Languages structured by the subword order and length relations*

Dietrich Kuske[1] and Georg Zetzsche[2]

[1] Technische Universität Ilmenau, Germany
dietrich.kuske@tu-ilmenau.de
[2] Max Planck Institute for Software Systems, Kaiserslautern, Germany
georg@mpi-sws.org

Abstract. We consider a language together with the subword relation, binary relations that relate the lengths of two words, and regular predicates. For such structures, we consider the extension of first-order logic by threshold- and modulo-counting quantifiers. Depending on the language, the used predicates, and the fragment of the logic, we determine four new combinations that yield decidable theories. These results extend earlier ones where only the language of *all* words *without the length relations* and fragments of *first-order logic* were considered.[3]

Keywords: Subword order · First-order logic · Counting quantifiers · Decidable theories

1 Introduction

The subword relation (sometimes called scattered subword relation) is one of the simplest nontrivial examples of a well-quasi ordering [7]. This property allows its prominent use in the verification of infinite state systems [4]. The subword relation can be understood as embeddability of one word into another. This embeddability relation has been considered for other classes of structures like trees, posets, semilattices, lattices, graphs etc. [14, 16, 15, 8–11, 24, 25].

In this paper, we study logical properties of a set of words ordered by the subword relation. We are mainly interested in general situations where we get a decidable logical theory. Regarding first-order logic, we already have a rather precise picture about the border between decidable and undecidable fragments: For the subword order alone, the $\exists^*$-theory is decidable [17] and the $\exists^*\forall^*$-theory is undecidable [12]. For the subword order together with regular predicates, the

* Part of the results were obtained when the second author was affiliated with LSV (ENS Paris-Saclay) and supported by a fellowship within the Postdoc-Program of the German Academic Exchange Service (DAAD) and by Labex DigiCosme, Université Paris-Saclay, project VERICONISS.

[3] In the conference version [19] of this paper, we did not consider the length relations, but the cover relation that can be expressed using the subword and the length relations.

two-variable theory is decidable [12] and the three-variable theory [12] as well as the $\exists^*$-theory are undecidable [6] (these two undecidabilities already hold if we only consider singleton predicates, i.e., constants).

Thus, to get a decidable theory, one has to restrict the expressiveness of first-order logic considerably. For instance, neither in the $\exists^*$-, nor in the two-variable fragment of first-order logic, one can express the cover relation "u is a proper subword of v and there is no word properly between these two". As another example, one cannot express threshold properties like "there are at most k subwords with a given property" in any of these two logics (for $k \geq 2$).

In this paper, we refine the analysis of logical properties of the subword order in four aspects:

- We restrict the universe from the set of all words to a given language L.
- As before, we may or may not add regular predicates or constants to the structure.
- Besides the subword order, we also consider predicates $\mathrm{ld}_{\geq \ell}$ expressing that the length difference between two words is at least the given constant ℓ.
- We add threshold and modulo counting quantifiers to the logic.

In other words, we consider reducts of the structure

$$(L, \sqsubseteq, (\mathrm{ld}_{\geq \ell})_{\ell \geq 0}, (K \cap L)_{K \text{ regular}}, (w)_{w \in L})$$

with L some language together with fragments of the logic C+MOD that extends first-order logic by threshold- and modulo-counting quantifiers.

In this spectrum, we identify four new cases of decidable theories:

1. The C+MOD-theory of the whole structure is decidable provided L is bounded and context-free (Theorem 3.5). A rather special case of this result follows from [6, Theorem 4.1]: If $L = (a_1^* a_2^* \cdots a_m^*)^\ell$ and if only regular predicates of the form $(a_1^* a_2^* \cdots a_m^*)^k$ are used, then the FO-theory is decidable (with $\Sigma = \{a_1, \ldots, a_m\}$).
2. The $\mathrm{C+MOD}^2$-theory of the whole structure is decidable whenever L is regular (Corollary 4.17). The decidability of the FO^2-theory without the length predicates follows from [12, Theorem 5.5], its complexity was bounded in [13], and this complexity bound was further improved and extended to the logic C^2 in [18].
3. The Σ_1-theory of the structure $(L, \sqsubseteq)$ is decidable provided L is regular (Theorem 5.1). For $L = \Sigma^*$, this is [17, Prop. 2.2].
4. The Σ_1-theory of the structure $(L, \sqsubseteq, (w)_{w \in L})$ is decidable provided L is regular and almost every word from L contains a non-negligible number of occurrences of every letter (see below for a precise definition, Theorem 6.2). Note that, by [6, Theorem 3.3], this theory is undecidable if $L = \Sigma^*$.

Our first result is shown by an interpretation of the structure in $(\mathbb{N}, +)$. Four ingredients are essential here: Parikh's theorem [20], the rationality of the subword relation [12], Nivat's theorem characterising rational relations [1], and the decidability of (an extension of) the FO+MOD-theory of $(\mathbb{N}, +)$ [5] (note that this decidability does not follow directly from Presburger's result since in his logic, one cannot make statements like "the number of witnesses $x \in \mathbb{N}$ satisfying ... is even").

Our second result extends a result from [12] that shows decidability of the FO^2-theory of the structure $(\Sigma^*, \sqsubseteq, (L)_{L \text{ regular}})$. Karandikar and Schnoebelen provide a quantifier elimination procedure which relies on two facts:

- The class of regular languages is closed under images of rational relations.
- The proper subword relation and the incomparability relation are rational.

Here, we follow a similar proof strategy. But while that proof had to handle the existential quantifier, only, we also have to deal with counting quantifiers. This requires us to develop a theory of counting images under rational relations, e.g., the set of words u such that there are at least two words v in the regular language K with (u, v) in the rational relation R. We show that the class of regular languages is closed under such counting images provided the rational relation R is unambiguous, a proof that makes heavy use of weighted automata [22]. To apply this to the subword relation and the length predicates, this also requires us to show that certain intersections of these relations (e.g., the set of pairs (u, v) such that u and v are no subwords of each other and of equal length) are unambiguous rational.

Our third result extends the decidability of the Σ_1-theory of $(\Sigma^*, \sqsubseteq)$ from [17]. The main point there was to prove that every finite partial order can be embedded into $(\Sigma^*, \sqsubseteq)$ if $|\Sigma| \geq 2$. This is certainly false if we restrict the universe, e.g., to $L = a^*$. However, such bounded regular languages are already covered by the first result, so we only have to handle unbounded regular languages L. In that case, we prove nontrivial combinatorial results regarding primitive words in regular languages and prefix-maximal subwords. These considerations then allow us to prove that, indeed, every finite partial order embeds into $(L, \sqsubseteq)$. Then, the decidability of the Σ_1-theory follows as in [17].

Regarding our fourth result, we know from [12] that decidability of the Σ_1-theory of $(L, \sqsubseteq, (w)_{w \in L})$ does not hold for every regular L. Therefore, we require that a certain fraction of the positions in a word carries the letter a (for almost all words from the language and for all letters). This allows us to conclude that every finite partial order embeds into $(L, \sqsubseteq)$ above each word. The second ingredient is that, for such languages, any Σ_1-sentence is effectively equivalent to such a sentence where constants are only used to express that all variables take values above a certain word w. These two properties together with some combinatorial arguments from the theory of well-quasi orders then yield the decidability.

In summary, we identify four classes of decidable theories related to the subword order. In this paper, we concentrate on these positive results, i.e., we did not try to find new undecidable theories. It would, in particular, be nice to understand what properties of the regular language L determine the decidability of the Σ_1-theory of the structure $(L, \sqsubseteq, (w)_{w \in L})$ (it is undecidable for $L = \Sigma^*$ and decidable for, e.g., $L = \{ab, baa\}^* \cup bb\{abb\}^*$). Another open question concerns the complexity of our decidability results.

2 Preliminaries

By $\mathbb{N}$, we denote the set of non-negative integers or natural numbers $\{0, 1, \dots\}$ and $[m, n]$ denotes the set $\{m, m+1, \dots, n-1, n\}$ for $m, n \in \mathbb{N}$ with $m \leq n$. A set $X \subseteq \mathbb{N}^n$ of n-tuples of natural number is *semi-linear*, if it is a finite union of sets of the form

$$\left\{\overline{m} + \sum_{1 \leq i \leq k} a_i \overline{n_i} \colon a_i \in \mathbb{N}\right\}$$

where $k, m, n_1, \dots, n_k \in \mathbb{N}^n$ are n-tuples of natural numbers. This is the case if there exists a first-order formula $\varphi(x)$ in the language of $(\mathbb{N}, +)$ such that X is the set of natural numbers satisfying φ. Finally, a set $X \subseteq \mathbb{N}^n$ is semi-linear iff it is a rational subset of the monoid $(\mathbb{N}^n, +)$.

For an alphabet Σ, we write Σ^* for the set of words over Σ, ε for the empty word, and Σ^+ for the set of non-empty words over Σ.

Throughout this paper, let Σ be some alphabet. A word $u = a_1 a_2 \dots a_m$ with $a_1, a_2, \dots, a_m \in \Sigma$ is a *subword* of a word $v \in \Sigma^*$ if there are words $v_0, v_1, \dots, v_m \in \Sigma^*$ with $v = v_0 a_1 v_1 a_2 v_2 \cdots a_m v_m$. In that case, we write $u \sqsubseteq v$; if, in addition, $u \neq v$, then we write $u \sqsubset v$ and call u a proper subword of v. If, for two words u and v, neither u is a subword of v nor *vice versa*, then the words u and v are *incomparable* and we write $u \parallel v$. For instance, $aa \sqsubset babbba$ and $aba \parallel aabb$.

For a non-negative integer $\ell \in \mathbb{N}$, let $\mathrm{ld}_{\geq \ell} \subseteq \Sigma^* \times \Sigma^*$ denote the set of all pairs (u, v) of words such that $|v| - |u| \geq \ell$ (ld stands for length difference).

Let $\mathcal{S} = (L, (R_i)_{i \in I}, (w_j)_{j \in J})$ be a *structure*, i.e., L is a set, $R_i \subseteq L^{n_i}$ is a relation of arity n_i (for all $i \in I$), and $w_j \in L$ for all $j \in J$. Then, formulas φ of the logic C+MOD are defined by the following grammar:

$$\varphi ::= (s = t) \mid R_i(s_1, \dots, s_{n_i}) \mid \neg\varphi \mid \varphi \vee \varphi \mid \exists x\, \varphi \mid \exists^{\geq k} x\, \varphi \mid \exists^{p \bmod q} x\, \varphi$$

where $s, t, s_1, \dots, s_{n_i}$ are variables or constants w_j with $j \in J$, $i \in I$, $k \in \mathbb{N}$, and $p, q \in \mathbb{N}$ with $p < q$. We call $\exists^{\geq k}$ a *threshold counting quantifier* and $\exists^{p \bmod q}$ a *modulo counting quantifier*. The semantics of these quantifiers is defined as follows:

- $\mathcal{S} \models \exists^{\geq k} x\, \alpha$ iff $|\{w \in L \mid \mathcal{S} \models \alpha(w)\}| \geq k$
- $\mathcal{S} \models \exists^{p \bmod q} x\, \alpha$ iff $|\{w \in L \mid \mathcal{S} \models \alpha(w)\}| \in p + q\mathbb{N}$

For instance, $\exists^{0 \bmod 2} x\, \alpha$ expresses that the number of elements of the structure satisfying α is even. Then $\left(\exists^{0 \bmod 2} x\, \alpha\right) \vee \left(\exists^{1 \bmod 2} x\, \alpha\right)$ holds iff only finitely many elements of the structure satisfy α. The fragment FO+MOD of C+MOD comprises all formulas not containing any threshold counting quantifier. First-order logic FO is the set of formulas from C+MOD not mentioning any counting quantifier. Let Σ_1 denote the set of first-order formulas of the form $\exists x_1\, \exists x_2 \dots \exists x_n \colon \psi$ where ψ is quantifier-free; these formulas are also called *existential*. Similarly, Σ_2 comprises all first-order formulas of the form $\exists x_1\, \exists x_2 \dots \exists x_m\, \forall y_1\, \forall y_2 \dots \forall y_n \colon \psi$ where ψ is quantifier-free.

The threshold quantifier $\exists^{\geq k}$ can be expressed using the existential quantifier, only, since the formulas $\exists^{\geq k} x\, \alpha$ and

$$\exists x_1, x_2, \ldots, x_k \,\forall x\colon \bigvee_{1 \leq i \leq k} x = x_i \to \alpha$$

are equivalent. Consequently, the logics FO+MOD and C+MOD are equally expressive.[4] The situation changes when we restrict the number of variables that can be used in a formula: let FO^2 and $\mathrm{C{+}MOD}^2$ denote the set of formulas from FO and C+MOD, respectively, that use the variables x and y, only. Then, the existence of ≥ 3 elements in the structure is expressible in $\mathrm{C{+}MOD}^2$, but not in $\mathrm{FO{+}MOD}^2$.

In this paper, we will consider the following structures:

- The largest one is $(L, \sqsubseteq, (\mathrm{ld}_{\geq \ell})_{\ell \in \mathbb{N}}, (K \cap L)_{K \text{ regular}}, (w)_{w \in L})$ for some $L \subseteq \Sigma^*$. The universe of this structure is the language L, we have the binary subword predicate, a binary predicate $\mathrm{ld}_{\geq \ell}$ for every non-negative integer $\ell \in \mathbb{N}$, a unary predicate $K \cap L$ for every regular language L, and we can use every word from L as a constant.[5]
- The other extreme is the structure $(L, \sqsubseteq)$ for some $L \subseteq \Sigma^*$ where we consider only the binary predicate $\sqsubseteq$.
- Finally, we will also prove results on the intermediate structure $(L, \sqsubseteq, (w)_{w \in L})$ that has a binary relation and any word from the language as a constant.

For any structure $\mathcal{S}$ and any of the logics $\mathcal{L}$, the *$\mathcal{L}$-theory* of $\mathcal{S}$ is the set of sentences (i.e., formulas without free variables) from $\mathcal{L}$ that hold in $\mathcal{S}$.

A language $L \subseteq \Sigma^*$ is *bounded* if there are a number $n \in \mathbb{N}$ and words $w_1, w_2, \ldots, w_n \in \Sigma^*$ such that $L \subseteq w_1^* w_2^* \cdots w_n^*$. Otherwise, it is *unbounded*.

