq_{\text{lex}}$ from $\mathbb{N}^*$. The *size* $|t|$ and *height* $\text{ht}(t)$ *of t* are defined as

$$|t| \;=\; |\text{pos}(t)| \quad \text{and} \quad \text{ht}(t) = \max_{p\in\text{pos}(t)} |p| \;.$$

For $p \in \text{pos}(t)$, the *label $t(p)$ of t at p*, the *subtree $t|_p$ of t at p* and the *substitution $t[t']_p$ of t' into t at p* are defined

- for $t \in \Sigma_0 \cup Z$ by $t(\varepsilon) = t|_\varepsilon = t$ and $t[t']_\varepsilon = t'$, and
- for $t = \sigma(t_1,\dots,t_k)$ by $t(\varepsilon) = \sigma,\ t(ip') = t_i(p'),\ t|_\varepsilon = t,\ t|_{ip'} = t_i|_{p'},\ t[t']_\varepsilon = t'$, and

  $$t[t']_{ip'} = \sigma(t_1,\dots,t_{i-1},t_i[t']_{p'},t_{i+1},\dots,t_k)$$

  for all $k \in \mathbb{N},\ \sigma \in \Sigma_k,\ t_1,\dots,t_k \in T_\Sigma(Z),\ i \in [k]$ and $p' \in \text{pos}(t_i)$.

For every subset $S \subseteq \Sigma \cup Z$, we let $\text{pos}_S(t) = \{p \in \text{pos}(t) \mid t(p) \in S\}$ and we abbreviate $\text{pos}_{\{s\}}(t)$ by $\text{pos}_s(t)$ for every $s \in \Sigma \cup Z$. Let $X = \{x_1, x_2, \dots\}$ be a fixed, countable set of formal variables. For $k \in \mathbb{N}$ we denote by X_k the subset $\{x_1,\dots,x_k\}$. For any $t \in T_\Sigma(X)$ we let

$$\text{var}(t) = \{x \in X \mid \text{pos}_x(t) \neq \emptyset\} \;.$$

Finally, for $t \in T_\Sigma(Z)$, a subset $V \subseteq Z$ and a mapping $\theta\colon V \to T_\Sigma(Z)$, we define the *substitution $t\theta$ applied to t* by $v\theta = \theta(v)$ for $v \in V$, $z\theta = z$ for $z \in Z \setminus V$, and

$$\sigma(t_1,\dots,t_k)\theta = \sigma\big(t_1\theta,\dots,t_k\theta\big)$$

for all $k \in \mathbb{N}$, $\sigma \in \Sigma_k$ and $t_1,\dots,t_k \in T_\Sigma(Z)$. If $V = \{v_1,\dots,v_n\}$, we write the substitution θ explicitly as $[v_1 \leftarrow \theta(v_1),\dots,v_n \leftarrow \theta(v_n)]$, and abbreviate it further to $[\theta(x_1),\dots,\theta(x_n)]$ if $V = X_n$.

Semirings and Tree Series

A *(commutative) semiring* [14, 15] 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

- the Boolean semiring $\mathbb{B} = \big(\{0,1\},\vee,\wedge,0,1\big)$,
- the semiring $\mathbb{N} = \big(\mathbb{N},+,\cdot,0,1\big)$,
- the semiring $\mathbb{Z} = \big(\mathbb{Z},+,\cdot,0,1\big)$,
- the tropical semiring $\mathbb{T} = \big(\mathbb{N}\cup\{\infty\},\min,+,\infty,0\big)$, and
- the arctic semiring $\mathbb{A} = \big(\mathbb{N}\cup\{-\infty\},\max,+,-\infty,0\big)$.

When there is no risk of confusion, we refer to a semiring $(\mathbb{S},+,\cdot,0,1)$ simply by its carrier set $\mathbb{S}$. We call $\mathbb{S}$ *zero-sum free* if $a+b = 0$ implies $a = b = 0$ for all $a,b \in \mathbb{S}$. All semirings listed above except for $\mathbb{Z}$ are zero-sum free. Let Σ be a ranked alphabet and Z a set. Any mapping $\varphi\colon T_\Sigma(Z) \to \mathbb{S}$ is called a *tree series* or *weighted tree language* over $\mathbb{S}$, and its *support* is the set $\text{supp}(\varphi) = \{t \in T_\Sigma(Z) \mid \varphi(t) \neq 0\}$.

Tree Homomorphisms

Given ranked alphabets Σ and Δ, let $h' \colon \Sigma \to T_\Delta(X)$ be a mapping that satisfies $h'(\sigma) \in T_\Delta(X_k)$ for all $k \in \mathbb{N}$ and $\sigma \in \Sigma_k$. We extend h' to $h \colon T_\Sigma \to T_\Delta$ by $h(\alpha) = h'(\alpha) \in T_\Delta(X_0) = T_\Delta$ for all $\alpha \in \Sigma_0$ and

$$h(\sigma(s_1,\dots,s_k)) = h'(\sigma)[x_1 \leftarrow h(s_1),\dots,x_k \leftarrow h(s_k)]$$

for all $k \in \mathbb{N}$, $\sigma \in \Sigma_k$, and $s_1,\dots,s_k \in T_\Sigma$. The mapping h is called the *tree homomorphism induced by* h', and we identify h' and its induced tree homomorphism h. We call h

- *nonerasing* if $h(\sigma) \notin X$ for all $\sigma \in \Sigma$,
- *nondeleting* if $\sigma \in \Sigma_k$ implies $\mathrm{var}(h'(\sigma)) = X_k$ for all $k \in \mathbb{N}$, and
- *input-finitary* if the preimage $h^{-1}(t)$ is finite for every $t \in T_\Delta$.

If a tree homomorphism $h \colon T_\Sigma \to T_\Delta$ is nonerasing and nondeleting, then for every $s \in h^{-1}(t)$, it is $|s| \le |t|$. In particular, h is then input finitary.

Consider a tree series $A \colon T_\Sigma \to \mathbb{S}$. Its *homomorphic image under* h is the tree series $h_A \colon T_\Delta \to \mathbb{S}$ defined for every $t \in T_\Delta$ by

$$h_A(t) = \sum_{s \in h^{-1}(t)} A(s) \ .$$

This definition relies on the tree homomorphism to be input-finitary, otherwise the above sum is not finite, so the value $h_A(t)$ might not be well-defined. For this reason, we will only consider nondeleting and nonerasing tree homomorphisms.

Weighted Tree Automata with Constraints

Recently it was shown [21, 22] that such homomorphic images of regular tree languages can be represented efficiently using *weighted tree automata with hom-constraints* (WTAh). These devices were first introduced for the unweighted case in [12] and defined for zero-sum free commutative semirings in [21].

Definition 1 (cf. [22, Definition 1])**.** *Let* $\mathbb{S}$ *be a commutative semiring. A* weighted tree automaton over $\mathbb{S}$ with hom-constraints (WTAh) *is a tuple* $\mathscr{A} = \big(Q,\Sigma,F,R,\mathrm{wt}\big)$ *such that* Q *is a finite set of states,* Σ *is a ranked alphabet,* $F \subseteq Q$ *is the set of final states,* R *is a finite set of rules of the form* (ℓ,q,E) *such that* $\ell \in T_\Sigma(Q) \setminus Q$*,* $q \in Q$ *and* E *is an equivalence relation on* $\mathrm{pos}_Q(\ell)$*, and* $\mathrm{wt} \colon R \to \mathbb{S}$ *assigns a weight to each rule.*

Rules of a WTAh are typically depicted as $r = \ell \xrightarrow{E}_{\mathrm{wt}(r)} q$. The components of such a rule are the *left-hand side* ℓ, the *target state* q, the set E of *hom-constraints* and the *weight* $\mathrm{wt}(r)$. A hom-constraint $(p,p') \in E$ is listed as " $p = p'$ ", and if p and p' are distinct, then p, p' are called *constrained positions*. The equivalence class of p in E is denoted $[p]_{\equiv_E}$. We typically omit the trivial constraints $(p,p) \in E$.

