The Weighted HOM-Problem over Fields

Andreea-Teodora Nász

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

Abstract. 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. In this paper, we prove the *weighted* HOM-problem for all fields decidable, provided that the tree homomorphism is *tetris-free* (a condition that generalizes injectivity). To this end, we reduce the problem to a property of the device representing the homomorphic image in question; to prove this property decidable, we then derive a pumping lemma for such devices from the well-known pumping lemma for regular tree series over fields, proved by Berstel and Reutenauer in 1982.

1 Introduction

The well-known model of finite-state automata has seen various extensions over the past decades. On the one hand, the qualitative evaluation of these acceptors was generalized to a quantitative one, leading to *weighted automata* [28]. Such devices assign a weight to each input word, and are thus suited to model numerical factors related to the input, such as costs, probabilities and consumption of resources or time. The research community focused on automata theory has studied weighted automata consistently and fruitfully [9,10,27]. Thereby, the favoured domains for weight calculations are often semirings [16,18], as they are both quite general and computationally efficient due to their distributivity.

Another dimension of generalization for finite-state automata targets their input, allowing them to handle more complex data structures such as infinite words [24], trees [5], graphs [4] and pictures [26]. In particular, *finite-state tree automata* and the *regular tree languages* they recognize were introduced independently in [7,30,31]. These devices find applications in a variety of areas like natural language processing [19], picture generation [8] and compiler construction [32]. Unsurprisingly, combining both types of generalizations leads to intricate yet fruitful research areas, and so several variants of *weighted tree automata* (WTA) and the *regular tree series* they recognize continue to be studied [11].

Tree homomorphisms are widely used in the context of term rewriting [13] and XML types [29]. A tree homomorphism is a structure-preserving transformation on trees which can duplicate subtrees, so the trees in the homomorphic image might have identical subtrees. Unfortunately, tree automata have limited memory, so they cannot ensure that certain subtrees are equal [12] (much like the classical (string) automata cannot ensure that the numbers of a's and of b's in a word are equal). Therefore, unlike in the word case, regular tree languages are not closed under tree homomorphisms. It was a long-standing open question if, given a regular tree language $\mathcal{L}$ and a tree homomorphism h as input, it is decidable whether $h(\mathcal{L})$ is again regular. This *HOM-problem* was finally solved in [6,14,15], with the help of a well-studied extension called *tree automata with constraints*; these devices can explicitely require certain subtrees to be equal, and can thus handle the duplications performed by h.

In the *weighted* HOM-problem, a regular tree series and a tree homomorphism are given as input. By its nature, this question requires a customized investigation for different semirings. Most recently, this problem was proved decidable for different scenarios [22,23], but in both cases, the semiring must be zero-sum free; this strong condition already excludes essential rings such as $\mathbb{Z}$. In this paper, we decide the weighted HOM-problem for all fields (and thus, all subspaces of fields), provided that the tree homomorphism is *tetris-free*, a property that generalizes injectivity.

The paper is structured as follows: In Section 2 we represent the homomorphic image of the input tree series by a WTA with constraints (WTAh). In Section 4 we show that, if the input tree homomorphism is tetris-free, then the weighted HOM-problem is equivalent to a certain decidable property of this WTAh. Proving said decidability relies on a pumping lemma for WTAh over fields, which we derive in Section 3. from the well-known pumping lemma for (regular) WTA over fields [1]. Finally, we present an example that illustrates why the approach is entirely unsuited for non-tetris-free tree homomorphisms.

2 Preliminaries and Technical Background

We denote the set $\{0, 1, 2, \ldots\}$ 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 $|w|$.

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 by $T_\Sigma(Z)$, is the smallest set T that satisfies (i) $\Sigma_0 \cup Z \subseteq T$ and (ii) $\sigma(t_1, \ldots, t_k) \in T$ for every $k \in \mathbb{N}$, $\sigma \in \Sigma_k$ and $t_1, \ldots, t_k \in T$. 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 $\mathrm{pos}(t) \subseteq \mathbb{N}^*$ of *positions of t* is defined by $\mathrm{pos}(t) = \{\varepsilon\}$ for every $t \in \Sigma_0 \cup Z$, and by $\mathrm{pos}\big(\sigma(t_1, \ldots, t_k)\big) = \{\varepsilon\} \cup \bigcup_{i \in [k]} \{ip \mid p \in \mathrm{pos}(t_i)\}$ for all $k \in \mathbb{N}$, $\sigma \in \Sigma_k$ and $t_1, \ldots, t_k \in T_\Sigma(Z)$. The set of positions of t inherits the lexicographic order $\leq_{\mathrm{lex}}$ from $\mathbb{N}^*$. The *size* $|t|$ *of* t is defined by $|t| = |\mathrm{pos}(t)|$ and the *height* $\mathrm{ht}(t)$ *of* t by $\mathrm{ht}(t) = \max_{p \in \mathrm{pos}(t)} |p|$. For $p \in \mathrm{pos}(t)$, 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 $t = \sigma(t_1, \ldots, t_k)$ by $t(\varepsilon) = \sigma$, $t(ip') = t_i(p')$, $t|_\varepsilon = t$, $t|_{ip'} = t_i|_{p'}$, $t[t']_\varepsilon = t'$, and finally $t[t']_{ip'} = \sigma(t_1, \ldots, t_{i-1}, t_i[t']_{p'}, t_{i+1}, \ldots, t_k)$ for every $k \in \mathbb{N}$, $\sigma \in \Sigma_k$,

$t_1, \ldots, t_k \in T_\Sigma(Z)$, $i \in [k]$ and $p' \in \mathrm{pos}(t_i)$. For every subset $S \subseteq \Sigma \cup Z$, we let $\mathrm{pos}_S(t) = \{p \in \mathrm{pos}(t) \mid t(p) \in S\}$ and we abbreviate $\mathrm{pos}_{\{s\}}(t)$ by $\mathrm{pos}_s(t)$ for every $s \in \Sigma \cup Z$. Let $X = \{x_1, x_2, \ldots\}$ be a fixed, countable set of formal variables. For $k \in \mathbb{N}$ we denote by X_k the subset $\{x_1, \ldots, x_k\}$. For any $t \in T_\Sigma(X)$ we let $\mathrm{var}(t) = \{x \in X \mid \mathrm{pos}_x(t) \neq \emptyset\}$. 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 $\sigma(t_1, \ldots, t_k)\theta = \sigma\big(t_1\theta, \ldots, t_k\theta\big)$ for all $k \in \mathbb{N}$, $\sigma \in \Sigma_k$ and $t_1, \ldots, t_k \in T_\Sigma(Z)$. If $V = \{v_1, \ldots, v_n\}$, we write θ explicitly as $[v_1 \leftarrow \theta(v_1), \ldots, v_n \leftarrow \theta(v_n)]$, or simply as $[\theta(x_1), \ldots, \theta(x_n)]$ if $V = X_n$.

A *(commutative) semiring* [16,17] is a tuple $(\mathbb{S}, +, \cdot, 0, 1)$ that satisfies the following conditions: $(\mathbb{S}, +, 0)$ and $(\mathbb{S}, \cdot, 1)$ are commutative monoids, $\cdot$ distributes over $+$, and $0 \cdot s = 0$ for all $s \in \mathbb{S}$. Examples include $\mathbb{N} = \big(\mathbb{N}, +, \cdot, 0, 1\big)$, $\mathbb{Z} = \big(\mathbb{Z}, +, \cdot, 0, 1\big)$, $\mathbb{Q} = \big(\mathbb{Q}, +, \cdot, 0, 1\big)$, the Boolean semiring $\mathbb{B} = \big(\{0, 1\}, \vee, \wedge, 0, 1\big)$ 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 $\mathbb{S}$. A semiring is a *field* if it is (i) a *ring*, i.e. there exists $-1 \in S$ such that $1 + (-1) = 0$, and (ii) a *semifield*, i.e. for every $a \in \mathbb{S} \setminus \{0\}$ there exists a multiplicative inverse a^{-1} such that $a \cdot a^{-1} \in \mathbb{S}$. Let $\mathbb{F}$ be a field, then $\mathbb{S}$ is a *subsemiring* of $\mathbb{F}$ if $\mathbb{S} \subseteq \mathbb{F}$ and the operations of $\mathbb{S}$ are embeddable in $\mathbb{F}$, i.e., $(+_\mathbb{F})|_\mathbb{S} = +_\mathbb{S}$, $(\cdot_\mathbb{F})|_\mathbb{S} = \cdot_\mathbb{S}$, $0_\mathbb{S} = 0_\mathbb{F}$ and $1_\mathbb{S} = 1_\mathbb{F}$. The semirings $\mathbb{N}$ and $\mathbb{Z}$ are subsemirings of $\mathbb{Q}$, but not $\mathbb{B}$.

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* over $\mathbb{S}$, and its *support* is the set $\mathrm{supp}(\varphi) = \{t \in T_\Sigma(Z) \mid \varphi(t) \neq 0\}$.

Given ranked alphabets Σ and Δ, let $h'\colon \Sigma \to T_\Delta(X)$ be a mapping such that for all $k \in \mathbb{N}$ and $\sigma \in \Sigma_k$, we have $h'(\sigma) \in T_\Delta(X_k)$. We extend h' to a mapping $h\colon T_\Sigma \to T_\Delta$ by $h(\alpha) = h'(\alpha) \in T_\Delta(X_0) = T_\Delta$ for all $\alpha \in \Sigma_0$, and by $h(\sigma(s_1, \ldots, s_k)) = h'(\sigma)[x_1 \leftarrow h(s_1), \ldots, x_k \leftarrow h(s_k)]$ for all $k \in \mathbb{N}$, $\sigma \in \Sigma_k$, and $s_1, \ldots, s_k \in T_\Sigma$. 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 $k \in \mathbb{N}$,
- *input-finitary* if the preimage $h^{-1}(t)$ is finite for every $t \in T_\Delta$, and
- *tetris-free* if it is nondeleting, nonerasing and for $s, s' \in T_\Sigma$, $h(s) = h(s')$ implies (i) $\mathrm{pos}(s) = \mathrm{pos}(s')$ and (ii) $h\big(s(p)\big) = h\big(s'(p)\big)$ for all $p \in \mathrm{pos}(s)$.

In other words, a nondeleting and nonerasing $h\colon T_\Sigma \to T_\Delta$ is tetris-free if we cannot combine the building blocks $h(\sigma)$, $\sigma \in \Sigma$ in different ways to build the same tree. Thus if we list all possible trees that can be generated from these building blocks, no tree will occur twice. This condition was introduced in [23] and generalizes injectivity: Intuitively, if a tree homomorphism h is tetris-free, then any non-injective behaviour of h is located entirely at the symbol level.

Example 1. Let $\Sigma = \{\alpha^{(0)}, \beta^{(0)}, \psi^{(2)}\}$ and $\Delta = \{a^{(0)}, f^{(3)}\}$. Consider the tree homomorphism $h\colon T_\Sigma \to T_\Delta$ that is induced by the mapping $h(\alpha) = h(\beta) = a$ and $h(\psi) = f(x_2, x_1, x_1)$. While h is not injective, it is tetris-free. However, the tree homomorphism $h'\colon T_\Sigma \to T_\Delta$ induced by $h'(\alpha) = a$, $h'(\beta) = f(a, a, a)$ and $h'(\psi) = f(x_2, x_1, a)$ is not: $\psi(\alpha, \alpha)$ and β violate the tetris-free condition.