For an alphabet Σ, a word $w \in \Sigma^*$, and a letter $a \in \Sigma$, let $|w|_a$ denote the number of occurrences of the letter a in the word w. The *Parikh vector* of w is the tuple $\Psi_\Sigma(w) = (|w|_a)_{a \in \Sigma} \in \mathbb{N}^\Gamma$. Note that Ψ_Σ is a homomorphism from the free monoid Σ^* onto the additive monoid $(\mathbb{N}^\Sigma, +) \cong (\mathbb{N}^{|\Sigma|}, +)$.

A word $w \in \Sigma^*$ is *primitive* if $w \neq \varepsilon$ cannot be written as a proper power of some word v, i.e., $w \notin vv^+$ for any word v. Two words u and v are *conjugate* if there are words x and y such that $u = xy$ and $v = yx$.

3 The C+MOD-theory with regular predicates

Let $L \subseteq \Sigma^*$ be a bounded and context-free language. The aim of this section is to prove that the full FO+MOD-theory of the structure

$$(L, \sqsubseteq, (\mathrm{ld}_{\geq \ell})_{\ell \in \mathbb{N}}, (K \cap L)_{K \text{ regular}}, (w)_{w \in L})$$

[4] If the threshold k is written in binary, then the second formula is exponentially longer, i.e., the use of the threshold quantifier allows to write formulas more succinctly.

[5] Note that u is a proper subword of v and there is no word in between holds iff $u \sqsubseteq v \wedge \mathrm{ld}_{\geq 1}(u, v) \wedge \neg \mathrm{ld}_{\geq 2}(u, v)$; hence the cover relation considered in the conference version [19] can be expressed in the more general setting of this paper.

is decidable. This is achieved by interpreting this structure in $(\mathbb{N}, +)$, i.e., in Presburger arithmetic whose C+MOD-theory is known to be decidable [5]. We start with four preparatory lemmas.

Lemma 3.1. *Let $w_1, \dots, w_n \in \Sigma^*$, $g\colon \mathbb{N}^n \to \Sigma^*$ be defined by $g(\overline{m}) = w_1^{m_1} w_2^{m_2} \cdots w_n^{m_n}$ for all $\overline{m} = (m_1, m_2, \dots, m_n) \in \mathbb{N}^n$, and let $\ell \in \mathbb{N}$. The set $\{(\overline{m}, \overline{n}) \in \mathbb{N}^n \times \mathbb{N}^n \mid (g(\overline{m}), g(\overline{n})) \in \mathrm{ld}_{\geq \ell}\}$ is effectively semilinear.*

Proof. Note that the length of $g(\overline{m})$ equals $\sum_{1 \leq i \leq n} m_i \cdot |w_i|$. Hence $(g(\overline{m}), g(\overline{n})) \in \mathrm{ld}_{\geq \ell}$ if, and only if,

$$\sum_{1 \leq i \leq n} n_i \cdot |w_i| - \sum_{1 \leq i \leq n} m_i \cdot |w_i| \geq \ell\,.$$

Here, the values $|w_i|$ and ℓ act as constants. Hence, this is an expression in the language of $(\mathbb{N}, +)$ in the variables m_i and n_i that can be constructed from the words w_i and the natural number ℓ. Consequently, the relation in question is effectively semilinear. ◻

Lemma 3.2. *Let $K \subseteq \Sigma^*$ be context-free, $w_1, \dots, w_n \in \Sigma^*$, and $g\colon \mathbb{N}^n \to \Sigma^*$ be defined by $g(\overline{m}) = w_1^{m_1} w_2^{m_2} \cdots w_n^{m_n}$ for all $\overline{m} = (m_1, m_2, \dots, m_n) \in \mathbb{N}^n$. The set $g^{-1}(K) = \{\overline{m} \in \mathbb{N}^n \mid g(\overline{m}) \in K\}$ is effectively semilinear.*

Proof. Let $\Gamma = \{a_1, a_2, \dots, a_n\}$ be an alphabet and define the monoid homomorphism $f\colon \Gamma^* \to \Sigma^*$ by $f(a_i) = w_i$ for all $i \in [1, n]$.

Since the class of context-free languages is effectively closed under inverse homomorphisms and under intersections with regular languages, the language

$$L = f^{-1}(K) \cap a_1^* a_2^* \cdots a_n^* = \{u \in a_1^* a_2^* \cdots a_n^* \mid f(u) \in K\}$$

is effectively context-free. Its Parikh image $\Psi_\Gamma(L) \subseteq \mathbb{N}^n$ is effectively semilinear [20].

It remains to be verified that $\Psi_\Gamma(L)$ equals the set from the lemma.

So let $\overline{m} \in \mathbb{N}^n$. Then $\overline{m} \in \Psi_\Gamma(L)$ iff there exists a word $u \in L$ with Parikh image $\overline{m}$. But the only word from $a_1^* a_2^* \dots a_n^*$ with this Parikh image is $a_1^{m_1} a_2^{m_2} \cdots a_n^{m_n}$. Consequently, $\overline{m} \in \Psi_\Gamma(L)$ iff $a_1^{m_1} a_2^{m_2} \cdots a_n^{m_n} \in f^{-1}(K)$. Since $f(a_1^{m_1} a_2^{m_2} \cdots a_n^{m_n}) = g(\overline{m})$, this is equivalent to $g(\overline{m}) \in K$. Thus, the semilinear set $\Psi_\Gamma(L)$ equals the set $g^{-1}(K)$ from the lemma. ◻

Lemma 3.3. *Let $w_1, \dots, w_n \in \Sigma^*$ and $g\colon \mathbb{N}^n \to \Sigma^*$ be defined by $g(\overline{m}) = w_1^{m_1} w_2^{m_2} \cdots w_n^{m_n}$ for all $\overline{m} = (m_1, m_2, \dots, m_n) \in \mathbb{N}^n$. The set $\{(\overline{m}, \overline{n}) \in \mathbb{N}^n \times \mathbb{N}^n \mid g(\overline{m}) \sqsubseteq g(\overline{n})\}$ is semilinear.*

Proof. Let $\Gamma = \{a_1, a_2, \dots, a_n\}$ be an alphabet and define the monoid-homomorphism $f\colon \Gamma^* \to \Sigma^*$ by $f(a_i) = w_i$ for all $i \in [1, n]$.

The subword relation

$$S = \{(U, V) \in \Sigma^* \times \Sigma^* \mid U \sqsubseteq V\}$$

on Σ^* is rational [12]. Since the class of rational relations is closed under inverse homomorphisms [1], also the relation

$$S_1 = \{(u, v) \in \Gamma^* \times \Gamma^* \mid f(v) \sqsubseteq f(v)\}$$

is rational. Let $T = \{(u, u) \mid u \in a_1^* a_2^* \cdots a_n^*\}$. Then T is a rational relation. Since the class of rational relations is closed under composition, also $S_2 = T \circ S_1 \circ T$ is rational [1]. Note that

$$S_2 = \{(u, v) \mid u, v \in a_1^* a_2^* \ldots a_n^*, f(v) \sqsubseteq f(v)\}\,.$$

By Nivat's theorem [1], there are a regular language R over some alphabet Δ and two homomorphisms $h_1, h_2 \colon \Delta^* \to \Gamma^*$ with

$$S_2 = \{\big(h_1(w), h_2(w)\big) \mid w \in R\}\,.$$

Since R is regular, its Parikh-image

$$\Psi_\Delta(R) = \{\Psi_\Delta(w) \mid w \in R\}$$

is semilinear [20].

The monoid $(\mathbb{N}^\Delta, +)$ is the commutative monoid freely generated by the vectors $\Psi_\Delta(b) = (0, \ldots, 0, 1, 0, \ldots, 0)$ for $b \in \Delta$. Since also $(\mathbb{N}^n, +)$ is commutative, we can define monoid homomorphisms $p_1, p_2 \colon \mathbb{N}^\Delta \to \mathbb{N}^n$ by $p_1(\Psi_\Delta(b)) = \Psi_\Gamma(h_1(b))$ and $p_2(\Psi_\Delta(b)) = \Psi_\Gamma(h_2(b))$ for all $b \in \Delta$.

Since the class of rational sets is closed under monoid homomorphisms [21, Prop. 1.1.7(i)], the image

$$H = \{\big(p_1(\Psi_\Delta(w)), p_2(\Psi_\Delta(w))\big) \mid w \in R\}$$

of the set $\Psi_\Delta(R)$ under the monoid homomorphism $(p_1, p_2) \colon \mathbb{N}^\Delta \to \mathbb{N}^n \times \mathbb{N}^n$ is semilinear.

Direct calculations yield $\Psi_\Gamma(h_i(w)) = p_i(\Psi_\Delta(w))$ for all $i \in \{1, 2\}$ and $w \in \Delta^*$.

Let $\overline{m}, \overline{n} \in \mathbb{N}^n$. Then $(\overline{m}, \overline{n}) \in H$ iff there exists $w \in R$ with $\overline{m} = p_1(\Psi_\Delta(w)) = \Psi_\Gamma(h_1(w))$ and $\overline{n} = p_2(\Psi_\Delta(w)) = \Psi_\Gamma(h_2(w))$. The existence of such a word $w \in R$ is equivalent to the existence of a pair $(u, v) \in S_2$ with $\overline{m} = \Psi_\Gamma(u)$ and $\overline{n} = \Psi_\Gamma(v)$.

Since all words appearing in S_2 belong to $a_1^* a_2^* \cdots a_n^*$, this last statement is equivalent to saying

$$(a_1^{m_1} a_2^{m_2} \cdots a_n^{m_n}, a_1^{n_1} a_2^{n_2} \cdots a_n^{n_n}) \in S_2\,.$$

But this is equivalent to saying

$$f(a_1^{m_1} a_2^{m_2} \cdots a_n^{m_n}) \sqsubseteq f(a_1^{n_1} a_2^{n_2} \cdots a_n^{n_n})\,.$$

Note that $g(\overline{m}) = f(a_1^{m_1} a_2^{m_2} \cdots a_n^{m_n})$ and similarly $g(\overline{n}) = f(a_1^{n_1} a_2^{n_2} \cdots a_n^{n_n})$. Thus, the last claim is equivalent to $g(\overline{m}) \sqsubseteq g(\overline{n})$.

In summary, we showed that the semilinear set H is the set from the lemma. $\square$

Lemma 3.4. *Let $w_1, w_2, \ldots, w_n \in \Sigma^*$, $L \subseteq w_1^* w_2^* \cdots w_n^*$ be context-free, and $g \colon \mathbb{N}^n \to \Sigma^*$ be defined by $g(\overline{m}) = w_1^{m_1} w_2^{m_2} \cdots w_n^{m_n}$ for every tuple $\overline{m} = (m_1, m_2, \ldots, m_n) \in \mathbb{N}^n$. Then there exists a semilinear set $U \subseteq \mathbb{N}^n$ such that g maps U bijectively onto L.*

Proof. By Lemma 3.2, the set $g^{-1}(L)$ is semilinear and satisfies, by its definition, $g(g^{-1}(L)) = L \cap a_1^* a_2^* \cdots a_n^* = L$.

Let T denote the semilinear set from Lemma 3.3.

Then let U denote the set of n-tuples $\overline{m} \in g^{-1}(L)$ such that the following holds for all $\overline{n} \in \mathbb{N}^n$:

If $(\overline{m}, \overline{n}) \in T$ and $(\overline{n}, \overline{m}) \in T$, then $\overline{m}$ is lexicographically smaller than or equal to $\overline{n}$.

This set U is semilinear since the class of semilinear relations is closed under first-order definitions. Now let $u \in L$. Since g maps $g^{-1}(L)$ onto L, there is $\overline{m} \in g^{-1}(L)$ with $g(\overline{m}) = u$. Since, on $g^{-1}(L) \subseteq \mathbb{N}^n$, the lexicographic order is a well-order, there is a lexicographically minimal such tuple $\overline{m}$. This tuple belongs to U and it is the only tuple from U mapped to u. $\square$

Now we can prove the main result of this section.

Theorem 3.5. *Let $L \subseteq \Sigma^*$ be context-free and bounded. Then the* FO+MOD-*theory of $\mathcal{S} = (L, \sqsubseteq, (\mathrm{ld}_{\geq \ell})_{\ell \in \mathbb{N}}, (K \cap L)_{K \text{ regular}}, (w)_{w \in L})$ is decidable.*

Proof. Since L is bounded, there are words $w_1, w_2, \ldots, w_n \in \Sigma^*$ such that $L \subseteq w_1^* \, w_2^* \cdots w_n^*$. For an n-tuple $\overline{m} = (m_1, m_2, \ldots, m_n) \in \mathbb{N}^n$ we define $g(\overline{m}) = w_1^{m_1} w_2^{m_2} \cdots w_n^{m_n} \in \Sigma^*$.

1. By Lemma 3.4, there is a semilinear set $U \subseteq \mathbb{N}^n$ that is mapped by g bijectively onto L.
2. For $\ell \in \mathbb{N}$, the set $\{(\overline{m}, \overline{n}) \colon |g(\overline{n})| - |g(\overline{m})| \geq \ell\}$ is effectively semilinear by Lemma 3.1.
3. The set $\{(\overline{m}, \overline{n}) \mid g(\overline{m}) \sqsubseteq g(\overline{n})\}$ is semilinear by Lemma 3.3.
4. For any regular language $K \subseteq \Sigma^*$ the set $\{\overline{m} \in \mathbb{N}^n \mid g(\overline{m}) \in K\} \subseteq \mathbb{N}^n$ is effectively semilinear by Lemma 3.2.

From these semilinear sets, we obtain first-order formulas $\lambda(\overline{x})$, $\delta_{\geq \ell}(\overline{x}, \overline{y})$ for $\ell \in \mathbb{N}$, $\sigma(\overline{x}, \overline{y})$, and $\kappa_K(\overline{x})$ in the language of $(\mathbb{N}, +)$ such that, for any $\overline{m}, \overline{n} \in \mathbb{N}^n$, we have

1. $(\mathbb{N}, +) \models \lambda(\overline{m}) \iff \overline{m} \in U$,
2. $(\mathbb{N}, +) \models \delta_{\geq \ell}(\overline{m}, \overline{n}) \iff |g(\overline{n})| - |g(\overline{m})| \geq \ell$, and
3. $(\mathbb{N}, +) \models \sigma(\overline{m}, \overline{n}) \iff g(\overline{m}) \sqsubseteq g(\overline{n})$, and
4. $(\mathbb{N}, +) \models \kappa_K(\overline{m}) \iff g(\overline{m}) \in K$.