Example 2. *Let* Σ *be the ranked alphabet* $\{a^{(0)}, g^{(1)}, k^{(2)}\}$*. Consider the WTAh* $\mathscr{A} = \big(Q,\Sigma,F,R,\mathrm{wt}\big)$ *over* $\mathbb{Z}$ *with* $Q = \{q, q_f\}$*,* $F = \{q_f\}$ *and the set of rules and weights*

$$R \;=\; \big\{\; a \to_1 q\,, \quad g(q) \to_2 q\,, \quad k\big(q,g(q)\big) \xrightarrow{1=21}_1 q_f \;\big\}\,.$$

The only constrained positions are 1 *and* 21 *in the rule with left-hand side* $k\big(q,g(q)\big)$*.*

The WTAh is a *weighted tree grammar* (WTG) if E is the identity relation for every rule $\ell \xrightarrow{E} q$, and a WTA in the classical sense [2] if additionally $\mathrm{pos}_\Sigma(\ell) = \{\varepsilon\}$. WTG and WTA are equally expressive, as WTG can be translated straightforwardly into WTA using additional states.

In this work, we are particularly interested in a specific subclass of WTAh, namely the *eq-restricted* WTAh [22]. In such a device, there is a designated *sink-state* whose sole purpose is to neutrally process copies of identical subtrees. More precisely, whenever different subtrees are mutually constrained, there is one leading copy among them that can be processed with arbitrary states and weights, while every other copy is handled exclusively by the weight-neutral sink-state.

Definition 3. *A WTAh $(Q,\Sigma,F,R,\mathrm{wt})$ is* eq-restricted *if it has a* sink state $\bot \in Q \setminus F$ *such that*

- *for all $\sigma \in \Sigma$, the rule $\sigma(\bot,\dots,\bot) \to_1 \bot$ belongs to R, and no other rule targets $\bot$, and*
- *for every rule $\ell \xrightarrow{E} q$ with $q \neq \bot$, the following conditions hold:*
  Let $\mathrm{pos}_Q(\ell) = \{p_1,\dots,p_n\}$ and $q_i = \ell(p_i)$ for $i \in [n]$.
  1. *For each $i \in [n]$, there exists $q' \in Q \setminus \{\bot\}$ with $\{q_j \mid p_j \in [p_i]_{\equiv_E}\} \setminus \{\bot\} = \{q'\}$.*
  2. *There exists exactly one $p_j \in [p_i]_{\equiv_E}$ such that $q_j = q'$.*

In other words, among each E-equivalence class of positions of a left-hand side ℓ, there is only one occurrence of a state different from $\bot$, every other related position is labelled by $\bot$. Moreover, $\bot$ processes every possible tree with weight 1. Whenever we consider an eq-restricted WTAh, we denote its state set by $Q\dot\cup\{\bot\}$ instead of $Q \ni \bot$ to point out the sink-state.

Example 4. *Recall the WTAc $\mathscr{A}$ from Example 2. It is not eq-restricted since the constrained positions* 1 *and* 21 *are both labeled by the same state, which is not a sink state. Instead, let us add a non-final state $\bot$ to Q, replace the rule $k(q,g(q)) \xrightarrow{1=21}_1 q_f$ with $k(q,g(\bot)) \xrightarrow{1=21}_1 q_f$ and add the required rules targeting $\bot$ to obtain an eq-restricted WTAh $\mathscr{A}'$. More precisely, we have the eq-restricted WTAh $\mathscr{A}' = (\{q,q_f,\bot\},\Delta,\{q_f\},R',\mathrm{wt}')$ with the set of rules and weights*

$$
\begin{aligned}
R' = \{\, & a \to_1 q, \quad g(q) \to_2 q, \quad k(q,g(\bot)) \xrightarrow{1=21}_1 q_f \,\} \\
\cup \{\, & a \to_1 \bot, \quad g(\bot) \to_1 \bot, \quad k(\bot,\bot) \to_1 \bot \,\} .
\end{aligned}
$$

Next, let us recall the semantics of WTAh from [22, Definitions 2 and 3].

Definition 5. *Let $\mathscr{A} = (Q,\Sigma,F,R,\mathrm{wt})$ be a WTAh. A* run of $\mathscr{A}$ *is a tree over the ranked alphabet $\Sigma \cup R$ where the rank of a rule is $\mathrm{rk}(\ell \xrightarrow{E} q) = |\mathrm{pos}_Q(\ell)|$, and it is defined inductively. Consider $t_1,\dots,t_n \in T_\Sigma$, $q_1,\dots,q_n \in Q$ and suppose that ρ_i is a run of $\mathscr{A}$ for t_i to q_i with weight $\mathrm{wt}(\rho_i) = a_i$ for each $i \in [n]$. Assume that there is a rule of the form $\ell \xrightarrow{E}_a q$ in R such that $\ell = \sigma(\ell_1,\dots,\ell_m)$, $\mathrm{pos}_Q(\ell) = \{p_1,\dots,p_n\}$ with $\ell(p_i) = q_i$ and that for all $p_i = p_j \in E$, it is $t_i = t_j$. Then the following is a run of $\mathscr{A}$ for the tree $t = \ell[t_1]_{p_1}\cdots[t_n]_{p_n}$ to q:*

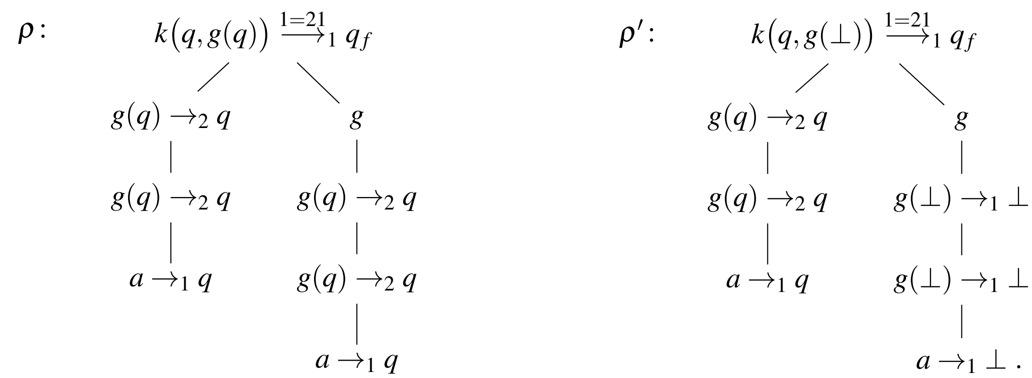
$$\rho = \big(\ell \xrightarrow{E}_a q\big)(\ell_1,\dots,\ell_m)[\rho_1]_{p_1}\cdots[\rho_n]_{p_n} .$$

Its weight $\mathrm{wt}(\rho)$ *is computed as $a \cdot \prod_{i\in[n]} a_i$. If $\mathrm{wt}(\rho) \neq 0$, then ρ is* valid*, and if in addition, $q \in F$ for its* target state *q, then ρ is* accepting*. We call $\mathscr{A}$* unambiguous *if for every $t \in T_\Sigma$ there is at most one accepting run. The value $\mathrm{wt}^q(t)$ is the sum of all weights $\mathrm{wt}(\rho)$ of runs of $\mathscr{A}$ for t to q. Finally, the tree series recognized by $\mathscr{A}$ is defined simply by*

$$[\![\mathscr{A}]\!]\colon\ T_\Sigma \to \mathbb{S}, \quad t \mapsto \sum_{q\in F} \mathrm{wt}^q(t) .$$

Since the weights of rules are multiplied, we can assume wlog. that $\mathrm{wt}(r) \neq 0$ for all $r \in R$, which we will do from now on. Finally, two WTAh are said to be *equivalent* if they recognize the same tree series.