If $h\colon T_\Sigma \to T_\Delta$ is nonerasing and nondeleting, then for every $s \in h^{-1}(t)$, we have $|s| \leq |t|$. In particular, h is then input-finitary. Let $A\colon T_\Sigma \to \mathbb{S}$ be a tree series. 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 relies on h to be input-finitary, otherwise the defining sum is not finite, so $h(A)(t)$ might not be well-defined. For this reason, we only consider nondeleting and nonerasing tree homomorphisms.

Recently it was shown [20,21] that such homomorphic images of regular tree languages can be represented efficiently using *weighted tree automata with hom-constraints* (WTAh) which were defined in [20], and first introduced for the Boolean case in [14]. All following concepts are illustrated in Example 5 below.

Definition 2 (cf. [21, Definition 1]). *Let $\mathbb{S}$ be a commutative semiring. A* weighted tree automaton over $\mathbb{S}$ with hom-constraints (WTAh) *is a tuple of the form $\mathcal{A} = \big(Q, \Sigma, F, R, \mathrm{wt}\big)$ where 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 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 *constraints* and the *weight* $\mathrm{wt}(r)$. A constraint $(p, p') \in E$ is listed as "$p = p'$ ", and if p is different from p', then p and p' are called *constrained positions*. The equivalence class of p in E is denoted $[p]_{\equiv_E}$. We generally omit the trivial constraints $(p, p) \in E$.

The WTAh is a *weighted tree grammar* (WTG) if $E = \emptyset$ (strictly speaking, E is the identity relation) for every rule $\ell \xrightarrow{E} q$, and a WTA in the classical sense [5] if additionally $\mathrm{pos}_\Sigma(\ell) = \{\varepsilon\}$. WTG and WTA are equally expressive, as WTG can be translated straightforwardly into WTA using additional states.

We are particularly interested in a specific subclass of WTAh, namely the *eq-restricted* WTAh [21]. In such a device, there is a designated *sink-state* whose sole purpose is to neutrally process copies of identical subtrees. More precisely, whenever subtrees are mutually constrained, there is one leading copy among them that can be processed as usual with arbitrary states and weights, while every other copy is handled exclusively by the weight-neutral sink-state.

Definition 3. *A WTAh $\big(Q, \Sigma, F, R, \mathrm{wt}\big)$ is* eq-restricted *if it has a so-called* sink state $\bot \in Q \setminus F$ *such that (i) $\sigma(\bot, \ldots, \bot) \to_1 \bot$ belongs to R for all $\sigma \in \Sigma$, and no other rules target $\bot$, and (ii) for every rule $\ell \xrightarrow{E} q$ with $q \neq \bot$, if $\mathrm{pos}_Q(\ell) = \{p_1, \ldots, p_n\}$ and $q_i = \ell(p_i)$ for $i \in [n]$, the following conditions hold:*

1. *For each $i \in [n]$, the set $\{q_j \mid p_j \in [p_i]_{\equiv_E}\} \setminus \{\bot\}$ is a singleton.*
2. *There exists exactly one $p_j \in [p_i]_{\equiv_E}$ such that $q_j \neq \bot$.*

In other words, among each E-equivalence class there is only one occurrence of a state different from $\bot$, and every other E-related position is labelled by $\bot$. Moreover, $\bot$ processes every possible tree with weight 1. We denote the state sets of WTAh by $Q \dot\cup \{\bot\}$ instead of $Q \ni \bot$ to point out the sink-state.

Next, let us recall the semantics of WTAh from [21, Definitions 2 and 3].

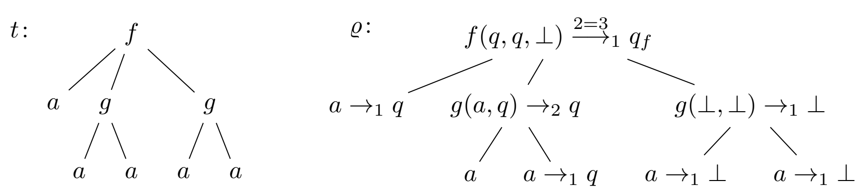
Definition 4. *Let $\mathcal{A} = \big(Q, \Sigma, F, R, \mathrm{wt}\big)$ be a WTAh. A* run of $\mathcal{A}$ *is a tree over the ranked alphabet $\Sigma \cup R$ where the rank of a rule is $\mathrm{rk}(\ell \xrightarrow{E} q) = \mathrm{rk}\,\big(\ell(\varepsilon)\big)$, and it is defined inductively. Consider $t_1, \ldots, t_n \in T_\Sigma$, $q_1, \ldots, q_n \in Q$ and suppose that ϱ_i is a run of $\mathcal{A}$ for t_i to q_i with weight $\mathrm{wt}(\varrho_i) = a_i$ for each $i \in [n]$. Assume there exists $\ell \xrightarrow{E}_a q$ in R such that $\ell = \sigma(\ell_1, \ldots, \ell_m)$, $\mathrm{pos}_Q(\ell) = \{p_1, \ldots, p_n\}$ with $\ell(p_i) = q_i$, and that $t_i = t_j$ for all $(p_i, p_j) \in E$. Let $t = \ell[t_1]_{p_1} \cdots [t_n]_{p_n}$, then $\varrho = \big(\ell \xrightarrow{E}_a q\big)(\ell_1, \ldots, \ell_m)[\varrho_1]_{p_1} \cdots [\varrho_n]_{p_n}$ is a run of $\mathcal{A}$ for t to q. Its* weight $\mathrm{wt}(\varrho)$ *is computed as $a \cdot \prod_{i \in [n]} a_i$. If $\mathrm{wt}(\varrho) \neq 0$, then ϱ is* valid*, and if in addition, $q \in F$ for its* target state *q, then ϱ is* accepting*. The value $\mathrm{wt}^q(t)$ is the sum of all weights $\mathrm{wt}(\varrho)$ of runs of $\mathcal{A}$ for t to q. Finally, the tree series $\|\mathcal{A}\| \colon T_\Sigma \to \mathbb{S}$ recognized by $\mathcal{A}$ is defined simply by $\|\mathcal{A}\| \colon t \mapsto \sum_{q \in F} \mathrm{wt}^q(t)$.*

Since the weights of rules are multiplied, we assume wlog $\mathrm{wt}(r) \neq 0$ for all $r \in R$.

Example 5. Let $\Delta = \{a^{(0)}, g^{(2)}, f^{(3)}\}$ and $\mathcal{A}' = \big(Q \dot\cup \{\bot\}, \Delta, F', R', \mathrm{wt}\big)$ be the WTGh over $\mathbb{Z}$ with $Q = \{q, q_f\}$, $F' = \{q_f\}$ and the set of rules and weights

$$\begin{aligned} R' \;=\; \big\{ \; & a \to_1 q\,, \quad g(a, q) \to_2 q\,, \quad\; f(q, q, \bot) \xrightarrow{2=3}_1 q_f\,, \\ & a \to_1 \bot\,, \quad g(\bot, \bot) \to_1 \bot\,, \;\; f(\bot, \bot, \bot) \to_1 \bot \; \big\}\,. \end{aligned}$$

The constrained positions $2, 3$ in the third rule satisfy (ii) from Definition 3, and the $\bot$-rules are as required in (i), so $\mathcal{A}'$ is eq-restricted. If we replace the third rule with $f(q, q, q) \xrightarrow{2=3}_1 q_f$, then the resulting WTAh is not eq-restricted any more. Let $t = f\big(a,\, g(a, a),\, g(a, a)\big) \in T_\Delta$. $\mathcal{A}'$ has a unique accepting run ϱ for t:



We have $\mathrm{wt}(\varrho) = 2$ despite $|\,\mathrm{pos}_g(t)| = 2$ because due to the eq-restriction, the duplicated subtree $t|_3$ is processed exclusively in the state $\bot$ with weight 1.

If a tree series is recognized by a WTA, it is called *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. The following example demonstrates this property.

Example 6. Consider $\Sigma = \{\alpha^{(0)}, \gamma^{(1)}, \psi^{(2)}\}$ and let $\mathcal{A} = \big(\{q, q_f\}, \Sigma, \{q_f\}, R, \mathrm{wt}\big)$ be the WTA over $\mathbb{Z}$ with the following set of rules:

$$R = \big\{\alpha \to_1 q,\; \gamma(q) \to_2 q,\; \psi(q, q) \to_1 q_f\big\}\,.$$

It is $\mathrm{supp}(\|\mathcal{A}\|) = \big\{\psi\big(\gamma^n(\alpha), \gamma^m(\alpha)\big) \mid n, m \in \mathbb{N}\big\} = \big\{s \in T_\Sigma \mid \mathrm{pos}_\psi(s) = \{\varepsilon\}\big\}$ and $\|\mathcal{A}\| \colon \psi\big(\gamma^n(\alpha), \gamma^m(\alpha)\big) \mapsto 2^{n+m} = 2^{|\mathrm{pos}_\gamma(s)|}$. Let $\Delta = \{a^{(0)}, g^{(2)}, f^{(3)}\}$ and let $h\colon T_\Sigma \to T_\Delta$ be the tetris-free tree homomorphism induced by

$$h(\alpha) = a\,, \quad h(\gamma) = g(a, x_1) \;\text{ and }\; h(\psi) = f\big(x_2, x_1, x_1\big)\;.$$

Then the eq-restricted WTAh $\mathcal{A}'$ from Example 5 recognizes $h(\|\mathcal{A}\|)$ defined by $\operatorname{supp}\big(h(\|\mathcal{A}\|)\big) = \big\{t \in T_\Delta \mid \operatorname{pos}_f(t) = \{\varepsilon\}\big\}$ and $h(\|\mathcal{A}\|)\colon t \mapsto 2^{|\operatorname{pos}_g(t) \setminus \operatorname{pos}_g(t|_3)|}$. The rules in R' are obtained from the rules in R by applying h to their left-hand sides, and the duplicated subtree at position 3 below f targets $\bot$ instead of q to avoid distorting the weight with an additional factor 2^n.

Formally, the following statement was shown in [20]. We have included the proof for better readability.

Lemma 7. (see [20, Theorem 19]) *Let $\mathcal{A} = \big(Q, \Sigma, F, R, \mathrm{wt}\big)$ be a WTA over a commutative semiring $\mathbb{S}$ and $h\colon T_\Sigma \to T_\Delta$ a nondeleting and nonerasing tree homomorphism. There is an eq-restricted WTAh $\mathcal{A}'$ that recognizes $h(\|\mathcal{A}\|)$.*

Proof. An eq-restricted WTAh $\mathcal{A}'$ for $h(\|\mathcal{A}\|)$ is constructed in two stages. First, we define $\mathcal{A}'' = \big(Q\dot\cup\{\bot\}, \Delta \cup \Delta \times R, F'', R'', \mathrm{wt}''\big)$ such that for every $r = \sigma(q_1, \ldots, q_k) \to_{\mathrm{wt}(r)} q$ in R and $h(\sigma) = u = \delta(u_1, \ldots, u_n)$, we include

$$r'' \;=\; \Big(\langle\delta, r\rangle(u_1, \ldots, u_n)[\![q_1, \ldots, q_k]\!] \xrightarrow{E}_{\mathrm{wt}''(r'')} q\Big) \;\in\; R''$$

with

$$E \;=\; \bigcup_{i \in [k]} \operatorname{pos}_{x_i}(u)^2$$

where the substitution $\langle\delta, r\rangle(u_1, \ldots, u_n)[\![q_1, \ldots, q_k]\!]$ replaces for every $i \in [k]$ only the $\leq_{\mathrm{lex}}$-minimal occurrence of x_i in $\langle\delta, r\rangle(u_1, \ldots, u_n)$ by q_i, and every other occurrence by $\bot$. For this rule we set $\mathrm{wt}''(r'') = \mathrm{wt}(r)$. Additionally, we let $r''_\delta = \delta(\bot, \ldots, \bot) \to \bot \in R''$ 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 [21, Theorem 4], we obtain an eq-restricted WTAh $\mathcal{A}' = \big(Q\dot\cup\{\bot\}, \Delta, F', R', \mathrm{wt}'\big)$ recognizing $h(\|\mathcal{A}\|)$. ☐

As illustrated in Examples 5 and 6, the WTAh $\mathcal{A}'$ for the homomorphic image of a WTA $\mathcal{A}$ replaces each symbol σ in a rule of $\mathcal{A}$ by $h(\sigma)$, and preserves the original state behaviour, only adding $\bot$ along the duplicated subtrees. Thus, we can define a mapping that traces the runs of $\mathcal{A}$ to the runs of $\mathcal{A}'$.