From a FO+MOD-formula $\varphi(x_1, \ldots, x_k)$ in the language of the structure $\mathcal{S}$, we now construct a formula $\varphi'(\overline{x_1}, \ldots, \overline{x_k})$ in the language of Presburger arithmetic $(\mathbb{N}, +)$ that uses existential and modulo-counting quantifiers over *tuples* of natural numbers (we will speak of FO+MOD(tuple)-formulas). To this aim, we set

– $\varphi' = \delta_{\geq\ell}(\overline{x},\overline{y})$ if $\varphi = \mathrm{ld}_{\geq\ell}(x,y)$,
– $\varphi' = \sigma(\overline{x},\overline{y})$ if $\varphi = (x \sqsubseteq y)$,
– $\varphi' = \kappa_K(\overline{x})$ if $\varphi = (x \in K)$,
– $\varphi' = (\exists \overline{x}\colon (\lambda(x) \wedge \alpha'))$ if $\varphi = (\exists x\colon \alpha)$, and
– $\varphi' = (\exists^{p \bmod q} \overline{x}\colon (\lambda(x) \wedge \alpha'))$ if $\varphi = (\exists^{p \bmod y} x\colon \alpha)$.

With the obvious semantics of FO+MOD(tuple)-formulas, we obtain for any sentence φ

$$\mathcal{S} \models \varphi \iff (\mathbb{N},+) \models \varphi'\,.$$

In [5], it was shown that the validity of FO+MOD(tuple)-sentences in $(\mathbb{N},+)$ is decidable (in twofold exponential space, i.e., is not much more complex than the validity of FO-sentences in $(\mathbb{N},+)$). □

Remark 3.6. Note that the size of the formula φ' in the above proof can be bounded by a polynomial in the size of φ and the sizes of the formulas λ, $\delta_{\geq\ell}$, σ, and κ_K. As we did not analyse the size of these formulas, we cannot bound the complexity of the above decision procedure.

4 The C+MOD^2-theory with regular predicates

It is the aim of this section to show that the C+MOD^2-theory of the structure $(L, \sqsubseteq, (\mathrm{ld}_{\geq\ell})_{\ell\in\mathbb{N}}, (K \cap L)_{K \text{ regular}}, (w)_{w\in L})$ is decidable for any regular language L. To this aim, we first show that the C+MOD^2-theory of

$$\mathcal{S} = (\Sigma^*, \sqsubseteq, , (\mathrm{ld}_{\geq\ell})_{\ell\in\mathbb{N}}, (L)_{L \text{ regular}})$$

is decidable. This decidability proof extends the proof from [12] for the decidability of the FO^2-theory of $(\Sigma^*, \sqsubseteq, (L)_{L \text{ regular}})$. It provides a quantifier-elimination procedure (see Section 4.3) that relies on the following two properties.

1. The class of regular languages is closed under *counting* images under *unambiguous* rational relations (Section 4.2) and
2. the proper subword and the incomparability relation in conjunction with length predicates $\mathrm{ld}_{\geq\ell}$ and $\mathrm{ld}_{=\ell} := (\mathrm{ld}_{\geq\ell} \setminus \mathrm{ld}_{\geq\ell+1})$ are *unambiguous* rational (Section 4.1).

4.1 Unambiguous rational relations

In this section, we define unambiguous rational relations and provide some examples that will turn out very useful in our study of the theory of the structure $\mathcal{S}$.

Recall that, by Nivat's theorem [1], a relation $R \subseteq \Sigma^* \times \Sigma^*$ is rational if there exist an alphabet Γ, a homomorphism $h\colon \Gamma^* \to \Sigma^* \times \Sigma^*$, and a regular language $S \subseteq \Gamma^*$ such that h maps S surjectively onto R. We call R an *unambiguous rational relation* if, in addition, h maps S *injectively* (and therefore bijectively) onto R.

Remark 4.1. Let M be a monoid. Then $\mathrm{URat}(M)$ [21, page 229] is the least set $\mathcal{C}$ of subsets of M such that the following hold:

- Any finite subset of M belongs to $\mathcal{C}$.
- If $K, L \in \mathcal{C}$ are disjoint, then $K \cup L \in \mathcal{C}$.
- If $K, L \in \mathcal{C}$ and all $(u, v), (u', v') \in K \times L$ with $uv = u'v'$ satisfy $(u, v) = (u', v')$, then $K \cdot L \in \mathcal{C}$.
- Suppose $K \in \mathcal{C}$ such that, for all $m, n \in \mathbb{N}$ and $u_1, \ldots, u_m, v_1, \ldots, v_n \in K$ with $u_1 u_2 \cdots u_m = v_1 v_2 \cdots v_n$, we have $m = n$ and $u_i = v_i$ for all $i \in [1, n]$. Then $K^* \in \mathcal{C}$.

Let $M = \Sigma^* \times \Sigma^*$. Then any relation from $\mathrm{URat}(M)$ is rational, but there are rational relations that do not belong to $\mathrm{URat}(M)$ [21, Fact 2.1.1].

If $M = \Sigma^*$ for some alphabet Σ, then $\mathrm{URat}(M)$ equals the class of regular languages [21, Thm. 2.1.3] over Σ. By [21, Prop. 2.1.13], it follows that a relation is unambiguous rational (as defined above) if, and only if, it belongs to $\mathrm{URat}(\Sigma^* \times \Sigma^*)$ – this justifies our above definition since, traditionally, relations in $\mathrm{URat}(\Sigma^* \times \Sigma^*)$ are called unambiguous rational.

The following example demonstrates that the class of unambiguous rational relations is neither closed under intersections nor, and this is different from the class of rational relations, under unions. Following this example, we will see that it is closed under *disjoint* unions.

Example 4.2. The relations $R_1 = \{(a^m b a^n, a^m) \mid m, n \in \mathbb{N}\}$ and $R_2 = \{(a^m b a^n, a^n) \mid m, n \in \mathbb{N}\}$ are unambiguous rational: take $\Gamma = \{x, y, z\}$, $S = x^* y z^*$ and the homomorphisms

$$x \mapsto (a, a) \; y \mapsto (b, \varepsilon) \; z \mapsto (a, \varepsilon)$$

(for the first relation) and

$$x \mapsto (a, \varepsilon) \; y \mapsto (b, \varepsilon) \; z \mapsto (a, a)\,.$$

Note that the intersection $R_1 \cap R_2 = \{(a^m b a^m, a^m) \mid m \in \mathbb{N}\}$ is not even rational, while the union $R = R_1 \cup R_2$ is rational (since the union of rational relations is always rational) [1]. But this union is not unambiguous rational: if it was unambiguous rational, then the set

$$\{u \in a^* b a^* \mid \exists^{\geq 2} v \colon (u, v) \in R\} = \{a^m b a^n \mid m \neq n\}$$

would be regular by Proposition 4.14 below.

Lemma 4.3. *Let $R_1, R_2 \subseteq \Sigma^* \times \Sigma^*$ be unambiguous rational and disjoint. Then $R_1 \cup R_2$ is unambiguous rational.*

Proof. There are disjoint alphabets Γ_1 and Γ_2, regular languages $S_i \subseteq \Gamma_i^*$ and homomorphisms $h_i \colon \Gamma_i^* \to \Sigma^* \times \Sigma^*$ such that S_i is mapped bijectively onto R_i. Let $\Gamma = \Gamma_1 \cup \Gamma_2$ and $S = S_1 \cup S_2$. Let the homomorphism $h \colon \Gamma^* \to \Sigma^* \times \Sigma^*$ be given by

$$h(a) = \begin{cases} h_1(a) & \text{if } a \in \Gamma_1 \\ h_2(a) & \text{if } a \in \Gamma_2 \end{cases}$$

for $a \in \Gamma$. Then h maps the regular language S bijectively onto $R_1 \cup R_2$. ☐

We next prepare the proof that the relations

$$\sqsubset \cap \, \mathrm{ld}_{\geq \ell} = \{(u,v) \mid u \sqsubset v \text{ and } |v| - |u| \geq \ell\}\,, \text{ and}$$
$$\sqsubset \cap \, \mathrm{ld}_{= \ell} = \{(u,v) \mid u \sqsubset v \text{ and } |v| - |u| = \ell\}$$

are unambiguous rational. Towards this aim, we first define a language Sub and a homomorphism proj that maps Sub bijectively onto the subword relation $\sqsubseteq$. In addition, we characterize when a word $w \in \mathrm{Sub}$ with $\mathrm{proj}(w) = (u,v)$ satisfies $\mathrm{ld}_{\geq \ell}(u,v)$.

Lemma 4.4. *Let Σ be some alphabet and $\Sigma' = \{a' \mid a \in \Sigma\}$ be a disjoint copy of Σ. Let further* $\mathrm{proj}\colon (\Sigma \cup \Sigma')^* \to \Sigma^* \times \Sigma^*$ *be the homomorphism defined by* $\mathrm{proj}(a) = (\varepsilon, a)$ *and* $\mathrm{proj}(a') = (a,a)$ *for all $a \in \Sigma$. Finally, let* Sub *be the set of words $w \in (\Sigma \cup \Sigma')^*$ of the form*

$$w = v_1 a_1' \, v_2 a_2' \cdots v_n a_n' \, v_{n+1}$$

for some $n \geq 0$, $v_1, v_2, \ldots, v_{n+1} \in \Sigma^*$, $a_1, \ldots, a_n \in \Sigma$ such that a_i does not appear in v_i for any $i \in [1,n]$.

(i) For any $u, v \in \Sigma^*$, we have $u \sqsubseteq v$ if, and only if, there is some word $w \in \mathrm{Sub}$ with $\mathrm{proj}(w) = (u,v)$.
(ii) If $w_1, w_2 \in \mathrm{Sub}$ satisfy $\mathrm{proj}(w_1) = \mathrm{proj}(w_2)$, then $w_1 = w_2$.
(iii) If $w, \overline{w} \in \mathrm{Sub}$ satisfy $\mathrm{proj}(w) = \mathrm{proj}(\overline{w})$, then $w = \overline{w}$.
(iv) If $w \in \mathrm{Sub}$ satisfies $\mathrm{proj}(w) = (u,v)$, then $|v| - |u| = \sum_{a \in \Sigma} |w|_a$.

Proof. (i) For the direction "$\Leftarrow$", let $w \in \mathrm{Sub}$ be arbitrary. Then there exist $n \geq 0$, $v_1, \ldots, v_{n+1} \in \Sigma^*$, and $a_1, \ldots, a_n \in \Sigma$ such that a_i does not appear in v_i for any $i \in [1,n]$ with $w = v_1 a_1' \, v_2 a_2' \cdots v_n a_n' \, v_{n+1}$. With $u = a_1 a_2 \cdots a_n$ and $v = v_1 a_1 \, v_2 a_2 \cdots v_n a_n \, v_{n+1}$, we get $\mathrm{proj}(w) = (u,v)$ and $u \sqsubseteq v$.

For the converse implication, suppose $u \sqsubseteq v$. Let $u = a_1 a_2 \cdots a_n$ with $a_i \in \Sigma$ for all $i \in [1,n]$. We construct a word $w \in \mathrm{Sub}$ with $\mathrm{proj}(w) = (u,v)$ by induction on n. If $n = 0$, we can set $w = v$ and obtain $w \in \mathrm{Sub}$ with $\mathrm{proj}(w) = (u,v)$. Now let $n > 0$. Let v_1 be the maximal prefix of v not containing a_1. Since $a_1 \sqsubseteq u \sqsubseteq v$, we obtain $v_1 \neq v$, hence $v_1 a_1$ is a prefix of v, let $\overline{v} \in \Sigma^*$ be the complementary suffix such that $v = v_1 a_1 \, \overline{v}$. Since a_1 does not appear in v_1, we get $a_2 a_3 \cdots a_n \sqsubseteq \overline{v}$. By the induction hypothesis, there exists $\overline{w} \in \mathrm{Sub}$ with $\mathrm{proj}(\overline{w}) = (a_2 \cdots a_n, \overline{v})$. Note that $w = v_1 a_1' \, \overline{w}$ is a word from Sub and satisfies

$$\mathrm{proj}(w) = \mathrm{proj}(v_1 a_1') \, \mathrm{proj}(\overline{w}) = (a_1, v_1 a_1) \, (a_2 \cdots a_n, \overline{v}) = (u,v)\,.$$

(ii) There are $m, n \geq 0$, $v_1, \ldots, v_{m+1}, \overline{v_1}, \ldots, \overline{v_{n+1}} \in \Sigma^*$, and $a_1, \ldots a_m, b_1, \ldots, b_n \in \Sigma$ such that a_i does not appear in v_i for any $i \in [1,m]$, b_i does not appear in $\overline{v_i}$ for any $i \in [1,m]$,

$$w = v_1 a_1' \, \cdots v_m a_m' \, v_{m+1}, \text{ and } \overline{w} = \overline{v_1} b_1' \, \cdots \overline{v_n} b_n' \, \overline{v_{n+1}}\,.$$

Then

$$\begin{aligned}(a_1 \cdots a_m, v_1 a_1 \, \cdots v_m a_m \, v_{m+1}) &= \text{proj}(w)\\ &= \text{proj}(\overline{w})\\ &= (b_1 \cdots b_n, \overline{v_1} b_1 \, \cdots \overline{v_n} b_n \, \overline{v_{n+1}})\end{aligned}$$

implies $m = n$ and $a_i = b_i$ for all $i \in [1, n]$.
We now proceed by induction on m. If $m = 0$, then

$$(\varepsilon, w) = \text{proj}(w) = \text{proj}(\overline{w}) = (\varepsilon, \overline{w})$$

ensures $w = \overline{w}$. Now let $m \geq 1$. Note that v_1 is the maximal prefix of $v_1 a_1 \, \cdots v_m a_m \, v_{m+1}$ not containing a_1 and, similarly, $\overline{v_1}$ is the maximal prefix of

$$\overline{v_1} b_1 \, \cdots \overline{v_m} b_m \, \overline{v_{m+1}} = v_1 a_1 \, \cdots v_m a_m \, v_{m+1}$$

not containing $b_1 = a_1$. Hence $\overline{v_1} = v_1$ and, by induction, $\overline{v_i} = v_i$ for all $i \in [1, n+1]$. Thus, indeed, $w = \overline{w}$.