Example 6. *Recall the WTAh $\mathscr{A}$ and $\mathscr{A}'$ from Examples 2 and 4 and consider the tree $k\big(g^2(a), g^3(a)\big)$. The accepting runs ρ and ρ' of $\mathscr{A}$ and $\mathscr{A}'$, respectively, for it are the following:*



It is $\mathrm{wt}(\rho) = 2^4$ *while* $\mathrm{wt}'(\rho') = 2^2$ *because in the eq-restricted WTAh $\mathscr{A}'$, every constrained subtree except for one (pending from position* 1*) is processed exclusively in the state $\perp$ with weight* 1.

Both WTAh are unambiguous, so it is impossible for different accepting runs with complementary weights to cancel out. Thus for a tree $t \in T_\Sigma$ it is $t \in \mathrm{supp}[\![\mathscr{A}]\!]$ iff. $\mathscr{A}$ has an accepting run for t, and the same is true for $\mathscr{A}'$. In fact, it is

$$\mathrm{supp}[\![\mathscr{A}]\!] \;=\; \mathrm{supp}[\![\mathscr{A}']\!] \;=\; \big\{k\big(g^n a, g^{n+1}(a)\big) \mid n \in \mathbb{N}\big\}\,.$$

If a tree series is recognized by a WTA, it is called *regular*, if it is recognized by some WTAh, then it is called *constraint-regular*, and if it is recognized by an eq-restricted WTAh, then it is called *hom-regular*. This choice of name hints at the fact that eq-restricted WTAh are tailored to represent homomorphic images of regular tree series. For an illustration of this feature, consider the following example.

Example 7. *Let $\Sigma = \{a^{(0)}, g^{(1)}, f^{(1)}\}$ and $A\colon T_\Sigma \to \mathbb{N}$ defined for every $s \in T_\Sigma$ by*

$$A(s) = \begin{cases} 2^n & \text{if } s = f\big(g^n(a)\big) \\ 0 & \text{else.} \end{cases}$$

A simple WTA recognizing the tree series A is $\mathscr{A} = \big(\{q, q_f\}, \Sigma, \{q_f\}, R, \mathrm{wt}\big)$ with the rules and weights $R = \big\{a \to_1 q,\ g(q) \to_2 q,\ f(q) \to_1 q_f\big\}$. Consider $\Delta = \{a^{(0)}, g^{(1)}, k^{(2)}\}$ and the input-finitary tree homomorphism $h\colon T_\Sigma \to T_\Delta$ induced by the mapping $h(a) = a$, $h(g) = g(x_1)$ and $h(f) = k\big(x_1, g(x_1)\big)$. The homomorphic image h_A is the tree series given for all $t \in T_\Delta$ by

$$h_A(t) = \begin{cases} 2^n & \text{if } t = k\big(g^n(a), g^{n+1}(a)\big) \\ 0 & \text{else.} \end{cases}$$

The natural eq-restricted WTAh that recognizes h_A is $\mathscr{A}' = \big(\{q, q_f, \perp\}, \Delta, \{q_f\}, R', \mathrm{wt}'\big)$ from Example 4 with

$$\begin{aligned} R' = \big\{\, & a \to_1 q, \quad g(q) \to_2 q, \quad k\big(q, g(\perp)\big) \xrightarrow{1=21}_1 q_f \,\big\} \\ \cup \big\{\, & a \to_1 \perp, \quad g(\perp) \to_1 \perp, \quad k(\perp, \perp) \to_1 \perp \,\big\}\,. \end{aligned}$$

The new rules in R' are obtained from the rules in R by applying the tree homomorphism to their left-hand sides. The duplicated subtree below k targets the sink state $\perp$ instead of q to avoid distorting the weight with an additional factor 2^n.

More formally, the following statement was shown in [22]. We include a condensed version of the proof as we will refer to a technical detail below.

Lemma 8 (cf. [22, Theorem 5])**.** *Let $\mathbb{S}$ be a commutative semiring, $\mathscr{A} = \big(Q,\Sigma,F,R,\mathrm{wt}\big)$ a WTA over $\mathbb{S}$ and $h\colon T_\Sigma \to T_\Delta$ a nondeleting and nonerasing tree homomorphism. There is an eq-restricted WTAh $\mathscr{A}'$ that recognizes $h_{[\![\mathscr{A}]\!]}$.*

Proof. An eq-restricted WTAh $\mathscr{A}'$ for $h_{[\![\mathscr{A}]\!]}$ is constructed in two stages.

First, we define $\mathscr{A}'' = \big(Q\dot\cup\{\perp\},\Delta\cup\Delta\times R,F'',R'',\mathrm{wt}''\big)$ such that for every $r = \sigma(q_1,\dots,q_k) \to_{\mathrm{wt}(r)} q$ in R and $h(\sigma) = u = \delta(u_1,\dots,u_n)$, we include

$$r'' \;=\; \Big(\langle\delta,r\rangle(u_1,\dots,u_n)[\![q_1,\dots,q_k]\!] \xrightarrow{E}_{\mathrm{wt}''(r'')} q\Big) \;\in\; R'' \qquad \text{with} \qquad E \;=\; \bigcup_{i\in[k]} \mathrm{pos}_{x_i}(u)^2$$

where the substitution $\langle\delta,r\rangle(u_1,\dots,u_n)[\![q_1,\dots,q_k]\!]$ replaces for every $i\in[k]$ only the $\le_{\mathrm{lex}}$-minimal occurrence of x_i in $\langle\delta,r\rangle(u_1,\dots,u_n)$ by q_i and all other occurrences by $\perp$. We set $\mathrm{wt}''(r'') = \mathrm{wt}(r)$. Additionally, we let $r''_\delta = \delta(\perp,\dots,\perp) \to \perp \in R''$ with $\mathrm{wt}''(r''_\delta) = 1$ for every $k\in\mathbb{N}$ and $\delta\in\Delta_k$. No other productions are in R''. Finally, we let $F'' = F$.

We can now delete the annotation: We use a deterministic relabeling to remove the second components of labels of $\Delta\times R$, adding up the weights of now identical rules. Since hom-regular languages are closed under relabelings [22, Theorem 4], we obtain an eq-restricted WTAh $\mathscr{A}' = \big(Q\dot\cup\{\perp\},\Delta,F',R',\mathrm{wt}'\big)$ recognizing $h_{[\![\mathscr{A}]\!]}$. □

The WTAh constructed for the homomorphic image of a WTA preserves the original state behaviour in its leading copies of duplicated subtrees. Using the notation from the proof of Lemma 8, we want to define a mapping that traces the runs of the input WTA to its homomorphic image.

Definition 9. *Let $\mathscr{A}$, h and $\mathscr{A}'$ be as in Lemma 8, let $r = \sigma(q_1,\dots,q_k) \to q \in R$ and $h(\sigma) = \delta(u_1,\dots,u_n)$. We let $h^R(r)$ be the rule $\delta(u_1,\dots,u_n)[\![q_1,\dots,q_k]\!] \xrightarrow{E} q$ of the WTAh $\mathscr{A}'$.*

The assignment h^R extends naturally to the runs of $\mathscr{A}$: For a run of the form $\rho = r = (\alpha\to q)$ with $\alpha\in\Sigma^0$, we set $h^R(\rho) = h^R(r)$. For a run of $\mathscr{A}$ of the form $\rho = r(\rho_1,\dots,\rho_k)$ with $r = \sigma(q_1,\dots,q_k)\to q$ and $h(\sigma) = \delta(u_1,\dots,u_n)$ we set

$$h^R(\rho) = \big(h^R(r)\big)(u_1,\dots,u_n)[\![h^R(\rho_1),\dots,h^R(\rho_k)]\!]\;;$$

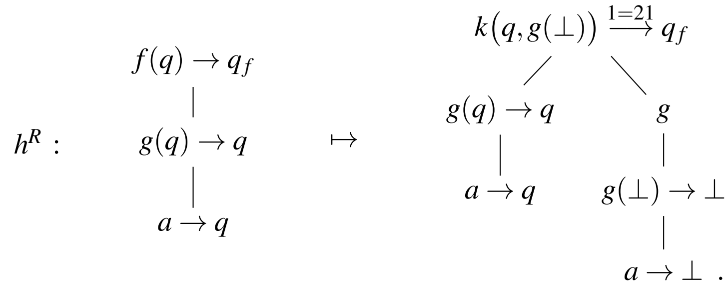
here, the substitution $[\![h^R(\rho_1),\dots,h^R(\rho_k)]\!]$ replaces for every $i\in[k]$ only the $\le_{lex}$-minimal occurrence of x_i in $\big(h^R(r)\big)(u_1,\dots,u_n)$ by $h^R(\rho_i)$ and all other occurrences by the respective unique run to $\perp$ for the unique tree that satisfies the constraint E.