Definition 8. **[23, Definition 9]** *Let $\mathcal{A} = \big(Q, \Sigma, F, R, \mathrm{wt}\big)$ be a WTA over a commutative semiring $\mathbb{S}$ and $h\colon T_\Sigma \to T_\Delta$ a nondeleting and nonerasing tree homomorphism. Let $\mathcal{A}'$ be the WTAh for $h(\|\mathcal{A}\|)$ provided by Lemma 7. Consider a rule $r = \sigma(q_1, \ldots, q_k) \to q$ of $\mathcal{A}$ and let $h(\sigma) = \delta(u_1, \ldots, u_n)$, then we set*

$$h^R(r) = \delta(u_1, \ldots, u_n)[\![q_1, \ldots, q_k]\!] \xrightarrow{E} q,$$

where the substitution $[\![q_1,\dots,q_k]\!]$ replaces for every $i \in [k]$ only the $\leq_{lex}$-minimal occurrence of x_i in $\delta(u_1,\dots,u_n)$ by q_i, and every other occurrence by $\perp$. The constraint set is defined as $E = \bigcup_{i\in[k]} \big[\operatorname{pos}_{x_i}\big(\delta(u_1,\dots,u_n)\big)\big]^2$.

The assignment h^R extends naturally to the runs of $\mathcal{A}$: For a run of the form $\vartheta = r = (\alpha \to q)$ with $\alpha \in \Sigma^0$, we set $h^R(\vartheta) = h^R(r)$. For $\vartheta = r(\vartheta_1,\dots,\vartheta_k)$ with $r = \sigma(q_1,\dots,q_k) \to q$ and $h(\sigma) = \delta(u_1,\dots,u_n)$ we set

$$h^R(\vartheta) = \big(h^R(r)\big)(u_1,\dots,u_n)[\![h^R(\vartheta_1),\dots,h^R(\vartheta_k)]\!] \ ;$$

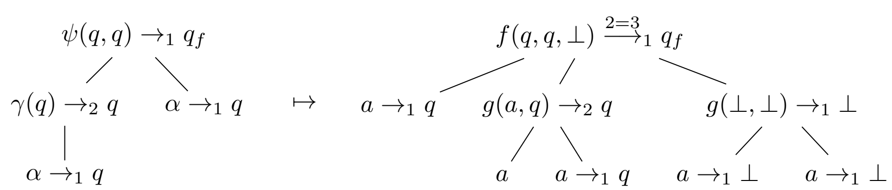
here, the substitution $[\![h^R(\vartheta_1),\dots,h^R(\vartheta_k)]\!]$ replaces for every $i \in [k]$ only the $\leq_{lex}$-minimal occurrence of x_i in $\big(h^R(r)\big)(u_1,\dots,u_n)$ by $h^R(\vartheta_i)$, and all other occurrences by the respective unique run to $\perp$ for the tree processed by ϑ_i.

Let us see how h^R acts on our example from above.

Example 9. Recall the WTA $\mathcal{A}$ and WTAh $\mathcal{A}'$ from Examples 5 and 6. We have

$$h^R : \quad \psi(q,q) \to q_f \quad \mapsto \quad f\big(q,q,\perp\big) \xrightarrow{2=3} q_f \ ,$$

and for the unique run of $\mathcal{A}$ for the tree $\psi\big(\gamma(\alpha),\alpha\big)$, the image under h^R is



The following statement is a direct consequence of the proof of Lemma 7.

Lemma 10. *The mapping h^R from Definition 8 is well-defined on R, although not necessarily injective. Its image is $h^R(R) = \{r' \in R' \mid r' \text{ targets some } q \neq \perp\}$. If ϑ is a run of $\mathcal{A}$ for $s \in T_\Sigma$, then $h^R(\vartheta)$ is a run of $\mathcal{A}'$ for $h(s)$; conversely, for every run ϱ of $\mathcal{A}'$ for some $t \in T_\Delta$ to some $q \neq \perp$, there exists $s \in h^{-1}(t)$ and a run ϑ of $\mathcal{A}$ for s to q such that $h^R(\vartheta) = \varrho$, but $\operatorname{wt}(\vartheta)$ and $\operatorname{wt}'(\varrho)$ may differ.*

3 A Pumping Lemma over Fields

The weighted HOM-problem takes a WTA $\mathcal{A}$ and a nondeleting, nonerasing tree homomorphism h as input, and asks whether $h(\|\mathcal{A}\|)$ is again regular. As mentioned earlier, the $\mathbb{N}$-variant of this problem was shown to be decidable in [22]. The proof presented there makes two assumptions on the semiring used for the weight calculations: First, it must be a subsemiring of a field, and second, it must be zero-sum free; the only common semiring that satisfies both conditions is $\mathbb{N}$. Remarkably, the strong condition of zero-sum freeness is only used to prove a pumping lemma for $h(\|\mathcal{A}\|)$. In this section, we derive an alternative pumping lemma over fields, provided that h is tetris-free. This way, we bypass the zero-sum freeness assumption, which allows us to lift the proof of [22] to the HOM-problem over fields, for tetris-free tree homomorphisms.

We begin by establishing a notation for the tree fragment read by a rule.

Definition 11. *Let $\mathcal{A}' = \big(Q \dot\cup \{\bot\}, \Delta, F, R, \mathrm{wt}\big)$ be an eq-restricted WTAh and let $r = \ell \xrightarrow{E} q$ be a rule of $\mathcal{A}'$ with some $q \neq \bot$. Let $\mathrm{pos}_{Q\setminus\{\bot\}}(\ell) = \{p_1, \ldots, p_k\}$. The* Δ-part of r *is the tree $\widehat{\ell} = \ell[\bot]_{p_1} \cdots [\bot]_{p_k} \in T_\Delta(\{\bot\})$.*

The Δ-part of a rule extracts the tree fragment from its left-hand side and overwrites every state label (for convenience simply with $\bot$). Note that ℓ can be easily recovered from $\widehat{\ell}$, E and the states $\ell(p_1), \ldots, \ell(p_k)$ in the correct order.

To prove the desired pumping lemma for our WTAh, we reduce it to the pumping lemma for WTA over fields proved by Berstel and Reutenauer in [1]. For this, we must construct a WTA related to the WTAh $\mathcal{A}'$ for $h(\|\mathcal{A}\|)$. The naive idea to simply use the input WTA $\mathcal{A}$ falls short: If h is not injective, there may be $s, s' \in \mathrm{supp}(\|\mathcal{A}\|)$ with $h(s) = h(s') \notin \mathrm{supp}(\|\mathcal{A}'\|)$ since in fields, different runs for $h(s)$ might cancel each other out, so we cannot lift the pumping lemma from $\mathcal{A}$ to $\mathcal{A}'$. Instead, we fabricate a new WTA that traces the behaviour of $\mathcal{A}'$ but ignores duplicated subtrees in order to remain regular. We will argue the well-definedness of this construction using some technical lemmas below.

Definition 12. *Let $\mathcal{A}' = \big(Q \dot\cup \{\bot\}, \Delta, F, R, \mathrm{wt}\big)$ be the eq-restricted WTAh from Lemma 7 for a WTA and a tetris-free tree homomorphism. Consider the ranked alphabet $\widehat{\Delta} = \{\widehat{\ell} \mid \ell$ is the left-hand side of some $r \in R\}$ with the rank function $\widehat{\mathrm{rk}}(\widehat{\ell}) = |\mathrm{pos}_{Q\setminus\{\bot\}}(\ell)|$. We define the WTA $\widehat{\mathcal{A}} = \big(Q \setminus \{\bot\}, \widehat{\Delta}, F, \widehat{R}, \widehat{\mathrm{wt}}\big)$ such that if $r = \ell \xrightarrow{E} q \in R$ with $q \neq \bot$ and $\mathrm{pos}_{Q\setminus\{\bot\}}(\ell) = \{p_1, \ldots, p_k\}$ ordered lexicographically with $\ell(p_i) = q_i$ for all $i \in [k]$, then $\widehat{\ell}(q_1, \ldots, q_k) \to q \in \widehat{R}$ with weight $\mathrm{wt}(r)$. No other rules are in $\widehat{R}$.*

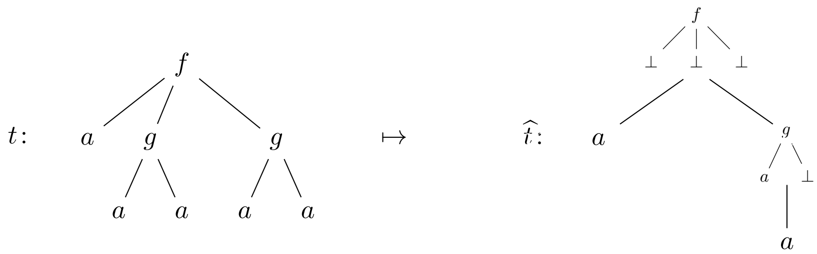
The translation $\mathcal{A}' \mapsto \widehat{\mathcal{A}}$ induces a mapping $t \mapsto \widehat{t}$ defined inductively as follows: Consider $t \in T_\Delta$, a run ϱ of $\mathcal{A}'$ for t with $\varrho(\varepsilon) = \ell \xrightarrow{E} q$ and let $\mathrm{pos}_{Q\setminus\{\bot\}}(\ell)$ be the set $\{p_1, \ldots, p_k\}$ in lexicographic order. Then $\widehat{t} = \widehat{\ell}\big(\widehat{t|_{p_1}}, \ldots, \widehat{t|_{p_2}}\big) \in T_{\widehat{\Delta}}$.

The WTA $\widehat{\mathcal{A}}$ reinterprets the trees $t \in T_\Delta$ as trees $\widehat{t} \in T_{\widehat{\Delta}}$ which, instead of symbols $\delta \in \Delta$, are now composed of the Δ-parts of the rules of $\mathcal{A}'$. As the WTA $\widehat{\mathcal{A}}$, without the instrument of constraints at hand, cannot ensure equality of subtrees, all $\bot$-processed copies are discarded, and $\bot$ is not a state anymore.