(iii) There are $n \geq 0$, $v_1, \ldots, v_{n+1} \in \Sigma^*$, and $a_1, \ldots a_n \in \Sigma$ such that a_i does not appear in v_i for any $i \in [1, n]$ and

$$w = v_1 a_1' \, \cdots v_n a_n' \, v_{n+1} \, .$$

From $\text{proj}(w) = (u, v)$, we obtain $u = a_1 a_2 \cdots a_n$ and $v = v_1 a_1 \, \cdots v_n a_n \, v_{n+1}$. Hence

$$|v| - |u| = |v_1 v_2 \cdots v_{n+1}| = \sum_{a \in \Sigma} |w|_a \, .$$

□

Next, we derive from the above lemma that, indeed, the relations $\sqsubset \cap \, \text{ld}_{\geq \ell}$ and $\sqsubset \cap \, \text{ld}_{= \ell}$ are effectively unambiguous rational.

Proposition 4.5. *For any alphabet Σ and any natural number $\ell \in \mathbb{N}$, the relations $\sqsubset \cap \, \text{ld}_{\geq \ell}$ and $\sqsubset \cap \, \text{ld}_{= \ell}$ are effectively unambiguous rational.*

Proof. Let Σ', Sub, and proj be defined as in Lemma 4.4. Furthermore, let $\text{LD}_{\geq \ell} \subseteq (\Sigma \cup \Sigma')^*$ denote the set of all words w with $\sum_{a \in \Sigma} |w|_a \geq \ell$. The languages

$$\text{Sub} = \left(\bigcup_{a \in \Sigma} \left(\left(\Sigma \setminus \{a\} \right)^* a' \right) \right)^* \Sigma^* \, .$$

and $\text{LD}_{\geq \ell}$ are effectively regular. Hence, the same applies to their intersection. By Lemma 4.5, the homomorphism proj maps $\text{Sub} \cap \text{LD}_{\geq \ell}$ bijectively onto $\sqsubseteq \cap \, \text{ld}_{\geq \ell}$. Hence this relation is effectively unambiguous rational. Now the first claim follows from $\sqsubset \cap \, \text{ld}_{\geq 0} = \sqsubseteq \cap \, \text{ld}_{\geq 1}$ and $\sqsubset \cap \, \text{ld}_{\geq \ell} = \sqsubseteq \cap \, \text{ld}_{\geq \ell}$ for all $\ell \geq 1$.

Similarly, the language $\text{Sub} \cap \text{LD}_{\geq \ell} \setminus \text{LD}_{\geq \ell + 1}$ is effectively regular and proj maps it bijectively onto $\sqsubseteq \cap \, \text{ld}_{= \ell}$. Now the second claim follows from $\sqsubset \cap \, \text{ld}_{= 0} = \emptyset$ and $\sqsubset \cap \, \text{ld}_{= \ell} = \sqsubseteq \cap \, \text{ld}_{= \ell}$ for all $\ell \geq 1$. □

Recall that, for any alphabet Σ, the incomparability relation $\parallel$ is defined as

$$\parallel = \{(u,v) \in \Sigma^* \times \Sigma^* \mid \text{neither } u \sqsubseteq v \text{ nor } v \sqsubseteq u\}\,.$$

Our next aim is to show that also the relations

$$\begin{aligned} \parallel \cap \operatorname{ld}_{\geq \ell} &= \{(u,v) \mid u \parallel v \text{ and } |v| - |u| \geq \ell\}\,, \text{ and} \\ \parallel \cap \operatorname{ld}_{= \ell} &= \{(u,v) \mid u \parallel v \text{ and } |v| - |u| = \ell\} \end{aligned}$$

are unambiguous rational. Here we proceed similarly to above: we first define a language Inc and a homomorphism proj that maps Inc bijectively onto the set of all pairs of word (u,v) with $u \parallel v$ and $|u| \leq |v|$. In addition, we characterize when a word $w \in \mathrm{Inc}$ with $\mathrm{proj}(w) = (u,v)$ satisfies $\operatorname{ld}_{\geq \ell}(u,v)$.

Lemma 4.6. *Let Σ be some alphabet, $\Sigma_i = \{a^{(i)} \mid a \in \Sigma\}$ for $i \in \{1,2\}$ be mutually disjoint copies of Σ and set $\Gamma = \Sigma \cup \Sigma_1 \cup \Sigma_2$. Let furthermore* $\mathrm{proj}\colon \Gamma^* \to \Sigma^* \times \Sigma^*$ *be the homomorphism defined by* $\mathrm{proj}(a) = \mathrm{proj}(a^{(2)}) = (\varepsilon, a)$ *and* $\mathrm{proj}(a^{(1)}) = (a, \varepsilon)$ *for all $a \in \Sigma$. Finally, let* Inc *be the set of words $w \in \Gamma^*$ of the form*

$$w = v_1 a_1 a_1^{(1)}\, v_2 a_2 a_2^{(1)}\, \cdots v_\ell a_\ell a_\ell^{(1)}\, v_{\ell+1} a_{\ell+1}^{(1)}\, b_1^{(1)} c_1^{(2)}\, b_2^{(1)} c_2^{(2)}\, \cdots\, b_n^{(1)} c_n^{(2)}$$

for some $\ell, n \geq 0$, $v_1, v_2, \ldots, v_{\ell+1} \in \Sigma^*$, $a_1, \ldots, a_{\ell+1} \in \Sigma$ such that a_i does not appear in v_i for any $i \in [\ell+1]$, $v_{\ell+1} \neq \varepsilon$, and $b_i, c_i \in \Sigma$ for all $i \in [1,n]$.

(i) For any $u, v \in \Sigma^*$, we have $u \parallel v$ and $|u| \leq |v|$ if, and only if, there is some word $w \in \mathrm{Inc}$ with $\mathrm{proj}(w) = (u,v)$.
(ii) If $w_1, w_2 \in \mathrm{Inc}$ satisfy $\mathrm{proj}(w_1) = \mathrm{proj}(w_2)$, then $w_1 = w_2$.
(iii) If $w \in \mathrm{Inc}$ satisfies $\mathrm{proj}(w) = (u,v)$, then $|v| - |u| = \sum_{a \in \Sigma} |w|_a - 1$.

Remark 4.7. Consider statement (i) and let $u = a$ and $v = bb$. Then the only word $w \in \mathrm{Inc}$ with $\mathrm{proj}(w) = (u,v)$ is $w = bba^{(1)}$. Thus, $\ell = n = 0$, $v_{\ell+1} = bb$, and $a_{\ell+1}^{(1)} = a^{(1)}$. This example shows that $|v_{\ell+1}| > 1$ is possible. In this aspect, statement (i) corrects [12, Lemma 5.2] where only letters $v_{\ell+1} \in \Sigma$ are considered.

Proof. (i) To prove the implication "$\Leftarrow$", let $w \in \mathrm{Inc}$. Then there are $\ell, n \geq 0$, $v_i \in \Sigma^*$ and $a_i \in \Sigma$ for $i \in [1, \ell+1]$, and $b_i, c_i \in \Sigma$ for $i \in [1,n]$ such that the conditions expressed in the lemma hold. With $\mathrm{proj}(w) = (u,v)$, we get

$$\begin{aligned} u &= a_1 a_2 \cdots a_\ell a_{\ell+1} b_1 b_2 \cdots b_n \text{ and} \\ v &= v_1 a_1\, v_2 a_2 \cdots v_\ell a_\ell\, v_{\ell+1}\, c_1 c_2 \cdots c_n\,. \end{aligned}$$

Note that $|u| = \ell+1+n$ and $|v| \geq \ell+|v_{\ell+1}|+n$. Since $v_{\ell+1} \neq \varepsilon$, this ensures $|u| \leq |v|$. It follows immediately that $v \not\sqsubseteq u$ holds. Towards a contradiction, suppose $u \sqsubseteq v$. Since v_1 does not contain a_1, the first letter of u, we obtain

$$a_2 \cdots a_\ell a_{\ell+1} b_1 \cdots b_n \sqsubseteq v_2 a_2 \cdots v_\ell a_\ell v_{\ell+1} c_1 \cdots c_n\,.$$

Proceeding recursively, this yields

$$a_{\ell+1}b_1\cdots b_n \sqsubseteq v_{\ell+1}c_1\cdots c_n\,.$$

Since $a_{\ell+1}$ does not appear in $v_{\ell+1}$, this implies

$$a_{\ell+1}b_1\cdots b_n \sqsubseteq c_1\cdots c_n\,,$$

i.e., the word on the left of length $1+n$ is a subword of the word on the right of length n, a contradiction. Hence, indeed, $u \not\sqsubseteq v$ which completes the proof of the implication "$\Leftarrow$".

Conversely, let $|u| \le |v|$ and $u \not\sqsubseteq v$. Let $u = a_1a_2\cdots a_m$. Suppose $\ell \ge 0$ such that

$$v = v_1a_1\,v_2a_2\,\cdots v_\ell a_\ell\,v''$$

for words $v_1,\dots,v_\ell,v'' \in \Sigma^*$ such that v_i does not contain a_i for any $i \in [1,\ell]$ and $|v''| \ge m-\ell$. Note that this holds for $\ell = 0$ (set $v'' = v$), but not for $\ell = m$ (since then $u \sqsubseteq v$). So let ℓ be maximal such that a factorization of v as above exists. Since $|v''| \ge m-\ell$, we can factorize $v'' = v_{\ell+1}\,v'$ with $v_{\ell+1} \ne \varepsilon$ and $|v'| = m-(\ell+1)$. Note that $a_{\ell+1}$ does not appear in $v_{\ell+1}$ since ℓ was chosen maximal. With $v' = c_1c_2\cdots c_{m-(\ell+1)}$ we set

$$w = v_1a_1a_1^{(1)}\,\cdots\,v_\ell a_\ell a_\ell^{(1)}\,v_{\ell+1}a_{\ell+1}^{(1)}\,a_{\ell+2}^{(1)}c_1^{(2)}\,a_{\ell+3}^{(1)}c_2^{(2)}\cdots a_m^{(1)}c_{m-(\ell+1)}^{(2)}\,.$$

Then $w \in \mathrm{Inc}$ and

$$\begin{aligned}\mathrm{proj}(w) &= (a_1a_2\cdots a_{\ell+1}a_{\ell+2}\cdots a_m, v_1a_1v_2a_2\cdots v_\ell a_\ell v_{\ell+1}c_1c_2\cdots c_{m-\ell-1})\\ &= (u,v)\,.\end{aligned}$$

This finishes the verification of the first claim.

(ii) Suppose $w \in \mathrm{Inc}$ with $\mathrm{proj}(w) = (a_1a_2\cdots a_m, v)$. We reconstruct the word w from this information thereby verifying claim (ii). By claim (i), $m \le |v|$ and $a_1\cdots a_m \not\sqsubseteq v$ (implying $m \ge 1$). From $w \in \mathrm{Inc}$ and $\mathrm{proj}(w) = (a_1\cdots a_m, v)$, we obtain a word $v_1 \in \Sigma^*$ not containing a_1 such that
(a) $v_1a_1a_1^{(1)}$ is a prefix of w or
(b) $v_1a_1^{(1)}$ is a prefix of w and $v_1 \ne \varepsilon$.
In case (a), there are $w' \in \mathrm{Inc}$ with $w = v_1a_1a_1^{(1)}w'$, $v' \in \Sigma^*$ with $v = v_1a_1v'$, and $\mathrm{proj}(w') = (a_2\cdots a_m, v')$. Proceeding recursively, we end up with case (b).
So assume $v_1a_1^{(1)}$ is a prefix of w and $v_1 \ne \varepsilon$. From $w \in \mathrm{Inc}$, this implies

$$\begin{aligned} w &= v_1a_1^{(1)}c_1^{(1)}d_1^{(2)}\cdots c_n^{(1)}d_n^{(2)}\,,\\ a_1\cdots a_m &= a_1c_1c_2\cdots c_n \ \text{ and}\\ v &= v_1d_1d_2\cdots d_n\,.\end{aligned}$$

Thus, the letters c_i and d_i are completely determined by the words $a_1\cdots a_m$, v_1, and v. Hence, also w is determined by these words.

(iii) Let $w \in \mathrm{Inc}$. Then there are $\ell, n \geq 0$, words v_i, and letters a_i for $i \in [1, \ell]$ as well as letters c_i, d_i for $i \in [1, n]$ such that

$$w = v_1 a_1 a_1^{(1)}\, v_2 a_2 a_2^{(1)}\, \cdots v_\ell a_\ell a_\ell^{(1)}\, v_{\ell+1} a_{\ell+1}^{(1)}\, b_1^{(1)} c_1^{(2)}\, b_2^{(1)} c_2^{(2)}\, \cdots\, b_n^{(1)} c_n^{(2)}\,.$$

With $\mathrm{proj}(w) = (u, v)$, we get

$$u = a_1 a_2 \cdots a_\ell\, a_{\ell+1}\, b_1 b_2 \cdots b_n\,, \text{ and}$$
$$v = v_1 a_1\, v_2 a_2\, \cdots v_\ell a_\ell\, v_{\ell+1} c_1 c_2 \cdots c_n$$

and therefore $|u| = \ell + 1 + n$ and $|v| = |v_1 v_2 \cdots v_{\ell+1}| + \ell + n$. Consequently, $|v| - |u| = |v_1 v_2 \cdots v_{\ell+1}| - 1 = \sum_{a \in \Sigma} |w|_a - 1$. □

Next, we derive from the above lemma that, indeed, the relations $\| \cap \mathrm{ld}_{\geq \ell}$ and $\| \cap \mathrm{ld}_{=\ell}$ are effectively unambiguous rational.

Proposition 4.8. *For any alphabet Σ and any $\ell \in \mathbb{N}$, the relations $\| \cap \mathrm{ld}_{\geq \ell}$ and $\| \cap \mathrm{ld}_{=\ell}$ are effectively unambiguous rational.*

Proof. Let Σ_1, Σ_2, Γ, Inc, and proj be defined as in Lemma 4.6. Furthermore, let $\mathrm{LD}_{\geq \ell} \subseteq \Gamma^*$ denote the set of all words w with $\sum_{a \in \Sigma} |w|_a \geq \ell$. The languages

$$\mathrm{Inc} = \left(\bigcup_{a \in \Sigma} \left(\left(\Sigma \setminus \{a\}\right)^* a a^{(1)} \right) \right)^* \left(\bigcup_{a \in \Sigma} \left(\left(\Sigma \setminus \{a\}\right)^+ a^{(1)} \right) \right) (\Sigma_1 \Sigma_2)^*$$

and $\mathrm{LD}_{\geq \ell}$ are effectively regular. Hence, the same applies to their intersection. By Lemma 4.6, the homomorphism proj maps $\mathrm{Inc} \cap \mathrm{LD}_{\geq \ell}$ bijectively onto $\| \cap \mathrm{ld}_{\geq \ell}$. Hence this relation is effectively unambiguous rational.