Using the notation above, the assignment $h^R\colon R\to R'$ is well-defined, but not necessarily injective, and its image is $h^R(R) = \{r'\in R' \mid r' \text{ targets some } q\ne\perp\}$. Let us see how it acts on our running example.

Example 10. *Recall the WTA $\mathscr{A}$ and WTAh $\mathscr{A}'$ from Example 7. The mapping h^R assigns*

$$h^R:\quad f(q)\to q_f \quad\mapsto\quad k\big(q,g(\perp)\big) \xrightarrow{1=21} q_f\,,$$

and for the unique run of $\mathscr{A}$ for the tree $f\big(g(a)\big)$, it is



When discussing the behaviour of a WTAh $\mathscr{A}$, we often argue with the help of runs ρ, so it is a nuisance that we might have $\mathrm{wt}(\rho) = 0$. This anomaly can occur even if $\mathrm{wt}(r) \neq 0$ for all rules r of $\mathscr{A}$ due to the presence of zero-divizors, that is, elements $s, s' \in \mathbb{S} \setminus \{0\}$ such that $s \cdot s' = 0$. Fortunately, we can avoid this altogether using a construction of [18], which is based on DICKSON's Lemma [4]. It was first lifted to tree automata in [7] and later to WTAh in [21, 22]. Here, we slightly adjust the proof of Lemma 3 in [22] such that it preserves the eq-restriction of the input WTAh.

Lemma 11. *(cf. [22, Lemma 3]) Let $\mathbb{S}$ be a commutative semiring. For every eq-restricted WTAh $\mathscr{A}$ over $\mathbb{S}$ there exists an eq-restricted WTAh $\mathscr{A}'$ equivalent to $\mathscr{A}$ such that $\mathrm{wt}_{\mathscr{A}'}(\rho') \neq 0$ for all runs ρ' of $\mathscr{A}'$. For each $t \in \mathrm{supp}[\![\mathscr{A}]\!]$, the accepting (i.e. of non-zero weight and targeting a final state) runs of $\mathscr{A}$ for t translate bijectively into the accepting runs ρ' of $\mathscr{A}'$ for t, and the weights are preserved.*

Proof. Let $\mathscr{A}$ be the eq-restricted WTAh $\big(Q \dot\cup \{\bot\}, \Sigma, F, R, \mathrm{wt}\big)$. Obviously, $(\mathbb{S}, \cdot, 1, 0)$ is a commutative monoid with zero. Let $(s_1, \ldots, s_n)$ be an enumeration of the finite set $\mathrm{wt}(R) \setminus \{1\} \subseteq \mathbb{S}$. We consider the monoid homomorphism $h \colon \mathbb{N}^n \to \mathbb{S}$, which is given for every $m_1, \ldots, m_n \in \mathbb{N}$ by

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

According to DICKSON's lemma [4], the set $\min\big(h^{-1}(0)\big)$ is finite, where the partial order is the standard pointwise order on $\mathbb{N}^n$. Hence there is $u \in \mathbb{N}$ such that $\min\big(h^{-1}(0)\big) \subseteq \{0, \ldots, u\}^n = U$. 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 eq-restricted WTAh $\mathscr{A}' = \big(Q' \dot\cup \{\bot\}, \Sigma, F', R', \mathrm{wt}'\big)$ such that $Q' = Q \times V$, $F' = F \times V$, and R' and wt' are given as follows. Consider a rule $r = \ell \xrightarrow{E} q \in R$, let $\mathrm{pos}_Q(\ell) = \{p_1, \ldots, p_k\}$ ordered lexicographically and let $q_i = \ell(p_i)$ for all $i \in [k]$. Note that we do not consider the leaves of ℓ that are labeled by $\bot$. For all choices of $v_1, \ldots, v_k \in V$ such that the value $v = 1_{\mathrm{wt}(r)} \oplus \bigoplus_{i=1}^{k} v_i$ is again in V, the production

$$\ell[\langle q_1, v_1\rangle]_{p_1} \ldots [\langle q_k, v_k\rangle]_{p_k} \xrightarrow{E} \langle q, v\rangle$$

belongs to R' and its weight is $\mathrm{wt}'(p') = \mathrm{wt}(r)$. No further rules are in R'.

By annotating the power vectors v_i to the states $q \neq \bot$, we suitably (for the purpose of zero-divisors) track the weight of runs as v. If attaching another rule adopted from R to so far valid runs of $\mathscr{A}'$ would evaluate the overall weight to zero, then we exclude this rule from R'. Consequently, every run of $\mathscr{A}'$ is valid. To preserve the eq-restriction, we only annotate power vectors v_i to the non-sink states. It is safe to omit $\bot$ in this construction since $\bot$ only ever processes the neutral weight 1. $\square$

From here on, we silently assume that each WTAh avoids zero-divizors.

A main result proved in this article is deciding regularity for unambiguous WTAh over zero-sum free commutative semirings. We achieve this by reducing the problem to the unweigted (i.e. boolean) case solved in [12]. For this, we must relate our WTAh model to the *tree automata with HOM equality constraints* used by [12] which differ slightly from our WTAh over the boolean semiring. Fortunately, the two are closely related and the translation is rather simple: We merely eliminate the sink state and drop the weight assignment.

Lemma 12. *Let $\mathbb{S}$ be a commutative semiring and $\mathscr{A} = \big(Q\dot\cup\{\bot\}, \Sigma, F, R, \mathrm{wt}\big)$ an eq-restricted WTAh over $\mathbb{S}$. If $\mathscr{A}$ is unambiguous or $\mathbb{S}$ is zero-sum free, then there is a* tree automaton with HOM equality constraints (TA_{hom}) *[12] $\mathscr{A}^{\mathbb{B}}$ that recognizes the tree language* $\mathrm{supp}[\![\mathscr{A}]\!]$*. If $\mathscr{A}$ is a WTA (i.e. without constraints), then $\mathscr{A}^{\mathbb{B}}$ is also a TA without constraints.*

Proof. Let $q \in Q$ and consider a rule $\ell \xrightarrow{E}_a q$ of $\mathscr{A}$. Suppose that $\{p_1^1, \ldots, p_{n_1}^1\}$, $\ldots$,$\{p_1^m, \ldots, p_{n_m}^m\}$ are the equivalence classes of E, and wlog. let p_1^i be the unique representative such that $\ell(p_1^i) \neq \bot$ for each $i \in [m]$. Then we include the unweighted rule

$$\ell\,[\ell(p_1^1)]_{p_2^1} \cdots [\ell(p_1^1)]_{p_{n_1}^1} \;\cdots\; [\ell(p_1^m)]_{p_2^m} \cdots [\ell(p_1^m)]_{p_{n_m}^m} \xrightarrow{E} q$$

in $R^{\mathbb{B}}$, that is, we replace every occurrence of $\bot$ by the unique state from Q that labels a related position. This is necessary because the definition of TA_{hom} requires E-related positions to be labelled with the same state. We proceed this way for every rule of $\mathscr{A}$, discarding the rules that target $\bot$, and obtain the (unweighted) TA_{hom} $\mathscr{A}^{\mathbb{B}} = \big(Q, \Sigma, F, R^{\mathbb{B}}\big)$. Since $\mathscr{A}$ avoids zero-divizors, the conditions in the statement are each sufficient to ensure that $t \in \mathrm{supp}[\![\mathscr{A}]\!]$ iff. there exists an run of $\mathscr{A}$ for t to a final state, so $\mathscr{A}^{\mathbb{B}}$ recognizes $\mathrm{supp}[\![\mathscr{A}]\!]$. ☐

Example 13. *Reconsider the WTAh $\mathscr{A}'$ from Example 7. To obtain the TA_{hom} $(\mathscr{A}')^{\mathbb{B}}$, we remove the sink state $\bot$, all rules that target $\bot$ and the weight assignment, and replace the rule $k\big(q, g(\bot)\big) \xrightarrow{1=21}_1 q_f$ with the unweighted rule $k\big(q, g(q)\big) \xrightarrow{1=21} q_f$.*

3 Deciding Regularity for Unambiguous WTAh

In this section, we prove that regularity is decidable for unambiguous eq-restricted WTAh over zero-sum free semirings. To this end, we reduce this problem to regularity in the unweighted case, which was proved decidable in [12].