Example 13. Recall the WTAh $\mathcal{A}'$ from Example 5. The ranked alphabet $\widehat{\Delta}$ is the set $\widehat{\Delta} = \{a^{(0)}, [g(a,\bot)]^{(1)}, [f(\bot,\bot,\bot)]^{(2)}\}$, and the WTA $\widehat{\mathcal{A}}$ is defined by $\widehat{\mathcal{A}} = \big(Q, \widehat{\Delta}, F', \widehat{R}, \widehat{\mathrm{wt}}\big)$ with the following set of rules and weights:

$$\widehat{R} \;=\; \big\{\; a \to_1 q\,, \quad [g(a,\bot)](q) \to_2 q\,, \quad [f(\bot,\bot,\bot)](q,q) \to_1 q_f \;\big\}\,.$$

For $t = f\big(a,\, g(a,a),\, g(a,a)\big) \in T_\Delta$ it is $\widehat{t} = [f(\bot,\bot,\bot)]\big(\,a,\; [g(a,\bot)](a)\,\big) \in T_{\widehat{\Delta}}$:



The following two lemmas are the basis for the correctness of our translation above. Unlike in a WTA where trees are read symbol-by-symbol, a rule of a WTAh processes an entire tree fragment; in general, there may be different ways to assemble a certain tree from these Δ-parts of the rules of the WTAh, but by definition, tetris-free tree homomorphisms exclude this ambiguity.

Lemma 14. *Let* $\mathcal{A}' = \big(Q\dot\cup\{\bot\}, \Delta, F, R, \mathrm{wt}\big)$ *be the eq-restricted WTAh from Lemma 7 for a WTA and a tetris-free tree homomorphism. For every* $t \in T_\Delta$*, the runs of* $\mathcal{A}'$ *for* t *differ only in the states they process, but neither in the* Δ*-part of the rules they use, nor in their constraints. In particular, the set of positions related to any* $p'' \in \mathrm{pos}(t)$ *by the constraints of the rules used in a run coincides for all runs of* $\mathcal{A}'$ *for* t*, i.e. it is uniquely determined by* t*.*

Proof. Let ϱ and ϱ' be runs of $\mathcal{A}'$ for some $t \in T_\Delta$. By Lemma 10, there are two runs ϑ and ϑ' of $\mathcal{A}$ for some s and s', respectively, such that $h(s) = h(s') = t$, and $h^R(\vartheta) = \varrho$ and $h^R(\vartheta') = \varrho'$. Since h is tetris-free, it is $\mathrm{pos}(s) = \mathrm{pos}(s')$ and $h\big(s(p)\big) = h\big(s'(p)\big)$ at every $p \in \mathrm{pos}(s)$. By the definition of h^R, these identical terms $h\big(s(p)\big)$ and $h\big(s'(p)\big)$ already determine the Δ-parts of the rules used by ϱ and ϱ'. Moreover, the constraint sets are implicit to these terms, therefore ϱ and ϱ' can only differ in the states they process. $\square$

The next lemma is again a consequence of the tetris-freeness.

Lemma 15. *Let* $\mathcal{A}'$ *be the eq-restricted WTAh from Lemma 7 for a WTA and a tetris-free tree homomorphism. If* $\mathcal{A}'$ *has two rules* r, r' *with the same* Δ*-parts, then their constraint sets coincide as well.*

Proof. We will infer the statement from a general property of tetris-free tree homomorphisms: Let $h\colon T_\Sigma \to T_\Delta$ be tetris-free, and let $\sigma, \tau \in \Sigma$ such that $h(\sigma)$ and $h(\tau)$ coincide on their Δ-positions – formally, $\mathrm{pos}_\Delta\big(h(\sigma)\big) = \mathrm{pos}_\Delta\big(h(\tau)\big)$, and $\big(h(\sigma)\big)(p) = \big(h(\tau)\big)(p)$ for all $p \in \mathrm{pos}_\Delta\big(h(\sigma)\big)$ – then a lready $h(\sigma) = h(\tau)$. To see this, note first that since $\mathrm{pos}_\Delta\big(h(\sigma)\big) = \mathrm{pos}_\Delta\big(h(\tau)\big)$ and the variables $x \in X$ are nullary symbols, it follows that $\mathrm{pos}\big(h(\sigma)\big) = \mathrm{pos}\big(h(\tau)\big)$. Next, let $\alpha \in \Sigma_0$ be a nullary symbol. By assumption, $h\big(\sigma(\alpha, \ldots, \alpha)\big) = h\big(\tau(\alpha, \ldots, \alpha)\big)$, since the same subtree $h(\alpha)$ is attached to every X-position of $h(\sigma)$ and $h(\tau)$, regardless of the particular variable. But since h was assumed to be tetris-free, we infer that (σ and τ have the same rank) $h(\sigma) = h(\tau)$. The claim of the lemma follows immediately by applying this property of h to Definition 8. $\square$

We are now ready prove that our translation $\mathcal{A}' \mapsto \widehat{\mathcal{A}}$ is correct:

Lemma 16. *The WTA $\widehat{\mathcal{A}}$ from Definition 12 is well-defined. The mapping $t \mapsto \widehat{t}$ induced by it is also well-defined and injective, and $\|\mathcal{A}'\|(t) = \|\widehat{\mathcal{A}}\|(\widehat{t})$.*

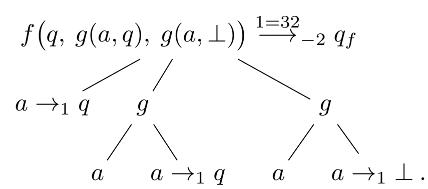
Proof. First, recall that ℓ can be recovered from $\widehat{\ell}$, E and the states $q_1, \dots, q_k$. While $\widehat{\ell}$ and $q_1, \dots, q_k$ are preserved in the rules of $\widehat{\mathcal{A}}$, E is uniquely determined by $\widehat{\ell}$ as stated in Lemma 15. Thus the weight function $\widehat{\mathrm{wt}}$ is well-defined.

Let $t \in T_\Delta$. By Lemma 14, all runs of $\mathcal{A}'$ for t have the same Δ-parts, and these are precisely the alphabet symbols for $\widehat{t} \in T_{\widehat{\Delta}}$. Thus, the mapping $t \mapsto \widehat{t}$ is well-defined. Since E (and thus the positioning of every direct subtree) is uniquely determined by $\widehat{\ell}$ via Lemma 15, the mapping is also injective. Finally, $\widehat{\mathcal{A}}$ preserves the state behaviour and weights, so every run of $\mathcal{A}'$ for t to some $q \neq \bot$ corresponds to a run of $\widehat{\mathcal{A}}$ for $\widehat{t}$ to q, and vice versa, which proves the claim.

For illustration purposes, consider cases where Lemmas 14 and 15 do not hold.

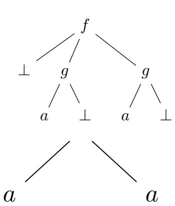
Example 17. Recall the WTAh $\mathcal{A}'$ recognizing $h(\|\mathcal{A}\|)$ from Examples 5 and 6. Since h is tetris-free, $\mathcal{A}'$ satisfies Lemmas 14 and 15. If we add $\varphi^{(2)}$ to the input alphabet Σ, extend $\mathcal{A}$ to, say, $\mathcal{B}$ by adding the rule $\varphi(q, q) \to_{-2} q_f$, and extend h to $h_\star$ via $h_\star(\varphi) = f\big(x_1,\ g(a, x_2),\ g(a, x_1)\big)$, then $h_\star$ is not tetris-free. The eq-restricted WTAh $\mathcal{B}'$ for $h_\star(\|\mathcal{B}\|)$ has the rule $f\big(q,\ g(a, q),\ g(a, \bot)\big) \xrightarrow{1=32}_{-2} q_f$, which allows an additional run $\varrho_\star$ for our tree $t = f\big(a,\ g(a, a),\ g(a, a)\big)$:

$\varrho_\star$:



It is $\mathrm{wt}(\varrho) + \mathrm{wt}(\varrho_\star) = 0$, hence $t \notin \mathrm{supp}(\|\mathcal{B}'\|)$. The rules at $\varrho(\varepsilon)$ and $\varrho_\star(\varepsilon)$ have different Δ-parts, so the statement in Lemma 14 does not hold. Indeed if we construct $\widehat{\mathcal{B}}$, we obtain the new symbol $\big[f\big(\bot,\ g(a, \bot),\ g(a, \bot)\big)\big]^{(2)} \in \widehat{\Delta}$ which provides a second tree $\widehat{t}_\star \in T_{\widehat{\Delta}}$ related to t:

$\widehat{t}_\star$:



So, while the translation $t \mapsto \widehat{t}$ is still injective, it is not well-defined anymore. Moreover, it is $\widehat{t}, \widehat{t}_\star \in \mathrm{supp}(\|\widehat{\mathcal{B}}\|)$, despite $t \notin \mathrm{supp}(\|\mathcal{B}'\|)$.

On the other hand, instead of $\varphi^{(2)}$ let us add $\beta^{(0)}$ and $\kappa^{(2)}$ to Σ, and b to Δ. We extend $\mathcal{A}$ to, say, $\mathcal{C}$ by adding the rules $\beta \to_1 q$ and $\kappa(q, q) \to_1 q_f$, and h

to $h^\star$ by setting $h^\star(\beta) = b$ and $h^\star(\kappa) = f(x_2,\, x_1,\, x_2)$. As before, $h^\star$ is not tetris-free. The WTAh $\mathcal{C}'$ has, compared to $\mathcal{A}'$, the additional rules $b \to_1 q,\ b \to_1 \perp$ and $f(q, q, \perp) \xrightarrow{1=3}_1 q_f$, so it does not satisfy Lemma 15. When constructing $\widehat{\mathcal{C}}$, we only add the symbol b to $\widehat{\Delta}$, but now there are two different rules whose Δ-part is $f(\perp, \perp, \perp)$. It is $h^\star\big(\kappa(\alpha, \beta)\big) = f(b, a, b) \neq f(b, a, a) = h^\star\big(\psi(\alpha, \beta)\big)$; however, we have $\widehat{f(b, a, b)} = \widehat{f(b, a, a)} = \big[f(\perp, \perp, \perp)\big](b, a)$. Not only is it unclear which weight the rule $\big[f(\perp, \perp, \perp)\big](q, q) \to q_f$ should have in $\widehat{\mathcal{C}}$, but because the translation $t \mapsto \widehat{t}$ is not injective, we cannot recover $\|\mathcal{C}'\|$ from $\|\widehat{\mathcal{C}}\|$ anymore.

Next, we want to derive a pumping lemma for our WTAh $\mathcal{A}'$, which will be the foundation for deciding the weighted HOM-problem over fields. To this end, we apply the well-known pumping lemma for WTA proved by Berstel and Reutenauer [1] to the WTA $\widehat{\mathcal{A}}$. We require one more definition: that of a context.

Definition 18. *Let Δ be a ranked alphabet and $\Box \notin \Delta$. Any $C \in T_\Delta(\{\Box\}) \setminus T_\Delta$ is called a* multi-context. *If $|\mathrm{pos}_\Box(C)| = 1$, then C is a* context. *For a multi-context C with $\mathrm{pos}_\Box(C) = \{p_1, \ldots, p_n\}$ and $t_1, \ldots, t_n \in T_\Sigma(\{\Box\})$, we abbreviate $C[t_1]_{p_1} \cdots [t_n]_{p_n}$ to $C[t_1, \ldots, t_n]$, and if $t_1 = \ldots = t_n = t$, we simply write $C[t]$.*

Let us now recall the pumping lemma for WTA over fields with a slight adjustment, namely that the pumping takes place below a certain position.