Similarly, the language $\mathrm{Inc} \cap \mathrm{LD}_{\geq \ell} \setminus \mathrm{LD}_{\geq \ell+1}$ is effectively regular and proj maps it bijectively onto $\| \cap \mathrm{ld}_{=\ell}$. □

4.2 Closure properties of the class of regular languages

Let $R \subseteq \Sigma^* \times \Sigma^*$ be a rational relation and $L \subseteq \Sigma^*$ a regular language. Then the language

$$K = \{u \mid \exists v \in L \colon (u, v) \in R\}$$

is effectively regular. Note that, for $k = 1$, it equals

$$\{u \mid \exists^{\geq k} v \in L \colon (u, v) \in R\}\,.$$

For arbitrary numbers $k \geq 2$, this language need not be regular as the following example shows.

Example 4.9. Consider the rational relation $R = \{(a^m b a^n, a^m), (a^m b a^n, a^n) \mid m, n \in \mathbb{N}\}$, $L = \Sigma^*$, and $k = 2$. Then the word u satisfies $\exists^{\geq 2} v \in \Sigma^* \colon (u, v) \in R$ iff it is of the form $u = a^m b a^m$ for some $m \in \mathbb{N}$. But the set of these words u is not regular. Similarly, u satisfies $\exists^{0 \bmod 2} v \in \Sigma^* \colon (u, v) \in R$ iff $u = a^m b a^m$ for some $m \in \mathbb{N}$ or u does not belong to $a^* b a^*$. Again, the set of these words u is not regular.

In this section, we will show that unambiguous rational relations preserve the regularity of a language under "counting images". More precisely, let $R \subseteq \Sigma^* \times \Sigma^*$ be an unambiguous rational relation and $L \subseteq \Sigma^*$ a regular language. We want to show that the languages of all words $u \in \Sigma^*$ with

$$|\{v \in L \mid (u,v) \in R\}| \geq k \tag{1}$$
$$\text{and } |\{v \in L \mid (u,v) \in R\}| \in p + q\mathbb{N}, \text{ respectively} \tag{2}$$

are effectively regular for all $k \in \mathbb{N}$ and all $0 \leq p < q$, respectively.

For these proofs, we need the following classical concepts from the theory of weighted automata (see [21, 2, 3] for recent surveys). Let S be a semiring. A function $r\colon \Sigma^* \to S$ is *recognizable over* S, if there are $n \in \mathbb{N}$, $\lambda \in S^{1\times n}$, a homomorphism $\mu\colon \Sigma^* \to S^{n\times n}$, and $\nu \in S^{n\times 1}$ with $r(w) = \lambda \cdot \mu(w) \cdot \nu$ for all $w \in \Sigma^*$. The triple (λ, μ, ν) is a *presentation* or a *weighted automaton for* r.

In the following, we consider the semiring $\mathbb{N}^\infty$, i.e., the set $\mathbb{N} \cup \{\infty\}$ together with the commutative operations $+$ and $\cdot$ (with $x + \infty = \infty$ for all $x \in \mathbb{N} \cup \{\infty\}$, $x \cdot \infty = \infty$ for all $x \in (\mathbb{N} \cup \{\infty\}) \setminus \{0\}$, and $0 \cdot \infty = 0$). In this semiring, we define an infinite addition in the natural way: $\sum_{i\in I} n_i = \infty$ if there are infinitely many $i \in I$ with $n_i > 0$; $\sum_{i\in I} n_i = \sum_{i\in I, n_i>0} n_i$ otherwise.

We next prove the following preservation property of the class of recognizable functions over the semiring $\mathbb{N}^\infty$. Suppose Γ and Σ are alphabets and $f\colon \Gamma^* \to \Sigma^*$ is a mapping. Let, furthermore, $\chi\colon \Gamma^* \to \mathbb{N}^\infty$ be a function. Then we define a new function $r\colon \Gamma^* \to \mathbb{N}^\infty$ setting

$$\chi \circ f^{-1}(u) = \sum_{\substack{w\in\Gamma^* \\ f(w)=u}} \chi(w)\,.$$

Since addition is commutative, the order of the summands does not matter, but there can be distinct words w_1, w_2 with $f(w_1) = f(w_2) = u$ and $k = \chi(w_1) = \chi(w_2)$. In that case, the summand k appears twice in this sum. Note that, in general, this sum is infinite. So the definition of $\chi \circ f^{-1}$ makes sense over the semiring $\mathbb{N}^\infty$, but not over arbitrary semirings where infinite sums might not be defined (e.g., $\mathbb{Z}_2$)

It is our aim to show that $\chi \circ f^{-1}$ is recognizable assuming χ to be recognizable and f to be a homomorphism (cf. Prop. 4.12 below). Proofs of the plain recognizability of this functions can be found in the literature, but those books do not consider the question of effectiveness. Therefore, we decided to present a construction here. This proof considers first two special cases of homomorphisms before it combines them for the general result.

A homomorphism $f\colon \Gamma^* \to \Sigma^*$ is *non-erasing* if $f(a) \neq \varepsilon$ for all $a \in \Gamma$ (or, equivalently, $|f(w)| \geq |w|$ for all $w \in \Gamma^*$). A proof of the non-effective version of the following lemma can be found in [21, Thm. IV.3.1(1)].

In our proof, we refer to the following notions. A *monomial* is a function $s\colon \Gamma^* \to \mathbb{N}^\infty$ such that $s(w) \neq 0$ for at most one word $w \in \Gamma^*$. Now let $r, s\colon \Gamma^* \to \mathbb{N}^\infty$. Their *sum* is the function $r + s\colon \Gamma^* \to \mathbb{N}^\infty$ with $(r + s)(w) = r(w) + s(w)$

for all $w \in \Gamma^*$. The *Cauchy-product* $r \odot s\colon \Gamma^* \to \mathbb{N}^\infty$ is defined by

$$(r \odot s)(w) = \sum_{w=uv} r(u) \cdot s(v)\,.$$

Here, we sum over finitely many elements of $\mathbb{N}^\infty$ since the word w can only be factorized in $|w| + 1$ many ways into two words u and v. Note that the sum is defined pointwise while the Cauchy-product is inspired by the product of multivariate polynomials with non-commuting variables. Finally, the *iteration* r^* is defined by

$$r^\infty(w) = \sum_{w=w_1 w_2 \cdots w_n} \prod_{1 \le i \le n} r(w_i)\,.$$

Note that the sum above is infinite since empty factors $w_i = \varepsilon$ are allowed.

Lemma 4.10 (cf. [21, Thm. IV.3.1(1)]). *Let Γ and Σ be alphabets, $f\colon \Gamma^* \to \Sigma^*$ a non-erasing homomorphism, and $\chi\colon \Gamma^* \to \mathbb{N}^\infty$ a recognizable function over $\mathbb{N}^\infty$. Then the function*

$$r = \chi \circ f^{-1}\colon \Sigma^* \to \mathbb{N}^\infty\colon u \mapsto \sum_{\substack{w \in \Gamma^* \\ f(w)=u}} \chi(w)$$

is effectively recognizable over $\mathbb{N}^\infty$.

Proof. Since χ is recognizable, it can be constructed from monomials $s\colon \Gamma^* \to \mathbb{N}^\infty$ using addition, Cauchy-product, and iteration applied to functions t with $t(\varepsilon) = 0$ (i.e., χ is *rational*) [22, Theorem 3.11].

For a monomial $s\colon \Gamma^* \to \mathbb{N}^\infty$ with $s(v) = 0$ for all $v \neq w$, set

$$s'(x) = \begin{cases} s(w) & \text{if } f(x) = w \\ 0 & \text{otherwise} \end{cases}$$

Since f is non-erasing, there are only finitely many words x with $s'(x) \neq 0$. Hence s' is the sum of finitely many monomials.

Now, in the construction of χ from above, replace any monomial s by the sum of monomials representing s'. This, way, one obtains a construction of r. Since f is non-erasing, also in this construction, iteration is only applied to functions t with $t(\varepsilon) = 0$. Hence, by [22, Theorem 3.1] again, r is recognizable. Analysing the proof of that theorem, one even obtains that a presentation for r can be computed from f and a presentation of χ. □

Another special case of the desired result concerns non-expanding homomorphisms: a homomorphism $f\colon \Gamma^* \to \Sigma^*$ is *non-expanding* if $f(a) \in \Sigma \cup \{\varepsilon\}$ for all $a \in \Gamma$ (or, equivalently, $|f(w)| \le |w|$ for all $w \in \Gamma^*$).

The following lemma can be found in [23, Thm. II.4.7], but the proof there is non-effective, again.

Lemma 4.11 (cf. [23, Thm. II.4.7]). *Let Γ and Σ be alphabets, $f\colon \Gamma^* \to \Sigma^*$ a non-expanding homomorphism, and $\chi\colon \Gamma^* \to \mathbb{N}^\infty$ a recognizable function over $\mathbb{N}^\infty$. Then the function*

$$r = \chi \circ f^{-1}\colon \Sigma^* \to \mathbb{N}^\infty : u \mapsto \sum_{\substack{w \in \Gamma^* \\ f(w)=u}} \chi(w)$$

is effectively recognizable over $\mathbb{N}^\infty$.

Proof. Let (λ, μ, ν) be a presentation of dimension n for χ.

For $\sigma \in \Sigma \cup \{\varepsilon\}$, let

$$\Gamma_\sigma = \{b \in \Gamma \mid f(b) = \sigma\}\,.$$

Since f is non-expanding, Γ is the disjoint union of these subalphabets. Furthermore, let $M \in (\mathbb{N}^\infty)^{n\times n}$ be the matrix defined by

$$M = \sum_{w \in \Gamma_\varepsilon^*} \mu(w)\,. \tag{3}$$

To define a presentation for the function r, we first define a homomorphism $\mu'\colon \Sigma^* \to \left(\mathbb{N}^\infty\right)^{n\times n}$ by

$$\mu'(a) = \sum_{b \in \Gamma_a} \big(\mu(b) \cdot M\big)$$

for all $a \in \Sigma$. Setting

$$\lambda' = \lambda \cdot M \text{ and } \nu' = \nu$$

defines the presentation (λ', μ', ν') of dimension n. Now let $u = a_1 a_2 \ldots a_m \in \Sigma^*$ with $a_i \in \Sigma$ for all $1 \le i \le m$. Then we get

$$\begin{aligned}
\lambda' \cdot \mu'(u) \cdot \nu' &= \lambda \cdot M \cdot \left(\prod_{1 \le i \le m} \sum_{b_i \in \Gamma_{a_i}} \big(\mu(b_i) \cdot M\big) \right) \cdot \nu \\
&= \lambda \cdot \sum \big(\mu(w_0) \mid w_0 \in \Gamma_\varepsilon^*\big) \cdot \left(\prod_{1 \le i \le m} \sum \big(\mu(w_i) \mid w_i \in \Gamma_{a_i} \Gamma_\varepsilon^*\big) \right) \cdot \nu \\
&= \lambda \cdot \sum \big(\mu(w) \mid w \in \Gamma_\varepsilon^* \Gamma_{a_1} \Gamma_\varepsilon^* \Gamma_{a_2} \cdots \Gamma_\varepsilon^* \Gamma_{a_m} \Gamma_\varepsilon^*\big) \cdot \nu \\
&= \lambda \cdot \sum \big(\mu(w) \mid w \in \Gamma^*, f(w) = u\big) \cdot \nu \\
&= \sum_{\substack{w \in \Gamma^* \\ f(w)=u}} \lambda \cdot \mu(w) \cdot \nu \\
&= r(u)\,.
\end{aligned}$$

Hence, (λ', μ', ν') is a presentation for the function r, i.e., r is recognizable.

It remains to be shown that the presentation (λ', μ', ν') is computable from the presentation (λ, μ, ν) and the homomorphism f. For this, it suffices to construct the matrix M effectively, i.e., to compute the infinite sum in Eq. (3). Using a pumping argument, one first shows the equivalence of the following two statements for all $i, j \in \{1, 2, \ldots, n\}$:

(a) There are infinitely many words $w \in \Gamma_\varepsilon^*$ with $\mu(w)_{ij} > 0$.
(b) There exists $w \in \Gamma_\varepsilon^*$ with $n < |w|$ and $\mu(u)_{ij} > 0$.
(c) There is a word $w \in \Gamma_\varepsilon^*$ with $n < |w| \leq 2n$ and $\mu(w)_{ij} > 0$.

Since statement (c) is decidable, we can evaluate Eq. (3) calculating

$$M_{ij} = \begin{cases} \infty & \text{if (c) holds} \\ \sum\big(\mu(w)_{ij} \mid w \in \Gamma_\varepsilon^*, |w| \leq n\} & \text{otherwise.} \end{cases}$$

□

We now come to the announced preservation result for arbitrary homomorphisms $f\colon \Gamma^* \to \Sigma^*$.

Proposition 4.12. *Let Γ and Σ be alphabets, $f\colon \Gamma^* \to \Sigma^*$ a homomorphism, and $\chi\colon \Gamma^* \to \mathbb{N}^\infty$ a recognizable function over $\mathbb{N}^\infty$. Then the following function r is effectively recognizable over $\mathbb{N}^\infty$:*

$$r = \chi \circ f^{-1}\colon \Sigma^* \to \mathbb{N}^\infty\colon u \mapsto \sum_{\substack{w \in \Gamma^* \\ f(w)=u}} \chi(w)$$

Proof. Let $\sigma \in \Sigma$ be arbitrary. Then define homomorphisms $f_1\colon \Gamma^* \to \Gamma^*$ and $f_2\colon \Gamma^* \to \Sigma^*$ by

$$f_1(a) = \begin{cases} \varepsilon & \text{if } f(a) = \varepsilon \\ a & \text{otherwise} \end{cases} \quad \text{and } f_2(a) = \begin{cases} f(a) & \text{if } f(a) \neq \varepsilon \\ \sigma & \text{otherwise} \end{cases}$$

for all letters $a \in \Gamma$. Then f_1 is non-expanding, f_2 is non-erasing, and $f = f_2 \circ f_1$.

By Lemma 4.11, the function $\chi \circ f_1^{-1}$ is effectively recognizable. By Lemma 4.10, also $\chi \circ f_1^{-1} \circ f_2^{-1} = \chi \circ f^{-1}$ is effectively recognizable. □

Let $R \subseteq \Sigma^* \times \Sigma^*$ be a relation and $L \subseteq \Sigma^*$ a language. Then R and L give rise to a counting function $\Sigma^* \to \mathbb{N}^\infty$ that maps $u \in \Sigma^*$ to *the number of words* $v \in L$ with $(u, v) \in R$. The following lemma demonstrates that this counting function is recognizable provided R is unambiguous rational and L is regular.