We begin by defining the *linearization* of eq-restricted WTAh, which was introduced for the boolean case in [12] and adapted to the weighted model in [23]. The linearization of a WTAh $\mathscr{A}$ by the number $\mathtt{h}$ is a WTG $\mathtt{lin}(\mathscr{A}, \mathtt{h})$ that approximates $\mathscr{A}$: It simulates all runs of $\mathscr{A}$ which only enforce the equality of subtrees of height at most $\mathtt{h}$. This is achieved by instantiating the constrained Q-positions of every rule $\ell \xrightarrow{E} q$ in $\mathscr{A}$ with compatible trees of height at most $\mathtt{h}$, while the Q-positions of ℓ that are unconstrained by E remain unchanged.

Formally, the linearization is defined following [12, Definition 7.1].

Definition 14 (cf. [23, Definition 12])**.** *Let $\mathbb{S}$ be a commutative semiring. Consider an eq-restricted WTAh $\mathscr{A} = (Q\dot\cup\{\bot\}, \Sigma, F, R, \mathrm{wt})$ over $\mathbb{S}$, and let $\mathtt{h} \in \mathbb{N}$ be a nonnegative integer. The* linearization of $\mathscr{A}$ by $\mathtt{h}$ *is the WTG $\mathtt{lin}(\mathscr{A}', \mathtt{h}) = \big(Q, \Sigma, F, R_{\mathtt{h}}, \mathrm{wt}_{\mathtt{h}}\big)$, where $R_{\mathtt{h}}$ and $\mathrm{wt}_{\mathtt{h}}$ are defined as follows.*

For $\ell' \in T_\Sigma(Q\dot\cup\{\bot\})$ and $q \in Q$, we include the rule $(\ell' \to q)$ in $R_{\mathtt{h}}$ iff. there exist a rule $(\ell \xrightarrow{E} q) \in R$, positions $p_1,\dots,p_k \in \mathrm{pos}_{Q\dot\cup\{\bot\}}(\ell)$, and trees $t_1,\dots,t_k \in T_\Sigma$ such that

- $\{p_1,\dots,p_k\} = \bigcup_{p\in \mathrm{pos}_\bot(\ell)}[p]_E$*, that is, E constrains exactly the positions $p_1,\dots,p_k$,*
- $(p_i,p_j) \in E$ *implies* $t_i = t_j$ *for all* $i,j \in [k]$,
- $\ell' = \ell[t_1]_{p_1}\cdots[t_k]_{p_k}$*, and*
- $\mathrm{wt}^{\ell(p_i)}(t_i) \neq 0$ *and* $\mathrm{ht}(t_i) \leq \mathtt{h}$ *for all* $i \in [k]$.

For every such production $\ell' \to q$ we set $\mathrm{wt}_{\mathtt{h}}(\ell' \to q)$ as the sum over all weights

$$\mathrm{wt}(\ell \xrightarrow{E} q)\cdot \prod_{i\in[k]} \mathrm{wt}^{\ell(p_i)}(t_i)$$

for all $(\ell \xrightarrow{E} q) \in R$, $p_1,\dots,p_k \in \mathrm{pos}_{Q\dot\cup\{\bot\}}(\ell)$ and $t_1,\dots,t_k \in T_\Sigma$ as above.

Note that the linearization is a WTG without constraints, so it recognizes a regular tree series. Let us apply this construction to our running example.

Example 15. *We recall the WTAh $\mathscr{A}'$ from Example 7 and set $\mathtt{h}=2$. The linearization of $\mathscr{A}'$ by 2 instantiates every constrained position by compatible trees of maximal height 2, keeping track of the weights, and removes $\bot$ and the rules that target it. More precisely, $\mathtt{lin}(\mathscr{A}',2) = \big(\{q,q_f\},\Delta,\{q_f\},R_2,\mathrm{wt}_2\big)$ with the set of rules and weights*

$$\begin{aligned} R_2 \;=\; \big\{\, & a \to_1 q, \qquad g(q) \to_2 q, \qquad k\big(a,g(a)\big) \to_1 q_f, \\ & k\big(g(a),g(g(a))\big) \to_2 q_f, \qquad k\big(g(g(a)),g(g(g(a)))\big) \to_4 q_f \,\big\}. \end{aligned}$$

This example illustrates that the larger we choose $\mathtt{h}$, the better $\mathtt{lin}(\mathscr{A}',\mathtt{h})$ approximates $[\![\mathscr{A}']\!]$. In this particular case however, there will always be a tree t such that $[\![\mathscr{A}']\!](t) \neq [\![\mathtt{lin}(\mathscr{A}',\mathtt{h})]\!](t)$, say, the tree $k\big(g^{\mathtt{h}+1}(a),g^{\mathtt{h}+2}(a)\big)$. For eq-restricted WTAh $\mathscr{A}$ over $\mathbb{B}$ or $\mathbb{N}$ it is known [12, 22] that $[\![\mathscr{A}]\!]$ is regular iff. $[\![\mathtt{lin}(\mathscr{A},\mathtt{h})]\!] = [\![\mathscr{A}]\!]$ for a certain parameter $\mathtt{h}$. For other semirings, a customized investigation is necessary, but unambiguous WTAh allow us to decide regularity by applying the boolean case directly. To this end, the following lemma is fundamental.

Lemma 16. *Let $\mathbb{S}$ be a commutative semiring, $\mathscr{A}$ an eq-restricted WTAh over $\mathbb{S}$ and $h \in \mathbb{N}$. For each $t \in \mathrm{supp}[\![\mathscr{A}]\!]$, there are at most as many accepting runs of $\mathtt{lin}(\mathscr{A},\mathtt{h})$ for t as there are accepting runs of $\mathscr{A}$ for t. In particular, if $\mathscr{A}$ is unambiguous, then so is its linearization, and for every $t \in \mathrm{supp}[\![\mathscr{A}]\!]$ it is either $[\![\mathtt{lin}(\mathscr{A},\mathtt{h})]\!](t) = [\![\mathscr{A}]\!](t)$, or there are no accepting runs of $\mathtt{lin}(\mathscr{A},\mathtt{h})$ for t.*

Proof. The linearization $\mathtt{lin}(\mathscr{A},\mathtt{h})$ is defined in such a way that it simulates every run ρ of $\mathscr{A}$ with the following property: Say ρ processes $t \in T_\Sigma$, then for every rule $\ell \xrightarrow{E} q$ used in ρ at position p (that is, at the root of $t|_p$), and for every position $\bar{p}$ constrained by E, it is $\mathrm{ht}(t|_{p\bar{p}}) \leq \mathtt{h}$. Different runs of $\mathscr{A}$ might be merged into the same run of $\mathtt{lin}(\mathscr{A},\mathtt{h})$, but for a particular run of $\mathscr{A}$ it is uniquely determined which run of $\mathtt{lin}(\mathscr{A},\mathtt{h})$ will incorporate it. □

We need yet another technical ingredient for the reduction to the unweighted case: to interchange the linearization of a WTAh and its projection onto the boolean TA_{hom}. The linearization for TA_{hom} was defined in [12, Definition 7] and indeed, the following holds.