Theorem 19 (cf. [1, Theorem 9.2]). *Let $\mathbb{F}$ be a field, Σ a ranked alphabet and $\mathcal{B}$ a WTA over $\mathbb{F}$ and Σ. There exists $N \in \mathbb{N}$ s.t. for every context C and $t_0 \in T_\Sigma$ such that $C[t_0] \in \mathrm{supp}(\|\mathcal{B}\|)$ and $\mathrm{ht}(t_0) \geq N$, there exists a sequence of pairwise distinct trees $t_1, t_2, \ldots$ such that $C[t_i] \in \mathrm{supp}(\|\mathcal{B}\|)$ for all $i \in \mathbb{N}$.*

From this, we obtain the desired pumping lemma for the WTAh $\mathcal{A}'$.

Proposition 20 (Pumping Lemma). *Let $\mathbb{F}$ be a field and $\mathcal{A}'$ the eq-restricted WTAh from Lemma 7 for a WTA over $\mathbb{F}$ and Σ, and a tetris-free tree homomorphism $h\colon T_\Sigma \to T_\Delta$. There exists $N \in \mathbb{N}$ such that for every multi-context C and $t_0 \in T_\Delta$ such that $t := C[t_0] \in \mathrm{supp}(\|\mathcal{A}'\|)$, $\mathrm{ht}(t_0) \geq N$, and $\mathrm{pos}_\Box(C)$ is an equivalence class of mutually constrained positions in t, there exist infinitely many pairwise distinct trees $t_1, t_2, \ldots$ such that $C[t_i] \in \mathrm{supp}(\|\mathcal{A}'\|)$ for all $i \in \mathbb{N}$.*

Proof. Recall from Lemma 14 that the equivalence relation of positions that are mutually constrained by a run of t, is uniquely determined by t. Let $\widehat{\mathcal{A}}$ be the WTA for $\mathcal{A}'$ from Definition 12 and let $\widehat{N}$ be the pumping constant for $\widehat{\mathcal{A}}$ from Theorem 19. We set $N = \widehat{N} \cdot \max_{\sigma \in \Sigma} \mathrm{ht}\big(h(\sigma)\big)$. Let t be as in the statement, then $\widehat{t} \in \mathrm{supp}(\widehat{\mathcal{A}})$ is of the form $\widehat{t} = \widehat{C}[\widehat{t_0}]$ with a context $\widehat{C}$ and $\mathrm{ht}(\widehat{t_0}) \geq \widehat{N}$. Thus by Theorem 19, there is a sequence of trees $\widehat{t}_1, \widehat{t}_2, \ldots$ such that $\widehat{C}[\widehat{t_i}] \in \mathrm{supp}(\|\widehat{\mathcal{A}}\|)$ for all $i \in \mathbb{N}$. In turn, each $\widehat{t_i}$ has a unique preimage t_i under the mapping $t \mapsto \widehat{t}$, and $\widehat{C}$ translates uniquely back to C. Thus we obtain t_i, $i \in \mathbb{N}$ with $\|\mathcal{A}'\|\big(C[t_i]\big) = \|\widehat{\mathcal{A}}\|\big(\widehat{C[t_i]}\big) = \|\widehat{\mathcal{A}}\|\big(\widehat{C}[\widehat{t_i}]\big) \neq 0$ for all $i \in \mathbb{N}$. $\Box$

4 The Tetris-Free Weighted HOM-Problem

In this section, we prove that the weighted HOM-problem over fields, restricted to tetris-free homomorphisms, is decidable. Formally, we show the following result.

Theorem 21. *Let $\mathbb{F}$ be a field, $\mathcal{A}$ a WTA over $\mathbb{F}$ and Σ, and $h\colon T_\Sigma \to T_\Delta$ a tetris-free tree homomorphism. It is decidable whether $h(\|\mathcal{A}\|)$ is regular.*

The approach to prove this result is quite natural: Nonregularity of $h\big(\|\mathcal{A}\|\big)$ is reduced to the following decidable property of the WTAh $\mathcal{A}'$ for $h(\|\mathcal{A}\|)$.

Definition 22 (see [22, Definition 10]). *Let $\mathcal{A}' = \big(Q\dot\cup\{\bot\}, \Delta, F, R, \mathrm{wt}\big)$ be the eq-restricted WTAh from Lemma 7 for a WTA over a field and a tetris-free tree homomorphism, and let N be the pumping constant of $\mathcal{A}'$. We say that $\mathcal{A}'$ has the* large duplication property (LDP) *if there exists $t \in \mathrm{supp}(\|\mathcal{A}'\|)$ with an accepting run ϱ, a position $p \in \mathrm{pos}_R(\varrho)$ where $\varrho(p)$ has a nontrivial constraint set E, and a position p' that is constrained by E such that $\mathrm{ht}(t|_{pp'}) \geq N$.*

A constraint that only acts on finitely many trees is expendable, since we can process these particular trees manually using additional states. If, however, $\mathcal{A}'$ has the LDP, then by our pumping lemma we obtain infinitely many trees to which a nontrivial constraint E applies, so we cannot bypass E. Thus, the LDP indicates that the constraints are indeed indispensable for representing $\|\mathcal{A}'\|$, and in turn these constraints cause nonregularity, as stated in Proposition 25.

The decision procedure of [22] for input $\mathcal{A}$ and h as above is now as follows.

1. Construct an eq-restricted WTAh $\mathcal{A}'$ recognizing $h(\|\mathcal{A}\|)$ via Lemma 7.
2. If $\mathcal{A}'$ has the LDP, then $h\big(\|\mathcal{A}\|\big)$ is not regular.
3. If $\mathcal{A}'$ does not have the LDP, then $h\big(\|\mathcal{A}\|\big)$ is regular.

For this procedure to be correct, the LDP must be (i) decidable and (ii) equivalent to the nonregularity of $\|\mathcal{A}'\|$. While proving (ii) only requires technical adaptations compared to [22], (i) presents new challenges since the pumping lemma for fields is weaker. We prove (i) indirectly by examining the WTA $\widehat{\mathcal{A}}$.

Proposition 23. (cf. [22, Lemma 11]) *Given as input a WTA $\mathcal{A}$ over a field, and a tetris-free tree homomorphism h, it is decidable whether the eq-restricted WTAh $\mathcal{A}'$ for $h(\|\mathcal{A}\|)$ from Lemma 7 has the LDP.*

Proof. Adopting the notation from Definition 22, let $t_0 = t|_{pp'}$. We will not decide the existence of a tree $t = C[t_0]$ in $\mathrm{supp}(\|\mathcal{A}'\|)$ as in the LDP directly, but instead decide whether its counterpart $\widehat{C[t_0]}$ exists in $\mathrm{supp}(\|\widehat{\mathcal{A}}\|)$. Consider thus the WTA $\widehat{\mathcal{A}}$ for $\mathcal{A}'$ constructed in Definition 12. We modify $\widehat{\mathcal{A}}$ by implementing a counter into its state set, which ensures that only trees of height less than N are attached to positions that are constrained in $\mathcal{A}'$. Then we check if any trees have been lost from $\mathrm{supp}(\|\widehat{\mathcal{A}}\|)$ in the process. If so, then the counterparts of these lost trees in $\mathcal{A}'$ confirm the LDP, otherwise $\mathcal{A}'$ does not have the LDP.

Formally, let $q \neq \bot$ and $\ell \xrightarrow{E} q$ a rule of $\mathcal{A}'$ with $\mathrm{pos}_{Q\setminus\{\bot\}}(\ell) = \{p_1, \dots, p_k\}$ ordered lexicographically, and $\ell(p_i) = q_i$ for all $i \in [k]$. Suppose that $p_{i_1}, \dots, p_{i_j}$

are the positions constrained by E. Then $\mathcal{A}'$ has the rule $\widehat{\ell}(q_1, \dots, q_k) \to q$, and we replace it by the collection of all $\widehat{\ell}\big(\langle q_1, n_1\rangle, \dots, \langle q_k, n_k\rangle\big) \to \langle q, n\rangle$ such that $n_1, \dots, n_k, n \in [N]$, $n = \min\big\{ \max_{i\in[k]}(n_i + |p_i|),\ N\big\}$ and $n_{i_1}, \dots, n_{i_j} < N$. All these new rules have the same weight as $\widehat{\ell}(q_1, \dots, q_k) \to q$. This operation is well-defined since by Lemma 15, the constraint E is uniquely determined by $\widehat{\ell}$. We proceed this way for every rule of $\widehat{\mathcal{A}}$ and denote the resulting WTA by $\widehat{\mathcal{B}}$.

Consider now the WTA recognizing $\|\widehat{\mathcal{A}}\| - \|\widehat{\mathcal{B}}\|$ defined by a disjoint union. Subtracting $\|\widehat{\mathcal{B}}\|$ removes all $\widehat{t}$ from $\mathrm{supp}(\|\widehat{\mathcal{A}}\|)$ s.t. all subtrees of $t \in \mathrm{supp}(\|\mathcal{A}'\|)$ pending from constrained positions are of height less than N; thus, the WTA for $\|\widehat{\mathcal{A}}\| - \|\widehat{\mathcal{B}}\|$ only accepts trees whose counterparts in $\mathcal{A}'$ satisfy the LDP. It remains to decide whether $\|\widehat{\mathcal{A}}\| - \|\widehat{\mathcal{B}}\|$ is the zero function by minimizing the WTA for it [2,3] and checking whether it has zero states. If indeed $\|\widehat{\mathcal{A}}\| = \|\widehat{\mathcal{B}}\|$, then $\mathcal{A}'$ has no tree that satisfies the condition of the LDP. If, however, there exists $\widehat{t} \in \mathrm{supp}(\|\widehat{\mathcal{A}}\|) \setminus \mathrm{supp}(\|\widehat{\mathcal{B}}\|)$, then its counterpart t satisfies the LDP. $\square$

Finally, we can complete the proof of Theorem 21. For proving the final proposition, we apply the following version of Ramsey's theorem [25]. For a set X, we denote by $\binom{X}{2}$ the set of all subsets of X of size 2.

Theorem 24. *Let $k \geq 1$ be an integer and $f \colon \binom{\mathbb{N}}{2} \to [k]$ a mapping. There exists an infinite subset $E \subseteq \mathbb{N}$ such that $f|_{\binom{E}{2}} \equiv i$ for some $i \in [k]$.*

Proposition 25. (cf. [22, Prop. 13 and Thm. 17]) *Let $\mathcal{A}$ be a WTA over a field $\mathbb{F}$, h a tetris-free tree homomorphism, and $\mathcal{A}'$ the WTAh for $h(\|\mathcal{A}\|)$ constructed in Lemma 7. Then $h(\|\mathcal{A}\|)$ is regular iff $\mathcal{A}'$ does not have the LDP.*

Proof. Let $\mathcal{A}' = \{Q \dot\cup \{\bot\}, \Delta, F, R, \mathrm{wt}\}$ and let N be its pumping constant.
Necessity. We begin with the easier direction of this reduction: Suppose first that $\mathcal{A}'$ does not have the LDP. Therefore, every constraint used in a run of some $t \in \mathrm{supp}(\|\mathcal{A}'\|)$ only applies to subtrees of height less than N. We will construct a WTG (without constraints), called the *linearization of* $\mathcal{A}'$, that is equivalent to $\mathcal{A}'$. It was first defined for the unweighted case in [14] and adapted to the weighted setting in [22]. Formally, the *linearization of* $\mathcal{A}'$ is the WTG $\mathit{lin}(\mathcal{A}') = \big(Q, \Delta, F, R_{lin}, \mathrm{wt}_{lin}\big)$, where R_{lin} and wt_{lin} are defined as follows.