Lemma 4.13. *Let $R \subseteq \Sigma^* \times \Sigma^*$ be an unambiguous rational relation and $L \subseteq \Sigma^*$ be regular. Then the following function r is effectively recognizable over $\mathbb{N}^\infty$:*

$$r\colon \Sigma^* \to \mathbb{N}^\infty\colon u \mapsto |\{v \in L \mid (u, v) \in R\}|$$

Proof. Since R is unambiguous rational, there are an alphabet Γ, homomorphisms $f, g\colon \Gamma^* \to \Sigma^*$, and a regular language $S \subseteq \Gamma^*$ such that

$$(f,g)\colon \Gamma^* \to \Sigma^* \times \Sigma^* \colon w \mapsto \big(f(w), g(w)\big)$$

maps S bijectively onto the relation R. Let $S_L = S \cap g^{-1}(L)$. Since L is regular, the language S_L is effectively regular. Furthermore, (f,g) maps S_L bijectively onto $R \cap (\Sigma^* \times L)$.

Since S_L is regular, the characteristic function

$$\chi\colon \Gamma^* \to \mathbb{N}^\infty \colon w \mapsto \begin{cases} 1 & \text{if } w \in S_L \\ 0 & \text{otherwise} \end{cases}$$

is effectively recognizable over $\mathbb{N}^\infty$ [22, Proposition 3.12].

By Proposition 4.12, also the function

$$r'\colon \Sigma^* \to \mathbb{N}^\infty \colon u \mapsto \sum_{\substack{w \in \Gamma^* \\ f(w)=u}} \chi(w)$$

is effectively recognizable. Note that, for $u \in \Sigma^*$, we get

$$\begin{aligned}
r'(u) &= \sum_{\substack{w \in \Gamma^* \\ f(w)=u}} \chi(w) \\
&= |\{w \in S_L \mid f(w) = u\}| && \text{since } \chi \text{ is the characteristic function of } S_L \\
&= |\{g(w) \mid w \in S_L, f(w) = u\}| && \text{since } (f,g) \text{ is injective on } S_L \subseteq S \\
&= |\{v \mid (u,v) \in R \cap (\Sigma^* \times L)\}| && \text{since } (f,g) \text{ maps } S_L \text{ onto } R \cap (\Sigma^* \times L) \\
&= |\{v \in L \mid (u,v) \in R\}|\,.
\end{aligned}$$

In other words, the effectively recognizable function r' is the function r from the statement of the lemma. □

We now come to the main result of this section.

Proposition 4.14. *Let $R \subseteq \Sigma^* \times \Sigma^*$ be an unambiguous rational relation and $L \subseteq \Sigma^*$ be regular. Then, for $k \in \mathbb{N}$ and for $p, q \in \mathbb{N}$ with $p < q$, the set H of words w satisfying* (1) *and* (2)*, respectively, is effectively regular*

Proof. By Lemma 4.13, the function $r\colon \Sigma^* \to \mathbb{N}^\infty \colon u \mapsto |\{v \in L \mid (u,v) \in R\}|$ is effectively recognizable over $\mathbb{N}^\infty$.

We construct a semiring $S_k^\infty = (\{0, 1, \ldots, k, \infty\}, \oplus, \odot, 0, 1)$ setting

$$x \oplus y = \begin{cases} x + y & \text{if } x + y \in S_k^\infty \\ k & \text{otherwise} \end{cases} \quad \text{and } x \odot y = \begin{cases} x \cdot y & \text{if } x \cdot y \in S_k^\infty \\ k & \text{otherwise} \end{cases}$$

for $x, y \in S_k^\infty$ (note that $x + y \notin S_k^\infty$ is equivalent to $k < x + y < \infty$).

Then the mapping

$$h\colon \mathbb{N}^\infty \to S_k^\infty : n \mapsto \begin{cases} \min\{k,n\} & \text{if } n \in \mathbb{N} \\ \infty & \text{if } n = \infty \end{cases}$$

is a semiring homomorphism. It follows that the function

$$h \circ r\colon \Sigma^* \to S_k^\infty : u \mapsto \min\{k, |\{u \in L \mid (u,v) \in R\}|\}$$

is effectively recognizable over S_k^∞ [22, Prop. 4.5]. Since the semiring S_k^∞ is finite, the language

$$(h \circ r)^{-1}(k) = \{u \in \Sigma^* \mid r(u) \geq k\}$$

is effectively regular [22, Prop. 4.5]. Since $r(u) \geq k$ iff $|\{v \in L \mid (u,v) \in R\}| \geq k$, this proves the first statement.

The second statement is shown similarly. We construct the semiring $\mathbb{Z}_q = (\{0,1,\dots,q,\infty\}, \oplus, \odot, 0, 1)$ with

$$x \oplus y = \begin{cases} x+y & \text{if } x+y \in \mathbb{Z}_q^\infty \\ x+y \bmod q & \text{otherwise} \end{cases} \quad \text{and } x \odot y = \begin{cases} x \cdot y & \text{if } x \cdot y \in \mathbb{Z}_q^\infty \\ x \cdot y \bmod q & \text{otherwise} \end{cases}$$

for $x, y \in S_k^\infty$. Beware that, with $q = 6$, we get $4 \oplus 2 = 6 \neq 0 = (4+2) \bmod q$ and similarly $3 \odot 2 = 6 \neq 0 = (3 \cdot 2) \bmod q$.

Let η denote the semiring homomorphism from $\mathbb{N}^\infty$ to $\mathbb{Z}_q^\infty$ with

$$\eta(n) = \begin{cases} \infty & \text{if } n = \infty \\ q & \text{if } n \in q\mathbb{N} \setminus \{0\} \\ n \bmod q & \text{otherwise.} \end{cases}$$

It follows that the function

$$\eta \circ r\colon \Sigma^* \to \mathbb{Z}_q^\infty : u \mapsto \eta(|\{v \in L \mid (u,v) \in R\}|)$$

is effectively recognizable over $\mathbb{Z}_q^\infty$ [22, Prop. 4.5]. Since the semiring $\mathbb{Z}_q^\infty$ is finite, the language

$$(\eta \circ r)^{-1}(p)$$

is effectively regular for all $p \in \mathbb{Z}_q^\infty$ [22, Prop. 6.3]. Then the claim follows since

$$H = \begin{cases} (\eta \circ r)^{-1}(p) & \text{if } 1 \leq p < q \\ (\eta \circ r)^{-1}(0) \cup (\eta \circ r)^{-1}(q) & \text{if } p = 0\,. \end{cases}$$

$\square$

4.3 Quantifier elimination for C+MOD^2

Our decision algorithm employs a quantifier elimination procedure, i.e., we will transform an arbitrary formula into an equivalent one that is quantifer-free. As usual, the heart of this procedure handles formulas $\psi = \mathrm{Q}y\,\varphi$ where Q is a quantifier and φ is quantifier-free. Since the logic C+MOD^2 has only two variables, any such formula ψ has at most one free variable. In other words, it defines a language K. The following lemma shows that this language is effectively regular, such that ψ is equivalent to the quantifier-free formula $x \in K$.

Lemma 4.15. *Let $\varphi(x, y)$ be a quantifier-free formula from* C+MOD^2*. Then the sets*

$$\{x \in \Sigma^* \mid \mathcal{S} \models \exists^{\geq k} y\,\varphi\} \text{ and } \{x \in \Sigma^* \mid \mathcal{S} \models \exists^{p \bmod q} y\,\varphi\}$$

are effectively regular for all $k \in \mathbb{N}$ and all $p, q \in \mathbb{N}$ with $p < q$.

Proof. Without changing the meaning of the formula φ, we can do the following replacements of atomic formulas:

- $x = y$ can be replaced by $x \sqsubseteq y \wedge y \sqsubseteq x$,
- $x \sqsubseteq x$ by $x \in \Sigma^*$ and $y \sqsubseteq y$ by $y \in \Sigma^*$,
- $\mathrm{ld}_{\geq 0}(x, x)$ by $x \in \Sigma^*$ and $\mathrm{ld}_{\geq \ell}(x, x)$ by $x \in \emptyset$ for $\ell \geq 1$, and
- $\mathrm{ld}_{\geq 0}(y, y)$ by $y \in \Sigma^*$ and $\mathrm{ld}_{\geq \ell}(y, y)$ by $y \in \emptyset$ for $\ell \geq 1$.

Since φ is quantifier-free, we can therefore assume that it is a Boolean combination of formulas of the form

- $x \in K$ and $y \in L$ for some regular languages K and L,
- $x \sqsubseteq y$ and $y \sqsubseteq x$, and
- $\mathrm{ld}_{\geq \ell}(x, y)$ and $\mathrm{ld}_{\geq \ell}(y, x)$ with $0 \leq \ell < \Lambda$

for some $\Lambda > 0$.

Let Θ denote the set of the following formulas:

- $x = y$
- $x \sqsubset y \wedge \mathrm{ld}_{=\ell}(x, y)$ for $1 \leq \ell < \Lambda$
  $y \sqsubset x \wedge \mathrm{ld}_{=\ell}(y, x)$ for $1 \leq \ell < \Lambda$
- $x \sqsubset y \wedge \mathrm{ld}_{\geq \Lambda}(x, y)$
  $y \sqsubset x \wedge \mathrm{ld}_{\geq \Lambda}(y, x)$
- $x \parallel y \wedge \mathrm{ld}_{=\ell}(x, y)$ for $0 \leq \ell < \Lambda$
  $y \parallel x \wedge \mathrm{ld}_{=\ell}(y, x)$ for $1 \leq \ell < \Lambda$
- $x \parallel y \wedge \mathrm{ld}_{\geq \Lambda}(x, y)$
  $y \parallel x \wedge \mathrm{ld}_{\geq \Lambda}(y, x)$

Note that any pair of words x and y satisfies precisely one of the formulas from Θ. Hence φ is equivalent to

$$\bigvee_{\theta \in \Theta} (\theta \wedge \varphi)\,.$$

In this formula, any occurrence of φ appears in conjunction with precisely one of the formulas $\theta \in \Theta$. Depending on this formula θ, we can simplify φ to φ_θ by replacing the atomic subformulas that relate x and y as follows:

– If $\theta = (x = y)$, then replace $x \sqsubseteq y$, $y \sqsubseteq x$, $\mathrm{ld}_{\geq 0}(x, y)$, and $\mathrm{ld}_{\geq 0}(y, x)$ by the valid formula $\top = (x \in \Sigma^*)$, and replace $\mathrm{ld}_{\geq m}(x, y)$ and $\mathrm{ld}_{\geq m}(y, x)$ for $m \geq 1$ by the contradictory formula $\perp = (x \in \emptyset)$.
– If $x \sqsubset y$ appears in θ, then replace $x \sqsubseteq y$ by $\top$, and replace $y \sqsubseteq x$ by $\perp$ (and symmetrically for $y \sqsubset x$).
– If $\mathrm{ld}_{=0}(x, y)$ or $\mathrm{ld}_{=0}(y, x)$ appears in θ, then replace $\mathrm{ld}_{\geq 0}(x, y)$ and $\mathrm{ld}_{\geq 0}(y, x)$ by $\top$, and replace $\mathrm{ld}_{\geq m}(x, y)$ and $\mathrm{ld}_{\geq m}(y, x)$ for $m \geq 1$ by $\perp$.
– If $\mathrm{ld}_{=\ell}(x, y)$ with $\ell > 0$ appears in θ, then replace $\mathrm{ld}_{\geq m}(x, y)$ for $m \leq \ell$ by $\top$ and replace $\mathrm{ld}_{\geq m}(x, y)$ for $m > \ell$ and $\mathrm{ld}_{\geq m}(y, x)$ for $m \geq 0$ by $\perp$ (and symmetrically for $\mathrm{ld}_{=\ell}(y, x)$).
– If $\mathrm{ld}_{\geq \Lambda}(x, y)$ appears in θ, then replace $\mathrm{ld}_{\geq m}(x, y)$ for $m \geq 0$ by $\top$ and $\mathrm{ld}_{\geq m}(y, x)$ for $m \geq 0$ by $\perp$ (and symmetrically for $\mathrm{ld}_{\geq \Lambda}(y, x)$).

Recall that φ does not contain any formula $\mathrm{ld}_{\geq \ell}(x, y)$ with $\ell \geq \Lambda$. Consequently, for any $\theta \in \Theta$, the formulas $\theta \wedge \varphi$ and $\theta \wedge \varphi_\theta$ are equivalent. Furthermore, φ_θ is a Boolean combination of formulas $x \in K$ and $y \in L$ for some regular languages K and L.

As a result, the formula φ is equivalent to

$$\bigvee_{\theta \in \Theta} (\theta \wedge \varphi_\theta)\,.$$

where the formulas φ_θ are Boolean combinations of formulas of the form $x \in K$ and $y \in L$ for some regular languages K and L.

Now let $k \in \mathbb{N}$. Since the formulas $\theta \in \Theta$ are mutually exclusive (i.e., $\theta_1(x, y) \wedge \theta_2(x, y)$ implies $\theta_1 = \theta_2$), we get

$$\exists^{\geq k} y\, \varphi \equiv \exists^{\geq k} y \bigvee_{\theta \in \Theta} (\theta \wedge \varphi_\theta) \equiv \bigvee_{(*)} \bigwedge_{\theta \in \Theta} \exists^{\geq k_\theta} y\, (\theta \wedge \varphi_\theta)$$

where the disjunction $(*)$ extends over all tuples $(k_\theta)_{\theta \in \Theta}$ of natural numbers $k_\theta \leq k$ with $\sum_{\theta \in \Theta} k_\theta = k$.