Lemma 17. *Consider an unambiguous, eq-restricted WTAh $\mathscr{A}$ over a commutative semiring. Let $\mathscr{A}^{\mathbb{B}}$ the TA_{hom} for $\mathrm{supp}[\![\mathscr{A}]\!]$ defined in Lemma 12 and $\mathrm{linearize}(\mathscr{A}^{\mathbb{B}},\mathtt{h})$ in turn the linearization of $\mathscr{A}^{\mathbb{B}}$ by $\mathtt{h}$ as introduced in [12, Definition 7.1]. Then it is $\mathtt{lin}(\mathscr{A},\mathtt{h})^{\mathbb{B}} = \mathrm{linearize}(\mathscr{A}^{\mathbb{B}},\mathtt{h})$.*

We are now ready for the main result of this section: the reduction of regularity for eq-restricted WTAh over zero-sum free semirings to the unweighted case.

Theorem 18. *Let $\mathbb{S}$ be a zero-sum free commutative semiring and $\mathscr{A}$ an unambiguous eq-restricted WTAh over $\mathbb{S}$. The tree series $[\![\mathscr{A}]\!]$ is regular iff.* $\mathrm{supp}[\![\mathscr{A}]\!]$ *is a regular tree language.*

Proof. Suppose first that $[\![\mathscr{A}]\!]$ is regular, thus there is a WTA $\mathscr{B}$ equivalent to $\mathscr{A}$. Since $\mathbb{S}$ is zero-sum free, we can apply Lemma 12 to $\mathscr{B}$ and obtain that $\mathrm{supp}[\![\mathscr{B}]\!] = \mathrm{supp}[\![\mathscr{A}]\!]$ is regular.

Next, suppose that $[\![\mathscr{A}]\!]$ is not regular. In particular, the regular WTG $\mathtt{lin}(\mathscr{A},\mathtt{h})$ is not equivalent to $\mathscr{A}$ for any $h \in \mathbb{N}$. Thus by Lemma 16, it is $\mathrm{supp}[\![\mathscr{A}]\!] \neq \mathrm{supp}[\![\mathtt{lin}(\mathscr{A},\mathtt{h})]\!]$. By Lemma 12, $\mathtt{lin}(\mathscr{A},\mathtt{h})^{\mathbb{B}}$ recognizes the regular language $\mathrm{supp}\,[\![\mathtt{lin}(\mathscr{A},\mathtt{h})]\!]$, and together with Lemma 17, it is

$$[\![\mathscr{A}^{\mathbb{B}}]\!] = \mathrm{supp}[\![\mathscr{A}]\!] \neq [\![\mathtt{lin}(\mathscr{A},\mathtt{h})^{\mathbb{B}}]\!] = [\![\mathit{linearize}(\mathscr{A}^{\mathbb{B}},\mathtt{h})]\!],$$

that is, the boolean linearization of the TA_{hom} $\mathscr{A}^{\mathbb{B}}$ is not equivalent to it for any $h \in \mathbb{N}$. This, however, implies that $[\![\mathscr{A}^{\mathbb{B}}]\!] = \mathrm{supp}[\![\mathscr{A}]\!]$ is not regular as proved in [22, Lemma 7.3]. ☐

Note that we only used zero-sum freeness of the semiring for the first part of the statement, as Lemma 12 holds for unambiguous WTAh over arbitrary commutative semirings. With this result, regularity of eq-restricted WTAh is decidable.

Corollary 19. *Let $\mathbb{S}$ be a zero-sum free commutative semiring. Given an unambiguous eq-restricted WTAh $\mathscr{A}$ over $\mathbb{S}$ as input, it is decidable whether $[\![\mathscr{A}]\!]$ is regular.*

Proof. By Theorem 18, $[\![\mathscr{A}]\!]$ is regular iff. $\mathrm{supp}[\![\mathscr{A}]\!]$ is regular. A TA_{hom} recognizing the latter can be constructed with Lemma 12, for which, in turn, regularity is decidable [12, Section 7]. ☐

4 A Sufficient Condition and the HOM-Problem

So far, the assumption we make for deciding regularity is imposed on the WTAh. Meanwhile the HOM-problem has a WTA $\mathscr{A}$ and a tree homomorphism h as input. In this section, we propose conditions on $\mathscr{A}$ and h which ensure that the strategy of the previous section is applicable to the corresponding instance of the HOM-problem. We begin with a condition for h which generalizes injectivity.

Definition 20. *Let Σ and Δ be ranked alphabets and $h\colon T_\Sigma \to T_\Delta$ a nondeleting and nonerasing tree homomorphism. We call h* tetris-free *if for all $s,s' \in T_\Sigma$ with $h(s) = h(s')$, it is* $\mathrm{pos}(s) = \mathrm{pos}(s')$ *and for all $p \in \mathrm{pos}(s)$, we have $h\big(s(p)\big) = h\big(s'(p)\big)$.*

In other words, $h\colon T_\Sigma \to T_\Delta$ is tetris-free if we cannot combine the building blocks $h(\sigma)$, $\sigma \in \Sigma$ in different ways to build the same tree. In contrast, Figure 1 below shows the well-known *Tetriminos*®© [19] violating (and thus naming) the tetris-free condition.

Let us discuss a short example and counter-example.

Example 21. *Let $\Sigma = \{a^{(0)}, b^{(0)}, g^{(1)}\}$ and $\Delta = \{c^{(0)}, k^{(2)}\}$. Consider the tree homomorphism $h\colon T_\Sigma \to T_\Delta$ induced by $h(a) = h(b) = c$ and $h(g) = k(x_1,x_1)$. While h is not injective, it is tetris-free. However, the tree homomorphism $\hat{h}\colon T_\Sigma \to T_\Delta$ induced by $\hat{h}(a) = c$, $\hat{h}(b) = k(c,c)$ and $\hat{h}(g) = k(x_1,c)$ is not: The trees $g(a)$ and b violate the tetris-free condition.*

Intuitively, if a tree homomorphism h is tetris-free, then any non-injective behaviour of h is located entirely at the symbol level. This allows the construction of the WTAh to cancel the non-injectivity of h. For this, however, we also need to make an assumption on the input WTA $\mathscr{A}$, which leads us to this augmented version of unambiguity for $\mathscr{A}$.

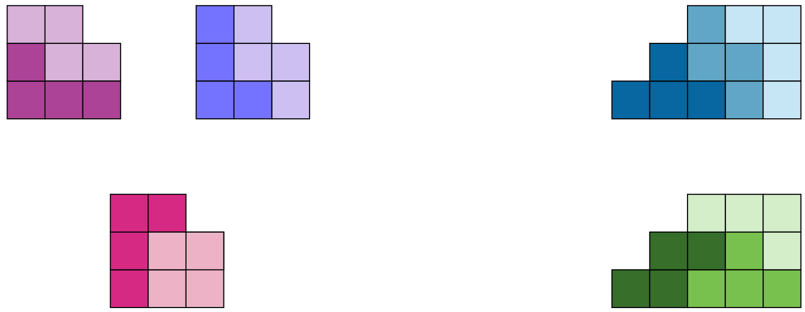


Figure 1: The game of *Tetris*®© [19] being non-tetris-free by nature.