For $\ell' \in T_\Delta(Q)$ and $q \in Q$, we include the rule $(\ell' \to q)$ in R_{lin} 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 $t_1, \dots, t_k \in T_\Delta$ such that

- $\{p_1, \dots, p_k\} = \bigcup_{p \in \mathrm{pos}_\bot(\ell)} [p]_{\equiv E}$, that is, $p_1, \dots, p_k$ are exactly the positions constrained by E,
- $(p_i, p_j) \in E$ 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 $\mathrm{ht}(t_i) < N$ for all $i \in [k]$.

For every such production $\ell' \to q$ we set $\mathrm{wt}_{lin}(\ell' \to q)$ 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 $(\ell \xrightarrow{E} q) \in R$, $p_1, \ldots, p_k \in \mathrm{pos}_{Q\dot\cup\{\bot\}}(\ell)$ and $t_1, \ldots, t_k \in T_\Delta$ as above. In other words, $lin(\mathcal{A}')$ simulates all runs of $\mathcal{A}'$ which only enforce the equality of subtrees of height less than N. This is achieved by instantiating the constrained Q-positions of every rule $\ell \xrightarrow{E} q$ in $\mathcal{A}'$ with compatible trees of height less than N, while the Q-positions of ℓ that are unconstrained by E remain unchanged. Since $lin(\mathcal{A}')$ has no constraints and is equivalent to $\mathcal{A}'$, we have found a regular representation of the tree series $h(\|\mathcal{A}\|)$.

Sufficiency. The other direction of the proof is significantly more challenging. We divide the statement into three parts: Recall that from the LDP for some tree $t = C[t_0]$ together with the pumping lemma (Proposition 20) we obtain a sequence of pairwise distinct trees $t_0, t_1, t_2, \ldots$ such that $t[t_i]_{pp'} \in \mathrm{supp}(\|\mathcal{A}'\|)$ and the position p' is constrained in the rule at p (in every run of $\mathcal{A}'$ for t). First, we decompose $\mathcal{A}'$ as $\mathcal{A}_1 + \mathcal{A}_2$ such that $\mathcal{A}_1$ isolates accepting runs of $\mathcal{A}'$ for these trees, which will be the basis for the nonregularity of $\mathcal{A}'$. As a second step, we identify a subsequence where $\mathcal{A}_2$ behaves almost like $\mathcal{A}_1$. Finally, in the third part, we prove nonregularity of $\|\mathcal{A}'\|$ by contradiction using linear algebra computations similar to the initial sigma algebra semantics for WTA [11], which we perform on the subsequence identified in the second step.

Before we begin, we want to introduce an alternative notation for runs which will come in handy in the remainder of the proof. Consider $t \in \mathrm{supp}(\|\mathcal{A}'\|)$ and a run ϱ of $\mathcal{A}'$ for t. Let $\mathrm{pos}_R(\varrho) = \{p_1, \ldots, p_m\}$ ordered lexicographically, but such that prefixes are larger, i.e. $p_m = \varepsilon$. Other than thinking of ϱ as a tree in $T_{\Delta\cup R}$, we can list the rules it applies as $\big(\varrho(p_1), p_1\big)\big(\varrho(p_2), p_2\big) \cdots \big(\varrho(p_m), p_m\big)$, sometimes simply denoted $(r_1, p_1)(r_2, p_2) \cdots (r_m, p_m)$ with $r_1, \ldots, r_m \in R$. Recall that for all runs of a fixed tree t, the positions $p_1, \ldots, p_m$ are the same, as are the positions among them where the target state is $\bot$.

<u>Part 1.</u> Consider $t = C[t_0]$ and p, p' as in the LDP such that p is of minimal length among all choices (where p' is labeled by a non-sink state in the rule applied at p by any run for t) and the trees $t_0, t_1, t_2, \ldots$ such that $C[t_i] \in \mathrm{supp}(\|\mathcal{A}'\|)$ for all $i \in \mathbb{N}$. We want to construct two eq-restricted WTAh $\mathcal{A}_1, \mathcal{A}_2$ such that $\mathcal{A}' = \mathcal{A}_1 + \mathcal{A}_2$, and $\mathcal{A}_1$ simulates runs of $\mathcal{A}'$ for these trees that coincide above p. Formally, there exist two eq-restricted WTAh $\mathcal{A}_1 = \big(Q_1\dot\cup\{\bot\}, \Delta, F_1, R_1, \mathrm{wt}_1\big)$ and $\mathcal{A}_2 = \big(Q_2\dot\cup\{\bot\}, \Delta, F_2, R_2, \mathrm{wt}_2\big)$ s.t. $\mathcal{A}'(s) = \mathcal{A}_1(s) + \mathcal{A}_2(s)$ for all $s \in T_\Delta$ and $F_1 = \{q_f\}$ for some $q_f \in Q_1$, and there exists exactly one rule in R_1, say $\ell_f \xrightarrow{E_f} q_f$, whose target state is q_f, and for this rule there exists $(p'', p_\bot) \in E_f$ with $\ell_f(p'') \neq \bot = \ell_f(p_\bot)$ such that $p'' \in \mathrm{pos}_\Box(C)$ and $C[t_i] \in \mathrm{supp}(\|\mathcal{A}_1\|)$ for infinitely many $i \in \mathbb{N}$.

To identify $\mathcal{A}_1$, consider all runs of $\mathcal{A}'$ for all trees $C[t_i]$ as in the LDP, and sort them into groups by the rules applied on the prefixes $[\varepsilon, p] := \{p_{i_1}, \ldots, p_{i_k}\}$ of p in $\mathrm{pos}_R(\varrho)$ (in ascending order where prefixes are larger). By evaluating each group, we obtain an expression of the form $\mathrm{wt}\big(\varrho|_{[\varepsilon,p]}\big) \cdot \mathrm{wt}^{q'}(t_i) \cdot \pi_{rest}$, where $\varrho|_{[\varepsilon,p]}$ is the partial run $\big(\varrho(p_{i_1}), p_{i_1}\big)\big(\varrho(p_{i_2}), p_{i_2}\big) \cdots \big(\varrho(p_{i_k}), p_{i_k}\big)$ unique to the group, q' stands for the respective non-sink state at pp', and π_{rest} contains the weights of the subtrees of t unrelated to $t|_{pp'}$ that are attached to the rules $\varrho(p_{i_1}), \ldots, \varrho(p_{i_k})$ at positions parallel to pp'. All duplication copies of t_i are processed in $\bot$ with weight 1, so their weight can be neglected. Since the field is not zero-sum free, in some of the groups, the values $\mathrm{wt}^{q'}(t_i) \cdot \pi_{rest}$ might be zero, but since for every particular tree $C[t_i]$, the sum of weights over all groups is $\|\mathcal{A}'\|\big(C[t_i]\big) \neq 0$, and there are only finitely many groups, for one of the groups there must be infinitely many trees $t_{j_0}, t_{j_1}, t_{j_2} \ldots$ such that $\mathrm{wt}^{q'}(t_{j_i}) \cdot \pi_{rest} \neq 0$ for all $i \in \mathbb{N}$ and the respective state q' of that group. We pick this subsequence as our new sequence $t_0, t_1, t_2, \ldots$, and in the following, we will join the partial run $\varrho|_{[\varepsilon,p]}$ of this group into the rule $\ell_f \xrightarrow{E_f} q_f$ from the statement at the beginning of Part 1.

From here on, the proof of Part 1 works the same as in [22]. Let $\varrho(p_{i_j})$ be of the form $\ell_{i_j} \xrightarrow{E_{i_j}} q_{i_j}$ for all $j \in [k]$. For a position $\bar{p}$ and a constraint set $\bar{E}$ we define the set $\bar{p}\bar{E} = \{(\bar{p}p', \bar{p}p_\perp) \mid (p', p_\perp) \in \bar{E}\}$. We want to join the respective left-hand sides $\ell_{i_1}, \ldots, \ell_{i_k}$ of the rules applied by ϱ on the path from ε to p, to create a new rule $\ell_{i_k}[\ell_{i_{k-1}}]_{p_{i_{k-1}}} \cdots [\ell_{i_1}]_{p_{i_1}} \xrightarrow{E_f} q_f$ with $E_f = \bigcup_{j \in [k]} p_{i_j} E_{i_j}$. Note that by the minimality of p, none of the positions $p_{1_i}, \ldots, p_{i_k}$ can occur in E_f. We define $\mathcal{A}_1 = \big(Q \dot\cup \{\perp\}, \Delta, F_1, R_1, \mathrm{wt}_1\big)$ with $Q_1 = Q \dot\cup \{q_f\}$, $F_1 = \{q_f\}$ and $R_1 = R \cup \{r_f\}$ where r_f is the rule $\ell_{i_k}[\ell_{i_{k-1}}]_{p_{i_{k-1}}} \cdots [\ell_{i_1}]_{p_{i_1}} \xrightarrow{E_f} q_f$ with the constraint set $E_f = \bigcup_{j \in [k]} p_{i_j} E_{i_j}$, and the weight function wt_1 is defined by $\mathrm{wt}_1(r_f) = \prod_{j \in [k]} \mathrm{wt}\left(\varrho(p_{i_j})\right)$, and otherwise $\mathrm{wt}_1(r) = \mathrm{wt}(r)$ for all $r \in R$.

Finally, we construct $\mathcal{A}_2$ such that $\mathcal{A}' = \mathcal{A}_1 + \mathcal{A}_2$. For this, we must simulate all runs of $\mathcal{A}'$ except for those covered by $\mathcal{A}_1$. For a compact definition of $\mathcal{A}_2$, we use $\square$ to denote a tree of height 0, and a term $\square[\ell_{i_k}]_{p_{i_k}} \cdots [\ell_{i_{j+1}}]_{p_{i_{j+1}}}[\ell']_{p_{i_j}}$ for $j = k$ is to be read as $\square[\ell']_{p_{i_j}}$. We let $q_f \notin Q \cup \{\perp\}$ be a new state and define $\mathcal{A}_2 = (Q_2 \dot\cup \{\perp\}, \Delta, F_2, R_2, \mathrm{wt}_2)$ with $Q_2 = Q \cup \{q_f\}$, $F_2 = \{q_f\} \cup F \setminus \{q_{i_k}\}$ (where q_{i_k} is the target state of r_{i_k} at the root of ϱ), and the following rules:

$$R_2 = R \cup \bigcup_{j \in [k]} \Big\{ \square[\ell_{i_k}]_{p_{i_k}} \cdots [\ell_{i_{j+1}}]_{p_{i_{j+1}}}[\ell']_{p_{i_j}} \xrightarrow{E_f} q_f \;\Big|\; r' = (\ell' \xrightarrow{E'} q_{i_j}) \in R \setminus \{r_{i_j}\}, E_f = p_{i_j} E' \cup \bigcup_{j'=j+1}^{k} p_{i_{j'}} E_{i_{j'}} \Big\} .$$

For a rule $r_f = \square[\ell_{i_k}]_{p_{i_k}} \cdots [\ell_{i_{j+1}}]_{p_{i_{j+1}}}[\ell']_{p_{i_j}} \xrightarrow{E_f} f$ constructed with r' as above we let $\mathrm{wt}_2(r_f) = \mathrm{wt}(r') \cdot \prod_{j'=j+1}^{k} \mathrm{wt}(r_{i_{j'}})$, and for every $r' \in R$ we let $\mathrm{wt}_2(r') = \mathrm{wt}(r')$. This way, $\mathcal{A}_2$ reconstructs all runs of $\mathcal{A}'$ except for the ones that coincide with ϱ on the path from ε to p. An illustration of this construction was given in [22, Example 15].