Hence it suffices to show that

$$\{x \in \Sigma^* \mid \exists^{\geq k} y\, (\theta \wedge \varphi)\} \tag{4}$$

is effectively regular for all $\theta \in \Theta$, all $k \in \mathbb{N}$, and all Boolean combinations φ of formulas of the form $x \in K$ and $y \in L$ where K and L are regular languages. Since the class of regular languages is closed under Boolean operations, we can find regular languages K_i and L_i such that

$$\varphi \equiv \bigvee_{i \in [1,n]} (x \in K_i \wedge y \in L_i)\,.$$

Note that this formula is equivalent to

$$\bigvee_{\substack{M,N \subseteq [1,n] \\ M \cap N \neq \emptyset}} \left(x \in \underbrace{\bigcap_{i \in M} K_i \setminus \bigcup_{i \notin M} K_i}_{=K_M} \wedge y \in \underbrace{\bigcap_{i \in N} L_i \setminus \bigcup_{i \notin N} L_i}_{=L_N} \right).$$

Since this disjunction is exclusive (i.e. any pair of words (x, y) satisfies at most one of the cases $M, N \subseteq [1, n]$ with $M \cap N \neq \emptyset$), the set from (4) equals the union of the sets

$$\{x \in \Sigma^* \mid \exists^{\geq k} y\,(\theta \wedge x \in K_M \wedge y \in L_N)\} \tag{5}$$

for $M, N \subseteq [1, n]$ not disjoint. Note that the set from (5) equals

$$X_{M,N,\theta} = K_M \cap \{x \in \Sigma^* \mid \exists^{\geq k} y \in L_N : \theta\}\,.$$

We next show that this set is effectively regular for any $\theta \in \Theta$ and any $M, N \subseteq [1, n]$ not disjoint:

- Suppose $\theta = (x = y)$. If $k = 1$, then we get $X_{M,N,\theta} = K_M \cap L_N$ and, if $k \neq 1$, then $X_{M,N,\theta}$ is empty. Hence, in any case, it is effectively regular.
- Suppose $\theta = (x \sqsubset y \wedge \mathrm{ld}_{=\ell}(x, y))$ with $1 \leq \ell < \Lambda$. Consider the relation $R_\theta = (\sqsubset \cap\, \mathrm{ld}_{=\ell})$. By Proposition 4.5, this relation is unambiguous rational. Hence, by Prop. 4.14, the language

  $$\big\{x\colon |\{y \in L_M \mid (x, y) \in R_\theta\}| \geq k\big\}$$

  is effectively regular. But this language equals the set of words

  $$\{x \mid \exists^{\geq k} y \in L_M : x \sqsubset y \wedge \mathrm{ld}_{=\ell}(x, y)\} = \{x \mid \exists^{\geq k} y \in L_M : \theta\}\,.$$

  Thus, $X_{M,N,\theta}$ is the intersection of two regular languages and therefore regular.
  For $\theta = (x \sqsubset y \wedge\, \mathrm{ld}_{\geq \Lambda}(x, y))$, we can argue analogously (just replace any occurrence of "$= \ell$" in the above paragraph by "$\geq \Lambda$").
- Suppose $\theta = (y \sqsubset x \wedge \mathrm{ld}_{=\ell}(y, x))$ with $1 \leq \ell < \Lambda$. Consider the unambiguous rational relation $R_h = (\sqsubset \cap\, \mathrm{ld}_{=\ell})$. It follows that also its inverse

  $$R_\theta = \{(x, y) \mid y \sqsubset x \wedge \mathrm{ld}_{\geq \ell}(y, x)\}$$

  is unambiguos rational. Hence, by Prop. 4.14, the language

  $$\big\{x\colon |\{y \in L_M \mid (x, y) \in R_\theta\}| \geq k\big\}$$

  is effectively regular. But this language equals the set of words

  $$\{x \mid \exists^{\geq k} y \in L_M : y \sqsubset x \wedge \mathrm{ld}_{=\ell}(y, x)\} = \{x \mid \exists^{\geq k} y \in L_M : \theta\}\,.$$

  Thus, $X_{M,N,\theta}$ is the intersection of two regular languages and therefore regular.
  For $\theta = (y \sqsubset x \wedge\, \mathrm{ld}_{\geq \Lambda}(y, x))$, we can argue analogously (just replace any occurrence of "$= \ell$" in the above paragraph by "$\geq \Lambda$").

The regularity of $X_{M,N,\theta}$ for the other formulas θ can be shown similarly (using Proposition 4.8 instead of Proposition 4.5.

Since the language from the claim of the lemma is a boolean combination of such languages $X_{M,N,\theta}$, the first claim is demonstrated.

To also demonstrate the regularity of the second language, let $p, q \in \mathbb{N}$ with $p < q$. Then $\exists^{p \bmod q} y\, \varphi$ is equivalent to

$$\bigvee_{(*)} \bigwedge_{\theta \in \Theta} \exists^{p_\theta \bmod q} y\, (\theta \wedge \varphi_\theta)$$

where the disjunction $(*)$ extends over all tuples $(p_\theta)_{\theta \in \Theta}$ of natural numbers $p_\theta \in \{0, 1, \ldots, q-1\}$ with $\sum_{\theta \in \Theta} p_\theta \equiv p \bmod q$. The rest of the proof proceeds *mutatis mutandis.* $\square$

Theorem 4.16. *Let $\mathcal{S} = (\Sigma^*, \sqsubseteq, (\mathrm{ld}_{\geq \ell})_{\ell \geq 0}, (L)_{L\ regular})$. Let $\varphi(x)$ be a formula from* C+MOD^2*. Then the set $\{x \in \Sigma^* \mid \mathcal{S} \models \varphi(x)\}$ is effectively regular.*

Proof. The only atomic formulas with a single variable x are $x \in L$ with L regular, $x = x$, $x \sqsubseteq x$, $\mathrm{ld}_{\geq 0}(x, x)$ (which are equivalent to $x \in \Sigma^*$), and $\mathrm{ld}_{\geq \ell}(x, x)$ with $\ell > 0$ (which is equivalent to $x \in \emptyset$). Therefore, the claim holds if φ is atomic. For more complicated formulas, the proof proceeds by induction using Lemma 4.15 and the effective closure of the class of regular languages under boolean operations. $\square$

Corollary 4.17. *Let $L \subseteq \Sigma^*$ be a regular language. Then the* C+MOD^2*-theory of the structure $\mathcal{S}_L = (L, \sqsubseteq, (\mathrm{ld}_{\geq \ell})_{\ell \geq 0}, (K \cap L)_{K\ regular}, (w)_{w \in L})$ is decidable.*

Proof. Let $\varphi \in$ C+MOD^2 be a sentence. We build φ_L by (1) restricting all quantifications to L, (2) replace $x\theta w$ by $\exists y \colon y \in \{w\} \wedge x\theta y$, and dually for $y\theta w$ for all $w \in L$ and all binary relations θ.

With $\mathcal{S}$ the structure from Theorem 4.16, we obtain $\mathcal{S} \models \varphi_L \iff \mathcal{S}_L \models \varphi$. By Theorem 4.16, the language $\{x \mid \mathcal{S} \models \varphi_L\}$ is regular (since φ_L is a sentence, it is $\emptyset$ of Σ^*). Hence φ_L holds iff this set is nonempty, which is decidable. $\square$

5 The Σ_1-theory

Recall that Σ_1 is the set of first-order formulas of the form $\exists x_1 \exists x_2 \ldots \exists x_m\, \psi$ where ψ is quantifier-free and $m \geq 0$. Furthermore, Σ_2 is the set of first-order formulas of the form $\exists x_1 \ldots \exists x_m\, \forall y_1 \ldots \forall y_n\, \psi$ where ψ is quantifier-free and $m, n \geq 0$. In particular, formulas from Σ_1 and Σ_2 can use more than the variables x and y, but no counting quantifiers $\exists^{\geq k}$ or $\exists^{p \bmod q}$.

Let L be regular and bounded. Then, by Theorem 3.5, we obtain in particular that the Σ_2-theory of $(L, \sqsubseteq)$ is decidable. Note that the regular language $L = \{a, b\}^*$ is not covered by this result since it is unbounded. And, indeed, the Σ_2-theory of $(\{a, b\}^*, \sqsubseteq)$ is undecidable [6]. On the positive side, we know that the Σ_1-theory of $(\{a, b\}^*, \sqsubseteq)$ is decidable [17].

In this section, we generalize this positive result to arbitrary regular languages, i.e., we prove the following result:

Theorem 5.1. *Let $L \subseteq \Sigma^*$ be regular. Then the Σ_1-theory of $\mathcal{S} = (L, \sqsubseteq)$ is decidable.*

The proof for the case $L = \{a, b\}^*$ in [17] essentially relies on the fact that each order $(\mathbb{N}^k, \leq)$, and thus every finite partial order, embeds into $(\{a, b\}^*, \sqsubseteq)$.

In the general case here, the situation is more involved. Take, for example, $L = \{ab, ba\}^*$. Then, orders as simple as $(\mathbb{N}^2, \leq)$ do not embed into $(L, \sqsubseteq)$: this is because the downward closure of any infinite subset of L contains all of L. Nevertheless, we will show, perhaps surprisingly, that every finite partial order embeds into $(L, \sqsubseteq)$. In fact, this holds whenever L is an unbounded regular language. The latter requires a result that we shall prove only later.

Proof (Theorem 5.1). By Theorem 3.5, we may assume that L is unbounded.

By Corollary 5.7 below, every finite partial order embeds into $(L, \sqsubseteq)$.

Hence the formula $\exists x_1 \exists x_2 \cdots \exists x_n \colon \psi$ with ψ quantifier-free holds in $(\Sigma^*, \sqsubseteq)$ if, and only if, it holds in some finite partial order of size at most n. Since there are only finitely many such partial orders, the result follows (cf. [17]). $\square$

The first proposition used in the above proof deals with the existence of certain primitive words for every unbounded regular language.

Proposition 5.2. *For every unbounded regular language $L \subseteq \Sigma^*$, there are $x, u, v, z \in \Sigma^*$ such that $|u| = |v|$, the word uv is primitive, and $x\{u, v\}^* z \subseteq L$.*

Proof. Let $\mathcal{A}$ be some deterministic finite automaton accepting L. Since L is not bounded, there is at least one strongly connected component that is not a cycle but belongs to some path from an initial to some accepting state. Hence there are a state s and words x, y_1, y_2, z such that

- x leads from the initial state to s and z from s to some accepting state and
- y_1 and y_2 lead from s back to s and are not prefixes of each other.

Set $p = y_1 y_2$ and $q = y_2 y_1$. Since y_1 and y_2 are not prefixes of each other, these two words are distinct, lead from s back to s, and are of the same positive length. Hence we have $|p| = |q| > 0$, $p \neq q$, and $x\{p, q\}^* z \subseteq L$. Set $r = pq$ and $s = pp$.

Then $|r| = |s|$ and $x\{r, s\}^* z \subseteq x\{p, q\}^* z \subseteq L$. Suppose r and s are conjugate. Since $s = p^2$, there exist words z_1 and z_2 such that $r = z_2 z_1 z_2 z_1$ with $p = z_2 z_1$, i.e., r is the square of some word $z_2 z_1$ of length $|p| = |q|$. But this contradicts $r = pq$ and $p \neq q$. Hence r and s are not conjugate.

Next let $n = |r|$, $u = r^n$ and $v = s^n$.

By contradiction, we show that uv is primitive.

Since we assume $uv = r^n s^n$ not to be primitive, there is a word $w \in \Sigma^+$ with $r^n s^n \in w w^+$. From $n = |r| = |s—$, we obtain

$$|w| \leq \tfrac{1}{2} |r^n s^n| = n^2 \,.$$

Define the natural number t as follows:

$$t = \begin{cases} n & \text{if } |w| < n \\ 1 & \text{if } |w| \geq n \end{cases}$$

In both cases, we get $n \leq |w^t| \leq n^2 \leq |r^n s^n|$.

Consequently, r and w^t are prefixes of $uv = r^n s^n$ of length n and $\geq n$, respectively. Hence r is a prefix, and therefore a factor, of w^t.

On the other hand, w^t and $v = s^n$ are suffixes of uv of length $\leq n^2$ and n^2, respectively. Hence w^t is a suffix, and therefore a factor, of $v = s^n$.

Taking these two facts together, we obtain that r is a factor of s^n of length $|s|$. Since r and s are not conjugate, this implies $r = s$ and therefore $pq = r = s = pp$. But this contradicts $p \neq q$. □

Recall that x is a subword of y if, intuitively, x can be found in y. We next make this notion precise in order to use the precise notion in the following proofs. So an *embedding* of x in y is an order-preserving injective mapping α from the set of positions $\{1, 2, \ldots, |x|\}$ of x into the set of positions $\{1, 2, \ldots, |y|\}$ of y such that the i-letter of x equals letter number $\alpha(i)$ in y. An embedding α of x into y is called *initial* if $\alpha(1) = 1$, i.e, the left-most position in x hits the left-most position in y. Symmetrically, it is *terminal* if $\alpha(|x|) = |y|$, i.e., the right-most position in x hits the right-most position in y.

We write $x \hookrightarrow y$ if x is a subword of y and every embedding α of x in y is terminal. In other words, $x \sqsubseteq y$, but no word xa with $a \in \Sigma$ is a subword of y or, equivalently $x \hookrightarrow y$ if x is a *prefix-maximal subword of y*.

Note that the relation $\hookrightarrow$ is reflexive, antisymmetric, but not transitive: $a \hookrightarrow ba \hookrightarrow aba$, but $a \not\hookrightarrow aba$. We next show that $\hookrightarrow$ is compatible with concatenation of words.

Lemma 5.3. *Let $x, x', y, y' \in \Sigma^*$ with $x \hookrightarrow y$ and $x' \hookrightarrow y'$. Then $xy \hookrightarrow x'y'$.*

Proof. Suppose $xy \not\hookrightarrow x'y'$. Since $xy \sqsubseteq x'y'$, there is $a \in \Sigma$ such that $xya \sqsubseteq x'y'$. Then, either $xb \sqsubseteq x'$ where b is the first letter of ya (contradicting $x \hookrightarrow x'$), or $ya \sqsubseteq y'$ (contradicting $y \hookrightarrow y'$). □

Clearly, w^n has an initial and a terminal embedding in w^{n+1} for any word $w \neq \varepsilon$ and any natural number n. But there can also be other embeddings (consider, e.g., $w = aa$ and $n = 2$, i.e., $w^n = aa\,aa$ and $w^{n+1} = aa\,aa\,aa$). The following lemma describes words w and natural numbers n such that these other embeddings do no exist.

Lemma 5.4. *Let w be primitive and $n > |w|$. Then every embedding of w^n into w^{n+1} is initial or terminal.*

We do not know whether the bound n given in this lemma is optimal or not.