Definition 22. *Let $\mathscr{A}$ be a WTA over a commutative semiring $\mathbb{S}$ and Σ, and $h\colon T_\Sigma \to T_\Delta$ a nondeleting and nonerasing, tetris-free tree homomorphism. We say that $\mathscr{A}$ is* h-unambiguous *if for all trees $s, s' \in T_\Sigma$ such that $h(s) = h(s')$, all accepting runs ρ, ρ' of $\mathscr{A}$ for s and s', respectively, and all $p \in \text{pos}(s)$, the target states of the rules applied in ρ and ρ' at p, respectively, coincide.*

Remark 23. *The condition of h-unambiguity is stronger than unambiguity: For $s = s' \in \text{supp}[\![\mathscr{A}]\!]$ we obtain that $\mathscr{A}$ has at most one accepting run for s (since runs of WTA are uniquely determined by the processed symbol and the target state at every position). A similar reasoning applies if we choose $s \neq s'$ with $h(s) = h(s')$: While the unique runs of $\mathscr{A}$ for s and s' may read different symbols, the states they pass through coincide at every position.*

Imposing these conditions on the input of the HOM-problem allows us to build on it with the arguments from the previous section.

Proposition 24. *Let $\mathscr{A}$ be a WTA over a commutative semiring $\mathbb{S}$ and Σ, and $h\colon T_\Sigma \to T_\Delta$ a nondeleting and nonerasing tree homomorphism. If h is tetris-free and $\mathscr{A}$ is h-unambiguous, then the eq-restricted WTAh $\mathscr{A}'$ for $h_{[\![\mathscr{A}]\!]}$ constructed in Lemma 8 is unambiguous.*

Proof. Let ϑ and ϑ' be accepting runs of $\mathscr{A}'$ for the same $t \in T_\Delta$. We prove the statement by contradiction, so assume that $\vartheta \neq \vartheta'$. Then there are two distinct runs ρ and ρ' of $\mathscr{A}$ such that $\vartheta = h^R(\rho)$ and $\vartheta' = h^R(\rho')$ as introduced in Definition 9. The mapping h^R does not modify the target states of runs, so both ρ and ρ' are accepting as well, and since $\mathscr{A}$ is unambiguous, they must process distinct trees s and s', respectively, with $h(s) = h(s')$. By the premises of the statement, at every $p \in \text{pos}(s) = \text{pos}(s')$ it is $h\big(s(p)\big) = h\big(s'(p)\big)$, and the target states of ρ and ρ' at p coincide, so although $\rho \neq \rho'$, after applying h it is $\vartheta = h^R(\rho) = h^R(\rho') = \vartheta'$, which contradicts our assumption. □

We want to illustrate the role played by our two conditions, the h-unambiguity and the tetris-freeness. Let us discuss this with the help of two counter-examples.

Example 25. *Consider the ranked alphabets $\Sigma = \{a^{(0)}, b^{(0)}, g^{(1)}\}$ and $\Delta = \{c^{(0)}, k^{(2)}\}$. Let $h\colon T_\Sigma \to T_\Delta$ be the tetris-free tree homomorphism from Example 21 induced by $h(a) = h(b) = c$ and $h(g) = k(x_1, x_1)$. Moreover, let $\mathscr{A} = \big(Q, \Sigma, Q, R, \text{wt}\big)$ be the WTA over the arctic semiring $\mathbb{A}$ with $Q = \{q_a, q_b\}$ and the following rules and weights:*

$$R = \{\, a \to_0 q_a\,, \quad b \to_0 q_b\,, \quad g(q_a) \to_1 q_a\,, \quad g(q_b) \to_2 q_b \,\}\,.$$

The WTA $\mathscr{A}$ is unambiguous, but not h-unambiguous, since the runs for a and b target different states despite $h(a) = h(b)$. Evaluating the weights in $\mathbb{A}$, we obtain the tree series $[\![\mathscr{A}]\!]$ defined by

$$[\![\mathscr{A}]\!] \,:\, s \mapsto \begin{cases} n & \text{if } s = g^n(a) \\ 2n & \text{if } s = g^n(b) \end{cases}$$

The WTAh $\mathscr{A}' = \big(Q \dot\cup \{\bot\}, \Delta, Q, R', \mathrm{wt}'\big)$ recognizing $h_{[\![\mathscr{A}]\!]}$ which is obtained from Lemma 8 has the following rules and weights:

$$R \;=\; \big\{\; c \to_0 q_a\,, \quad c \to_0 q_b\,, \quad k(q_a,\bot) \xrightarrow{1=2}_1 q_a\,, \quad k(q_b,\bot) \xrightarrow{1=2}_2 q_b \;\big\}$$
$$\cup\; \big\{\; c \to_0 \bot\,, \quad k(\bot,\bot) \to_0 \bot \;\big\}\,.$$

Because of the different target states, the rules $c \to q_a$ and $c \to q_b$ are not merged in $\mathscr{A}'$, therefore $\mathscr{A}'$ is not unambiguous.

On the other hand, let $\hat{h}$ be the homomorphism from Example 21 induced by $\hat{h}(a) = c$, $\hat{h}(b) = k(c,c)$ and $\hat{h}(g) = k(x_1,c)$. Recall that h is not tetris-free because $h\big(g(a)\big) = h(b)$ although $\mathrm{pos}\,\big(g(a)\big) \neq \mathrm{pos}(b)$. Moreover, consider the WTA $\hat{\mathscr{A}} = \big(\{q\}, \Sigma, \{q\}, \hat{R}, \hat{\mathrm{wt}}\big)$ over $\mathbb{N}$ with the following rules and weights:

$$\hat{R} \;=\; \big\{\; a \to_2 q\,, \quad b \to_3 q\,, \quad g(q) \to_1 q \;\big\}\,.$$

The WTA $\hat{\mathscr{A}}$ only has one state, so it is deterministic and thus unambiguous. It recognizes the tree series $[\![\hat{\mathscr{A}}]\!]$ defined by

$$[\![\hat{\mathscr{A}}]\!] \,:\, s \mapsto 2|\mathrm{pos}_a(s)| + 3|\mathrm{pos}_b(s)|\,.$$

However, the WTAh $\hat{\mathscr{A}}' = \big(\{q,\bot\}, \Delta, \{q\}, \hat{R}', \hat{\mathrm{wt}}'\big)$ for $h_{[\![\mathscr{A}]\!]}$ has the following rules and weights:

$$\hat{R}' \;=\; \big\{\; c \to_2 q\,, \quad k(c,c) \to_3 q\,, \quad k(q,c) \to_1 q \;\big\}$$
$$\cup\; \big\{\; c \to_1 \bot\,, \quad k(\bot,\bot) \to_1 \bot \;\big\}\,.$$

Since $\hat{h}$ performs no duplications, the rules targeting $\bot$ are not used in any accepting run, so we can safely ignore them. Although this time, no two rules of $\hat{\mathscr{A}}'$ (that are used in an accepting run) share a left-hand side, the tree $k(c,c) = \hat{h}\big(g(a)\big) = \hat{h}(b)$ still has two different runs, which stem directly from the non-tetris-freeness of $\hat{h}$.

As a concequence of Proposition 24, our restricted version of the HOM-problem is decidable.

Corollary 26. *Let $\mathbb{S}$ be a zero-sum free, commutative semiring. For a nondeleting and nonerasing, tetris-free tree homomorphism h and an h-unambiguous WTA $\mathscr{A}$ over $\mathbb{S}$ as input, it is decidable whether the tree series $h_{[\![\mathscr{A}]\!]}$ is regular.*

5 Conclusion and Future Work

Homomorphic images of regular tree series can be represented using an extension of WTA, the so-called *eq-restricted WTAh* [22]. In this paper, we have shown that regularity is decidable for unambiguous devices of this type over zero-sum free commutative semirings. For this, we reduced this question to the unweighted setting, where regularity is known to be decidable [12]. Moreover, we have phrased a condition on the input WTA $\mathscr{A}$ and tree homomorphism h that ensures unambiguity of the WTAh representing the image $h_{[\![\mathscr{A}]\!]}$. Thus the *HOM-problem* over zero-sum free semirings which, given $\mathscr{A}$ and h as input, asks whether $h_{[\![\mathscr{A}]\!]}$ is regular, is decidable if the input satisfies our condition.