Part 2. Next we identify a subsequence of $t_0, t_1, t_2, \ldots$ such that $C[t_i, t_j, \ldots, t_j]$ is not in $\mathrm{supp}(\|\mathcal{A}_2\|)$ if $i \neq j$. Recall that by the minimality assumption in Part 1, no prefix of pp' is a constrained position in t. Let $\{w_1, \ldots, w_r\}$ be the set of all positions equality constrained to $p'' = pp'$ by E_f in $\mathcal{A}'$, where $w_1 = p''$. i.e. $\{w_1, \ldots, w_r\} = \mathrm{pos}_\square(C)$. Since $\ell_f \xrightarrow{E_f} q_f$ is the only rule of $\mathcal{A}_1$ that targets the final state q_f, we have $\|\mathcal{A}_1\|\big(C[t_i, t_j, \ldots, t_j]\big) \neq 0$ iff $i = j$. Of course, $\mathcal{A}_2$ might have valid runs for $C[t_i]$, but also for $C[t_i, t_j, \ldots, t_j]$ with $i \neq j$. We will show that there is a subsequence of $t_0, t_1, t_2, \dots$ where also $\|\mathcal{A}_2\|\big(C[t_i, t_j, \dots, t_j]\big) = 0$ if $i \neq j$. Example 26 below shows an illustration of this selection.

For a run $\vartheta = \big(r_1, p_1\big) \cdots \big(r_m, p_m\big)$ of $\mathcal{A}_2$ and a set S, let $\{p_{i_1}, \dots, p_{i_n}\}$ be the set $\{p_1, \dots, p_m\} \cap S$, then $\vartheta|_S$ denotes the restricted run $\big(r_{i_1}, p_{i_1}\big) \cdots \big(r_{i_m}, p_{i_m}\big)$; its weight is $\mathrm{wt}_2(\vartheta|_S) = \prod_{j \in [n]} \mathrm{wt}_2(r_{i_j})$, and we define for all $k, h \in \mathbb{N}$:

$$\Theta_{kh} = \big\{\vartheta|_{\mathrm{pos}(C)} \mid \vartheta \text{ is accepting run of } \mathcal{A}_2 \text{ for } C[t_k, t_h, \dots, t_h]\big\} \ .$$

We now employ RAMSEY's theorem in the following way. For $k, h \in \mathbb{N}$ with $k < h$, we consider the mapping $\{k, h\} \mapsto \Theta_{kh}$. This mapping has a finite range as every Θ_{kh} is a set of finite words over the alphabet $R_2 \times \mathrm{pos}(C)$ of length at most $|\mathrm{pos}(C)|$. Thus, by RAMSEY's theorem, we obtain a subsequence $(t_{i_j})_{j \in \mathbb{N}}$ with $\Theta_{i_k i_h} = \Theta_<$ for all $k, h \in \mathbb{N}$ and some set $\Theta_<$. For simplicity, we assume that $\Theta_{kh} = \Theta_<$ for all $k, h \in \mathbb{N}$ with $k < h$. In the same fashion, we may select a further subsequence and assume that $\Theta_{kh} = \Theta_>$ for all $k, h \in \mathbb{N}$ with $k > h$. Finally, the mapping $k \mapsto \Theta_{kk}$ also has a finite range, so by the pigeonhole principle, we may select a further subsequence and assume that $\Theta_{kk} = \Theta_=$ for all $k \in \mathbb{N}$ and some set $\Theta_=$. In the following, we show that $\Theta_< = \Theta_> = \emptyset$. For this, we prove that $\Theta_< = \Theta_> \subseteq \Theta_=$; since $\mathcal{A}'$ satisfies Lemma 14 all runs of $\mathcal{A}'$ for $C[t_k]$ enforce equality for all positions in $\mathrm{pos}_\square(C)$. Moreover, the (absolute) positions constrained by runs of $\mathcal{A}_2$ are the same as in the corresponding runs of $\mathcal{A}'$. Therefore the set $\Theta_=$ is disjoint from both $\Theta_<$ and $\Theta_>$, so overall we have $\Theta_< = \Theta_> = \emptyset$.

Assume thus that $\Theta_< \neq \emptyset$. Let $(r_1, p_1) \cdots (r_m, p_m) \in \Theta_<$ with $r_i = \ell_i \xrightarrow{E_i} q_i$ for every $i \in [m]$. Moreover, we will abbreviate $C_{kh} = C[t_k, t_h, t_h, \dots, t_h]$, $C_{k\square} = C[t_k, \square, \square, \dots, \square]$, and $C_{\square h} = C[\square, t_h, t_h, \dots, t_h]$ for $k, h \in \mathbb{N}$. We show that every constraint from every E_i is satisfied on all C_{kh} with $k, h \geq 1$, not just for $k < h$. More precisely, let $i \in [m]$, $(u', v') \in E_i$, and $(u, v) = (p_i u', p_i v')$. We show $C_{kh}|_u = C_{kh}|_v$ for all $k, h \geq 1$. Note that by assumption, $C_{kh}|_u = C_{kh}|_v$ is true for all $k, h \in \mathbb{N}$ with $k < h$. We show our statement by a case distinction depending on the position of u and v in relation to $\{w_1, \dots, w_r\} = \mathrm{pos}_\square(C)$.

1. If both u and v are parallel to w_1, then $C_{ij}|_u$ and $C_{ij}|_v$ do not depend on i. Thus, $C_{0j}|_u = C_{0j}|_v$ for all $j \geq 1$ implies the statement.
2. If u is in prefix-relation with w_1 and v is parallel to w_1, then $C_{ij}|_v$ does not depend on i. If $u \leq w_1$, then by our assumption that $(t_i)_{i \in \mathbb{N}}$ are pairwise distinct, we obtain the contradiction $C_{02}|_v = C_{02}|_u \neq C_{12}|_u = C_{12}|_v$, where $C_{02}|_v = C_{12}|_v$ should hold. Thus, we have $w_1 \leq u$ and in particular, $C_{ij}|_u$ does not depend on j. Thus, for all $i, j \geq 1$ we obtain

$$C_{ij}|_u = C_{i,i+1}|_u = C_{i,i+1}|_v = C_{0,i+1}|_v = C_{0,i+1}|_u = C_{0j}|_u = C_{0j}|_v = C_{ij}|_v \ .$$

   If v is in prefix-relation with w_1 and u is parallel to w_1, then we come to the same conclusion by formally exchanging u and v in this argumentation.
3. If u and v are both in prefix-relation with w_1, then u and v being parallel to each other implies $w_1 \leq u$ and $w_1 \leq v$. In particular, both u and v are parallel

to all $w_2, \ldots, w_m$. Thus, we obtain, as in the first case, that $C_{ij}|_u$ and $C_{ij}|_v$ do not depend on j and the statement follows from $C_{i,i+1}|_u = C_{i,i+1}|_v$ for all $i \in \mathbb{N}$.

Let $k, h \geq 1$ and $\vartheta_C \in \Theta_<$. Moreover, let $q \in Q_2$ and let $\vartheta_{k,k+1}$ and $\vartheta_{h-1,h}$ be runs of $\mathcal{A}_2$ for $C_{k,k+1}$ and $C_{h-1,h}$, respectively, to q such that

$$\vartheta_C = \vartheta_{k,k+1}|_{\mathrm{pos}(C)} = \vartheta_{h-1,h}|_{\mathrm{pos}(C)} \ .$$

Let $\vartheta_k = \vartheta_{k,k+1}|_{\mathrm{pos}(C_{k,k+1})\setminus\mathrm{pos}(C_{\square,k+1})}$ and $\vartheta_h = \vartheta_{h-1,h}|_{\mathrm{pos}(C_{h-1,h})\setminus\mathrm{pos}(C_{h-1,\square})}$, then we can reorder $\vartheta = \vartheta_k\vartheta_h\vartheta_C$ to a run of $\mathcal{A}_2$ for C_{kh}, as all equality constraints from ϑ_k are satisfied by the assumption on $\vartheta_{k,k+1}$, all equality constraints from ϑ_h are satisfied by the assumption on $\vartheta_{h-1,h}$, and all equality constraints from ϑ_C are satisfied by our case distinction. Considering the special cases $k = 2$, $h = 1$, and $k = h = 1$, and the definitions of $\Theta_>$ and $\Theta_=$, we obtain $\vartheta_C \in \Theta_{21} = \Theta_>$ and $\vartheta_C \in \Theta_{11} = \Theta_=$, and hence, $\Theta_< \subseteq \Theta_>$ and $\Theta_< \subseteq \Theta_=$.

The converse inclusion $\Theta_> \subseteq \Theta_<$ follows with an analogous reasoning: Suppose again that $\Theta_> \neq \emptyset$ and consider as before some $(r_1, p_1) \cdots (r_m, p_m) \in \Theta_>$, a pair (u', v') constrained in r_i for some $i \in [m]$, and let $(u, v) = (p_i u', p_i v')$. By assumption, (u, v) is satisfied by C_{kh} for all $k > h$. Again, we distinguish three cases depending on the position of u and v compared to w_1. In the first and third case, we draw the conclusions from $C_{j+1,j}|_u = C_{j+1,j}|_v$ for all $j \geq 0$ and $C_{i0}|_u = C_{i0}|_v$ for all $i \geq 1$, respectively. In the second case, if u is in prefix-relation with w_1 and v is parallel to w_1, we first see that u is not a prefix of w_1. Otherwise we would again have a contradiction via $C_{20}|_v = C_{20}|_u \neq C_{10}|_u = C_{10}|_v$. Then we argue similarly that for all $i, j \geq 1$ we have

$$C_{ij}|_u = C_{i0}|_u = C_{i0}|_v = C_{j+1,0}|_v = C_{j+1,0}|_u = C_{j+1,j}|_u = C_{j+1,j}|_v = C_{ij}|_v \ .$$

In conclusion, we obtain $\Theta_< = \Theta_> \subseteq \Theta_=$. Since by Lemma 14, all partial runs in $\Theta_=$ enforce constraints on all $w_1, \ldots, w_r$, it is $\Theta_= \cap \Theta_< = \Theta_= \cap \Theta_> = \emptyset$, and thus we conclude $\Theta_< = \Theta_> = \emptyset$ as desired.