Notably, the zero-sum freeness of the semiring is only used in Theorem 18 to show that if the tree series recognized by an unambiguous eq-restricted WTAh $\mathscr{A}$ is regular, then its support is also regular. It is plausible that the zero-sum freeness is not needed: Its purpose is to ensure that different accepting runs of $\mathscr{A}$ for the same tree t cannot cancel out, leaving $t \notin \operatorname{supp} [\![\mathscr{A}]\!]$ despite $\mathscr{A}$ having accepting runs for t. This, however, should not be a concern if $\mathscr{A}$ is unambiguous. To discard the zero-sum freeness assumption, it suffices to prove this simple statement: *If $\mathscr{A}$ is an unambiguous eq-restricted WTAh and $[\![\mathscr{A}]\!]$ is regular, then there is an unambiguous WTA equivalent to $\mathscr{A}$.* In fact, the linearization of $\mathscr{A}$, which is unambiguous by Lemma 16, is a promising candidate. Thus we conjecture that Theorem 18 holds for arbitrary commutative semirings, as do then Corollaries 19 – stating that regularity is decidable for unambiguous eq-restricted WTAh – and 26 – stating that the HOM-problem is decidable under our assumptions on $\mathscr{A}$ and h.

Recently, the disambiguation of weighted (string) automata from [24] was lifted to trees [30]. Here, the authors assume a variation of the *twins property* which restricts the behaviour of related states of a WTA. This allows them to construct an equivalent unambiguous WTA. A natural question is whether this proof can be adjusted to provide even an h-unambiguous WTA, say, by refining the twins property with respect to h. That way, we could lift our result to a larger class of input WTA.

References

[1] Symeon Bozapalidis & Antonios Kalampakas (2008): *Graph automata*. Theoretical Computer Science 393(1-3), pp. 147–165, doi:10.1016/j.tcs.2007.11.022.

[2] H. Comon, M. Dauchet, R. Gilleron, C. Löding, F. Jacquemard, D. Lugiez, S. Tison & M. Tommasi (2007): *Tree Automata — Techniques and Applications*.

[3] Carles Creus, Adrià Gascón, Guillem Godoy & Lander Ramos (2012): *The HOM problem is EXPTIME-complete*. In: Proc. 27th Annual IEEE Symp. Logic in Computer Science, IEEE, pp. 255–264, doi:10.1109/LICS.2012.36.

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

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

[6] Frank Drewes (2006): *Grammatical picture generation*. Springer.

[7] Manfred Droste & Doreen Heusel (2015): *The supports of weighted unranked tree automata*. Funda. Inform. 136(1–2), pp. 37–58, doi:10.3233/FI-2015-1143.

[8] Manfred Droste & Dietrich Kuske (2021): *Weighted automata.*

[9] Zoltán Fülöp & Heiko Vogler (2009): *Weighted tree automata and tree transducers*. In: Handbook of Weighted Automata, chapter 9, Springer, pp. 313–403, doi:10.1007/978-3-642-01492-5_9.

[10] Ferenc Gécseg & Magnus Steinby (2015): *Tree Automata*. Technical Report 1509.06233, arXiv.

[11] Dora Giammarresi & Antonio Restivo (1992): *Recognizable picture languages*. International Journal of Pattern Recognition and Artificial Intelligence 6(02n03), pp. 241–256, doi:10.1142/S021800149200014X.

[12] Guillem Godoy & Omer Giménez (2013): *The HOM problem is decidable*. J. ACM 60(4), pp. 1–44, doi:10.1145/2508028.2501600.

[13] Guillem Godoy, Omer Giménez, Lander Ramos & Carme Àlvarez (2010): *The HOM problem is decidable*. In: *Proc. 42nd ACM symp. Theory of Computing*, ACM, pp. 485–494, doi:10.1145/1806689.1806757.

[14] Jonathan S. Golan (1999): *Semirings and their Applications*. Kluwer Academic, Dordrecht, doi:10.1007/978-94-015-9333-5.

[15] Udo Hebisch & Hanns J. Weinert (1998): *Semirings*. World Scientific, doi:10.1142/3903.

[16] Udo Hebisch & Hanns Joachim Weinert (1998): *Semirings: algebraic theory and applications in computer science*. 5, World Scientific, doi:10.1142/9789812815965_bmatter.

[17] Dan Jurafsky & James H. Martin (2008): *Speech and language processing*. Prentice Hall.

[18] Daniel Kirsten (2011): *The support of a recognizable series over a zero-sum free, commutative semiring is recognizable*. *Acta Cybernet.* 20(2), pp. 211–221, doi:10.14232/actacyb.20.2.2011.1.

[19] Tetris Holding; The Tetris Company LLC (1985): *Tetris*. Available at https://tetris.com.

[20] Christof Loding & Wolfgang Thomas (2000): *Alternating Automata and Logics over Infinite Words*. In Jan van Leeuwen, Osamu Watanabe, Masami Hagiya, Peter D. Mosses & Takayasu Ito, editors: *Theoretical Computer Science: Exploring New Frontiers of Theoretical Informatics*, Springer Berlin Heidelberg, doi:10.1007/3-540-44929-9_36.

[21] Andreas Maletti & Andreea-Teodora Nász (2022): *Weighted tree automata with constraints*. In: *Developments in Language Theory: 26th International Conference, DLT 2022, Tampa, FL, USA, May 9–13, 2022, Proceedings*, Springer, pp. 226–238, doi:10.1007/978-3-031-05578-2_18.

[22] Andreas Maletti & Andreea-Teodora Nász (2023): *Weighted Tree Automata with Constraints*. Technical Report 2302.03434, arXiv. Available at https://arxiv.org/abs/2302.03434.

[23] Andreas Maletti, Andreea-Teodora Nász & Erik Paul (2023): *Weighted HOM-Problem for Nonnegative Integers*. arXiv:2305.04117.

[24] Mehryar Mohri & Michael D. Riley (2017): *A disambiguation algorithm for weighted automata*. *Theoretical Computer Science* 679, pp. 53–68, doi:10.1016/j.tcs.2016.08.019. Available at https://www.sciencedirect.com/science/article/pii/S0304397516304376. Implementation and Application of Automata.

[25] J. Mongy-Steen (1981): *Transformation de noyaux reconnaissables d'arbres. Forêts RATEG*. Ph.D. thesis, Université de Lille.

[26] Dominique Perrin (1984): *Recent results on automata and infinite words*. In: *Proc. 11th Int. Symp. Mathematical Foundations of Computer Science*, *LNCS* 176, Springer, pp. 134–148, doi:10.1007/BFb0030294.

[27] Azriel Rosenfeld (2014): *Picture languages: formal models for picture recognition*. Academic Press.

[28] Arto Salomaa & Matti Soittola (1978): *Automata-Theoretic Aspects of Formal Power Series*. Texts and Monographs in Computer Science, Springer, doi:10.1007/978-1-4612-6264-0.

[29] Marcel Paul Schützenberger (1961): *On the definition of a family of automata*. *Inform. and Control* 4(2–3), pp. 245–270, doi:10.1016/S0019-9958(61)80020-X.

[30] Kevin Stier & Markus Ulbricht (2021): *Disambiguation of Weighted Tree Automata*. In Yo-Sub Han & Sang-Ki Ko, editors: *Descriptional Complexity of Formal Systems*, Springer International Publishing, Cham, pp. 163–175, doi:10.1007/978-3-030-93489-7_14.

[31] James W. Thatcher (1967): *Characterizing Derivation Trees of Context-Free Grammars through a Generalization of Finite Automata Theory*. *J. Comput. Syst. Sci.* 1(4), pp. 317–322, doi:10.1016/S0022-0000(67)80022-9.

[32] James W. Thatcher & Jesse B. Wright (1968): *Generalized finite automata theory with an application to a decision problem of second-order logic*. *Math. Systems Theory* 2(1), pp. 57–81, doi:10.1007/BF01691346.

[33] Reinhard Wilhelm, Helmut Seidl & Sebastian Hack (2013): *Compiler Design - Syntactic and Semantic Analysis*. Springer, doi:10.1007/978-3-642-17540-4.