Part 3. Finally, we derive a representation of $\|\mathcal{A}'\|$ that allows us to prove its nonregularity. For $k \in \mathbb{N}$ let $\nu_k = \|\mathcal{A}_2\|(C_{kk})$, then it is $\|\mathcal{A}_2\|(C_{kh}) = \delta_{kh}\nu_k$, where δ_{kh} denotes the Kronecker delta. As mentioned above, we similarly have $\|\mathcal{A}_1\|(C_{kh}) \neq 0$ iff $k = h$, so we can overall write $\|\mathcal{A}'\|(C_{kh}) = \delta_{kh}\mu_k$ with $\mu_k \neq 0$ for all $k \in \mathbb{N}$. If $\|\mathcal{A}'\|$ is regular, then by the initial sigma algebra semantics [11] we can assume a representation $\|\mathcal{A}'\|(C_{kh}) = g(\kappa_k, \kappa_h, \ldots, \kappa_h)$ for all $k, h \in \mathbb{N}$, where κ_h is a finite vector of weights over $\mathbb{F}$ where each entry corresponds to the sum of all runs for t_h to a specific state of a WTA recognizing $\|\mathcal{A}'\|$ by the regularity assumption, and g is a multilinear map encoding the weights of the runs for C depending on the specific input states at the $\square$-positions and the target state at ε. Let dim be the number of entries in κ_h, then all κ_h, $h \in \mathbb{N}$ are elements of the finite-dimensional vector space $\mathbb{F}^{dim}$. We choose $K \in \mathbb{N}$ such that $\kappa_1, \ldots, \kappa_K$ are a generating set of the $\mathbb{F}$-vector space spanned by κ_i, $i \in \mathbb{N}$. Then there are coefficients $\alpha_1, \ldots, \alpha_K \in \mathbb{F}$ such that $\kappa_{K+1} = \sum_{i\in[K]} \alpha_i\kappa_i$. Thus we compute

$$0 \neq \mu_{K+1} = \|\mathcal{A}'\|(C_{K+1,K+1}) = g(\kappa_{K+1}, \kappa_{K+1}, \ldots, \kappa_{K+1})$$

$$= \sum_{i \in [K]} \alpha_i g(\kappa_i, \kappa_{K+1}, \ldots, \kappa_{K+1})$$
$$= \sum_{i \in [K]} \alpha_i \|\mathcal{A}'\|(C_{i,K+1}) = \sum_{i \in [K]} \alpha_i \delta_{i,K+1} \mu_i = 0,$$

so our assumption that $\|\mathcal{A}'\|$ is regular led to a contradiction. $\square$

To illustrate why it is necessary to identify such a subsequence in the proof of Proposition 25, consider the following simple example.

Example 26. Consider the WTAh $\mathcal{A}' = \big(\{q, q_f, \bot\}, \{a^{(0)}, g^{(1)}, f^{(2)}\}, \{q_f\}, R, \mathrm{wt}\big)$ over $\mathbb{Q}$ with the following rules:

$$R = \{\ a \to_1 q, \quad g(q) \to_2 q, \quad f(q, \bot) \xrightarrow{1=2}_1 q_f, \quad f\big(q, g(\bot)\big) \xrightarrow{1=21}_1 q_f\ \} \ \cup\ R_\bot$$

where $R_\bot = \{a \to_1 \bot, g(\bot) \to_1 \bot, f(\bot, \bot) \to_1 \bot\}$. The WTAh $\mathcal{A}'$ represents the image of a WTA under a suitable tetris-free tree homomorphism, and the context $C = f(\Box, \Box)$ and the sequence $t_i = g^i(a)$ satisfy the conditions of the LDP. For constructing $\mathcal{A}_1$ we can simply choose the third rule as it is. However, the corresponding WTAh $\mathcal{A}_2$ does not satisfy $\|\mathcal{A}_2\|\big(C[t_i, t_j]\big) = 0$ if $i = j$. Instead, for every i, it is $\|\mathcal{A}_2\|\big(C[t_i, t_{i+1}]\big) \neq 0$. However, if we choose the subsequence (t_{2i}), $i \in \mathbb{N}$, then indeed $\|\mathcal{A}_2\|\big(C[t_{2i}, t_{2j}]\big) = 0$ (in particular) for all $i \neq j$ as required for the computations in Part 3.

Restrictihg the HOM-problem over fields to tetris-free tree homomorphisms is of essence: On the one hand, we use this assumption to construct a well-defined WTA $\widehat{\mathcal{A}}$ when proving that the LDP is decidable in Proposition 23. On the other hand, the statement of Proposition 25, which reduces the weighted HOM-problem to the LDP, also does not hold if h is not tetris-free:

Example 27. Consider $\mathcal{A}' = \big(\{q, q_f, q'_f, \bot\}, \{a^{(0)}, g^{(1)}, f^{(2)}\}, \{q_f, q'_f\}, R, \mathrm{wt}\big)$ with

$$\begin{aligned} R = \{ \quad & a \to_1 q, \qquad g(q) \to_1 q, \qquad f(q, q) \to_3 q_f, \\ & f(q, \bot) \xrightarrow{1=2}_2 q_f, \qquad f(q, \bot) \xrightarrow{1=2}_{-2} q'_f \quad \} \ \cup\ R_\bot \end{aligned}$$

where $R_\bot = \{a \to_1 \bot, g(\bot) \to_1 \bot, f(\bot, \bot) \to_1 \bot\}$.

The WTAh $\mathcal{A}'$ represents the image of a WTA under a suitable tree homomorphism, but not under any tetris-free one since $\mathcal{A}'$ does not satisfy Lemma 14. It is easy to see that $\mathcal{A}'$ has the LDP, e.g. with $C = f(\Box, \Box)$ and the sequence $t_i = g^i(a)$. However, the accepting runs for $C[t_i]$ that use constraints cancel each other out. Despite $\mathcal{A}'$ having the LDP, $\|\mathcal{A}'\|$ is the regular tree series with $\mathrm{supp}(\|\mathcal{A}'\|) = \big\{t \mid \mathrm{pos}_f(t) = \{\varepsilon\}\big\}$ and $\|\mathcal{A}'\|\colon f\big(g^i(a), g(^j(a)\big) \mapsto 3$ for all $i, j \in \mathbb{N}$. Thus, without the tetris-free assumption, Proposition 25 does not hold.

5 Conclusion

In this paper, we have proved that the weighted HOM-problem over fields for tetris-free tree homomorphisms is decidable. Formally, for a WTA $\mathcal{A}$ over a field, and a tetris-free tree homomorphism h as input, it is decidable whether $h(\|\mathcal{A}\|)$ is again regular. A tree homomorphism is tetris-free if its non-injective behaviour is located only at the symbol level, thus this property generalizes injectivity.

Our proof strategy is similar to [22]: We have reduced the HOM-problem to a decidable property of (the WTAh that recognizes) $h(\|\mathcal{A}\|)$. The homomorphism h has the ability to duplicate subtrees of its input trees, and we have shown that $h(\|\mathcal{A}\|)$ is regular iff h duplicates only finitely many subtrees of trees accepted by $\mathcal{A}$. This limited duplication is in turn decidable, and proving its decidability is our main contribution. For this, we presented a pumping lemma for the WTAh recognizing $h(\|\mathcal{A}\|)$, by translating it into a WTA and applying the pumping lemma for WTA over fields proved in [1].

The analogous decision problem, without the tetris-free restriction, is also decidable for WTA over $\mathbb{N}$ [22]. However, since fields allow zero-sums, the proof strategy fails without the tetris-free restriction, as our last example illustrates.

References

1. Berstel, J., Reutenauer, C.: Recognizable formal power series on trees. Theoret. Comput. Sci. **18**(2), 115–148 (1982)
2. Bozapalidis, S.: Effective construction of the synthetic algebra of a recognizable series on trees. Acta Informatica **28**(4), 351–363 (1991)
3. Bozapalidis, S., Alexandrakis, A.: Représentations matricielles des séries d'arbre reconnaissables. RAIRO-Theoretical Informatics and Applications **23**(4), 449–459 (1989)
4. Bozapalidis, S., Kalampakas, A.: Graph automata. Theoretical Computer Science **393**(1-3), 147–165 (2008). https://doi.org/10.1016/j.tcs.2007.11.022
5. Comon, H., Dauchet, M., Gilleron, R., Löding, C., Jacquemard, F., Lugiez, D., Tison, S., Tommasi, M.: Tree automata — Techniques and applications (2007)
6. Creus, C., Gascón, A., Godoy, G., Ramos, L.: The HOM problem is EXPTIME-complete. In: Proc. 27th Annual IEEE Symp. Logic in Computer Science. pp. 255–264. IEEE (2012). https://doi.org/10.1109/LICS.2012.36
7. Doner, J.: Tree acceptors and some of their applications. J. Comput. System Sci. **4**(5), 406–451 (1970). https://doi.org/10.1016/S0022-0000(70)80041-1
8. Drewes, F.: Grammatical picture generation. Springer (2006)
9. Droste, M., Kuich, W., Vogler, H.: Handbook of weighted automata. Springer (2009)
10. Droste, M., Kuske, D.: Weighted automata. (2021)
11. Fülöp, Z., Vogler, H.: Weighted tree automata and tree transducers. In: Handbook of Weighted Automata [9], chap. 9, pp. 313–403. https://doi.org/10.4171/Automata-1/4
12. Gécseg, F., Steinby, M.: Tree automata. Tech. Rep. 1509.06233, arXiv (2015)
13. Gilleron, R., Tison, S.: Regular tree languages and rewrite systems. Fundamenta informaticae **24**(1-2), 157–175 (1995)
14. Godoy, G., Giménez, O.: The HOM problem is decidable. J. ACM **60**(4), 1–44 (2013)
15. Godoy, G., Giménez, O., Ramos, L., Àlvarez, C.: The hom problem is decidable. In: Proc. 42nd ACM Symp. Theory of Computing. pp. 485–494. ACM (2010). https://doi.org/10.1145/1806689.1806757

16. Golan, J.S.: Semirings and their Applications. Kluwer Academic, Dordrecht (1999). https://doi.org/10.1007/978-94-015-9333-5
17. Hebisch, U., Weinert, H.J.: Semirings. World Scientific (1998). https://doi.org/10.1142/3903
18. Hebisch, U., Weinert, H.J.: Semirings: algebraic theory and applications in computer science, vol. 5. World Scientific (1998)
19. Jurafsky, D., Martin, J.H.: Speech and language processing. Prentice Hall (2008)
20. Maletti, A., Nász, A.T.: Weighted tree automata with constraints. In: Developments in Language Theory: 26th International Conference, DLT 2022, Tampa, FL, USA, May 9–13, 2022, Proceedings. pp. 226–238. Springer (2022)
21. Maletti, A., Nász, A.T.: Weighted tree automata with constraints. Theory of Computing Systems pp. 1–28 (2023)
22. Maletti, A., Nász, A., Paul, E.: Weighted hom-problem for nonnegative integers (2023). https://doi.org/10.48550/arXiv.2305.04117
23. Nász, A.T.: Solving the weighted hom-problem with the help of unambiguity. EPTCS 386 p. 200
24. Perrin, D.: Recent results on automata and infinite words. In: Proc. 11th Int. Symp. MFCS. LNCS, vol. 176, pp. 134–148. Springer (1984)
25. Ramsey, F.P.: On a problem of formal logic. Proc. London Math. Soc **30** (1930)
26. Rosenfeld, A.: Picture languages: formal models for picture recognition. Academic Press (2014)
27. Salomaa, A., Soittola, M.: Automata-Theoretic Aspects of Formal Power Series. Texts and Monographs in Computer Science, Springer (1978). https://doi.org/10.1007/978-1-4612-6264-0
28. Schützenberger, M.P.: On the definition of a family of automata. Inform. and Control **4**(2–3), 245–270 (1961). https://doi.org/10.1016/S0019-9958(61)80020-X
29. Schwentick, T.: Automata for xml—a survey. Journal of Computer and System Sciences **73**(3), 289–315 (2007)
30. Thatcher, J.W.: Characterizing derivation trees of context-free grammars through a generalization of finite automata theory. J. Comput. Syst. Sci. **1**(4), 317–322 (1967). https://doi.org/10.1016/S0022-0000(67)80022-9
31. Thatcher, J.W., Wright, J.B.: Generalized finite automata theory with an application to a decision problem of second-order logic. Math. Systems Theory **2**(1), 57–81 (1968). https://doi.org/10.1007/BF01691346
32. Wilhelm, R., Seidl, H., Hack, S.: Compiler Design - Syntactic and Semantic Analysis. Springer (2013). https://doi.org/10.1007/978-3-642-17540-4