Proof. Let α be an embedding of w^n into w^{n+1}. Consider the n copies of w in the word w^n. We call such a copy *gapless* if its image in w^{n+1} under α is contiguous. Since the length difference between w^n and w^{n+1} is only $|w| < n$, there has to be at least one gapless copy of w, say the ith copy. The image of this copy is a contiguous subword of w^{n+1} that spells w and occurs at some position $i \cdot |w| + j$ with $j \in \{0, \ldots, |w|\}$. Since w is primitive, it can occur as a contiguous subword in w^{n+1} only at positions that are divisible by $|w|$, hence $j \in \{0, |w|\}$. If $j = 0$, then α is initial and if $j = |w|$, then α is terminal. □

Lemma 5.5. *Let $u, v \in \Sigma^*$ be words such that $|u| = |v|$ and uv is primitive. Then, for all $\ell, n \in \mathbb{N}$ with $n > |uv| + \ell + 2$, we have*

(i) $(uv)^n \hookrightarrow v(uv)^n$,
(ii) $(uv)^\ell v(uv)^{n-(\ell+1)} \hookrightarrow (uv)^n$, *and*
(iii) $(uv)^\ell v(uv)^{n-(\ell+1)} \hookrightarrow v(uv)^n$.

Proof. Note that (iii) implies (ii) since any non-terminal embedding of $(uv)^\ell v(uv)^{n-(\ell+1)}$ into $(uv)^n$ induces a non-terminal embedding of the same word into $v(uv)^n$. It therefore remains to be shown that (i) and (iii) hold.

For any of these two claims, we define words x, y, x', and y' as follows:

(i) Set $x = uv$, $y = (uv)^{n-1}$, $x' = v$, and $y' = (uv)^n$.
(iii) Set $x = (uv)^\ell v$, $y = (uv)^{n-(\ell+1)}$, $x' = v(uv)^\ell$, and $y' = (uv)^{n-\ell}$.

In any of these cases, xy is the word on the left of the corresponding claim and $x'y'$ the word on the right. Hence, we have to verify $xy \hookrightarrow x'y'$. First, $xy \sqsubseteq x'y'$ is immediate. Now, let α be some embedding of xy into $x'y'$. Since x is no subword of x', the embedding α induces a non-initial embedding α' of y into $y' = y(uv)$. Since uv is primitive and $n-1, n-(\ell+1) > |uv|$, the embedding α' is terminal by Lemma 5.4. Hence also the embedding α is terminal. Since α was abritrary, we get $xy \hookrightarrow x'y'$ in cases (i) and (iii) as desired. $\square$

Proposition 5.6. *Let $u, v \in \Sigma^*$ be distinct such that uv is primitive and $|u| = |v|$. Then every finite partial order embeds into $(\{u,v\}^*, \sqsubseteq)$.*

Proof. Observe that every finite partial order with m elements embeds into $(\{0,1\}^m, \leq)$ where $\leq$ is the componentwise order. Hence, it suffices to construct an embedding φ of this partial order into $(\{u,v\}^*, \sqsubseteq)$.

So let $n = |uv| + m + 3$ and define, for a tuple $t = (t_1, \ldots, t_m) \in \{0,1\}^m$,

$$\varphi(t_1, \ldots, t_m) = v^{t_1}(uv)^n \cdots v^{t_m}(uv)^n\,.$$

Then, clearly, $s \leq t$ implies $\varphi(s) \sqsubseteq \varphi(t)$.

Conversely let $s = (s_1, \ldots, s_m)$ and $t = (t_1, \ldots, t_m)$ be two vectors from $\{0,1\}^m$ with $\varphi(s) \sqsubseteq \varphi(t)$. Towards a contradiction, suppose $s \not\leq t$. Then there is some $i \in [1,m]$ with $s_i = 1$, $t_i = 0$ and $s_j \leq t_j$ for all $j \in [1, i-1]$.

By induction, we construct for all $k \in [i,m]$ some $\ell_k \in [1,k]$ with

$$v^{s_1}(uv)^n \cdots v^{s_k}(uv)^{n-\ell_k} \hookrightarrow v^{t_1}(uv)^n \cdots v^{t_k}(uv)^n. \tag{6}$$

From Lemma 5.5(i), we get $v^{s_j}(uv)^n \hookrightarrow v^{t_j}(uv)^n$ for every $j \in [1, i-1]$ since $s_j \leq t_j$. Furthermore $v^{s_i}(uv)^{n-1} = v(uv)^{n-1} \hookrightarrow (uv)^n = v^{t_i}(uv)^n$ according to Lemma 5.5(ii) with $\ell = 0$. Therefore, compatibility of $\hookrightarrow$ with concatenation (Lemma 5.3) yields

$$v^{s_1}(uv)^n \cdots v^{s_i}(uv)^{n-1} \hookrightarrow v^{t_1}(uv)^n \cdots v^{t_i}(uv)^n$$

which settles the base case $k = i$ with $\ell_k = 1$. So let $k \in [i, m-1]$ and $\ell_k \in [1,k]$ such that (6) holds. We distinguish three cases.

1. Suppose $s_{k+1} = 0$. Set $\ell_{k+1} = \ell_k$. We have

$$(uv)^n \hookrightarrow v^{t_{k+1}}(uv)^n$$

since either $t_{k+1} = 0$ and the two words are the same, or $t_{k+1} = 1$ and $(uv)^n \hookrightarrow v(uv)^n$ by Lemma 5.5(i). So together with the induction hypothesis (6), compatibility of $\hookrightarrow$ with concatenation yields

$$\begin{aligned}
&v^{s_1}(uv)^n \cdots v^{s_k}(uv)^n v^{s_{k+1}}(uv)^{n-\ell_{k+1}} \\
&\qquad = v^{s_1}(uv)^n \cdots v^{s_k}(uv)^{n-\ell_k} \;\; (uv)^n \\
&\qquad \hookrightarrow v^{t_1}(uv)^n \cdots v^{t_k}(uv)^n \qquad v^{t_{k+1}}(uv)^n\,,
\end{aligned}$$

where the equality is due to $s_{k+1} = 0$ and $\ell_{k+1} = \ell_k$. Since $\ell_{k+1} = \ell_k \in [1, k] \subseteq [1, k+1]$, this proves (6) for $k+1$.
2. Suppose $s_{k+1} = 1$ and $t_{k+1} = 0$. Set $\ell_{k+1} = \ell_k + 1$. By Lemma 5.5(ii), we have

$$(uv)^{\ell_k} v(uv)^{n-\ell_{k+1}} = (uv)^{\ell_k} v(uv)^{n-(\ell_k+1)} \hookrightarrow (uv)^n\,.$$

So together with the induction hypothesis (6), compatibility of $\hookrightarrow$ with concatenation implies

$$\begin{aligned}
&v^{s_1}(uv)^n \cdots v^{s_k}(uv)^n v^{s_{k+1}}(uv)^{n-\ell_{k+1}} \\
&\qquad = v^{s_1}(uv)^n \cdots v^{s_k}(uv)^{n-\ell_k} \;\; (uv)^{\ell_k} v(uv)^{n-\ell_{k+1}} \\
&\qquad \hookrightarrow v^{t_1}(uv)^n \cdots v^{t_k}(uv)^n \qquad (uv)^n \\
&\qquad = v^{t_1}(uv)^n \cdots v^{t_k}(uv)^n \qquad v^{t_{k+1}}(uv)^n
\end{aligned}$$

where the first equality holds by $s_{k+1} = 1$ and the last one by $t_{k+1} = 0$. Since $\ell_{k+1} = \ell_k + 1 \in [2, k+1] \subseteq [1, k+1]$, this proves (6) for $k+1$.
3. Finally suppose $s_{k+1} = t_{k+1} = 1$. Set $\ell_{k+1} = \ell_k + 1$. Then Lemma 5.5(iii) tells us that

$$(uv)^{\ell_k} v(uv)^{n-\ell_{k+1}} = (uv)^{\ell_k} v(uv)^{n-(\ell_k+1)} \hookrightarrow v(uv)^n\,.$$

So together with the induction hypothesis (6), compatibility of $\hookrightarrow$ with concatenation implies

$$\begin{aligned}
&v^{s_1}(uv)^n \cdots v^{s_k}(uv)^n v^{s_{k+1}}(uv)^{n-(\ell+1)} \\
&\qquad = v^{s_1}(uv)^n \cdots v^{s_k}(uv)^{n-\ell} \;\; (uv)^{\ell} v(uv)^{n-(\ell+1)} \\
&\qquad \hookrightarrow v^{t_1}(uv)^n \cdots v^{t_k}(uv)^n \qquad v(uv)^n \\
&\qquad = v^{t_1}(uv)^n \cdots v^{t_k}(uv)^n \qquad v^{t_{k+1}}(uv)^n
\end{aligned}$$

where the equalities follow from $s_{k+1} = t_{k+1} = 1$. Since $\ell_{k+1} = \ell_k + 1 \in [2, k+1] \subseteq [1, k+1]$, this proves (6) for $k+1$.

This completes the induction. Therefore, we have in particular

$$v^{s_1}(uv)^n \cdots v^{s_m}(uv)^{n-\ell_m} \hookrightarrow v^{t_1}(uv)^n \cdots v^{t_m}(uv)^n = \varphi(t)$$

for some $\ell_m > 0$. Since the left-hand side is a proper prefix of $\varphi(s)$, this contradicts $\varphi(s) \sqsubseteq \varphi(t)$. $\square$

Corollary 5.7. *Let $L \subseteq \Sigma^*$ be unbounded and regular. Then every finite partial order embeds into $(L, \sqsubseteq)$.*

Proof. By Proposition 5.2, there are words $x, u, v, z \in \Sigma^*$ with $|u| = |v|$, uv primitive, and $x\{u, v\}^* z \subseteq L$. By Proposition 5.6, any finite partial order embeds into $(\{u, v\}^*, \sqsubseteq)$ and therefore into $(x\{u, v\}^* z, \sqsubseteq)$ which is a substructure of $(L, \sqsubseteq)$. □

This completes the proof of the main result of this section, i.e., of Theorem 5.1.

6 The Σ_1-theory with constants

By Theorem 5.1, the Σ_1-theory of $(L, \sqsubseteq)$ is decidable for all regular languages L. If L is bounded, then even the Σ_1-theory of $(L, \sqsubseteq, (w)_{w \in L})$ is decidable (Theorem 3.5). This result does not extend to all regular languages since, e.g., the Σ_1-theory of $(\Sigma^*, \sqsubseteq, (w)_{w \in \Sigma^*})$ is undecidable [6]. In this section, we present another class of regular languages L (besides the bounded ones) such that

$$(L, \sqsubseteq, (w)_{w \in L})$$

has a decidable Σ_1-theory.

Let $L \subseteq \Sigma^*$ be some language. Then *almost all words from L have a non-negligible number of occurrences of every letter* if there exists a positive real number ε such that for all $a \in \Sigma$ and all but finitely many words $w \in L$, we have $|w|_a \geq \varepsilon \cdot |w|$.

An example of such a regular language is $\{ab, ba\}^*$ (this class contains all finite languages, is closed under union and concatenation and under iteration, provided every word of the iterated language contains every letter).

For $w \in \Sigma^*$, let $w{\uparrow}$ denote the set of superwords of w, i.e., the upward closure of $\{w\}$ in $(\Sigma^*, \sqsubseteq)$.

The basic idea is, as in the proof of Theorem 5.1, to embed every finite partial order into $(L, \sqsubseteq)$. The following lemma refines this embedability. Furthermore, it shows that $L \setminus w{\uparrow}$ is finite in this case.

Lemma 6.1. *Let $L \subseteq \Sigma^*$ be an unbounded regular language such that almost all words from L have a non-negligible number of occurrences of every letter. Let $w \in \Sigma^*$.*

Then $L \setminus w{\uparrow}$ is finite and $L \cap w{\uparrow}$ is unbounded and regular.

Proof. Let $M = (Q, \Sigma, \iota, \delta, F)$ be a deterministic finite automaton accepting L. We show that all words $v \in L$ of length at least $|Q| \cdot |w|$ belong to $w{\uparrow}$ implying that $L \setminus w{\uparrow}$ is finite. So let $v \in L$ with $|v| \geq |Q| \cdot |w|$. Then we can factorize the word v into $v = v_0 v_1 \cdots v_{|w|+1}$ with $v_i \neq \varepsilon$ for all $i \in \{0, \ldots, |w| + 1\}$ such that $\delta(\iota, v_0) = \delta(\iota, v_0 v_1 \cdots v_i)$ for all $1 \leq i \leq |w|$. With $q = \delta(\iota, v_0)$, we obtain $\delta(q, v_i) = q$ for all $1 \leq i \leq |w|$ and therefore $v_0 v_1 \cdots v_{i-1} v_i^* v_{i+1} \cdots v_{|w|+1} \subseteq L$.

Suppose the letter $a \in \Sigma$ does not appear in the word v_i. Then

$$\frac{|v_0 \cdots v_{i-1}\, v_i^n\, v_{i+1} \cdots v_{|w|+1}|_a}{|v_0 \cdots v_{i-1}\, v_i^n\, v_{i+1} \cdots v_{|w|+1}|} \longrightarrow_{n\to\infty} 0\,.$$

But this contradicts the assumption that almost all words from L have a non-negligible number of occurrences of every letter. Hence the word v_i contains all letters from Σ. Since this holds for all $1 \le i \le |w|$, we obtain $w \sqsubseteq v_1 v_2 \ldots v_{|w|} \sqsubseteq v$ and therefore $v \in w{\uparrow}$. Hence, indeed, $L \setminus w{\uparrow}$ is finite.

Next, let $w{\uparrow} = \{u_1, u_2, \cdots, u_m\}$. Suppose $L \cap w{\uparrow}$ is bounded, say $L \cap w{\uparrow} \subseteq w_1^* w_2^* \cdots w_n^*$. Then $L \subseteq w{\uparrow} \cup (L \cap w{\uparrow})$ is a subset of

$$u_1^* \cdots u_m^*\, w_1^* \cdots w_n^*$$

and therfore bounded. But this contradicts our assumption on L, hence $L \cap w{\uparrow}$ is unbounded. $\square$

Theorem 6.2. *Let $L \subseteq \Sigma^*$ be an unbounded regular language such that almost all words from L have a non-negligible number of occurrences of every letter. Then the Σ_1-theory of $(L, \sqsubseteq, (w)_{w\in L})$ is decidable.*

Proof. We want to show that satisfiability in $(L, \sqsubseteq, (w)_{w\in L})$ is decidable for quantifier-free formulas, i.e., for positive Boolean combinations φ of literals of the following forms (where x and y are arbitrary variables and w an arbitrary word from L):

(i) $x \sqsubseteq w$ (iii) $w \sqsubseteq x$ (v) $x \sqsubseteq y$ (vii) $x = w$
(ii) $x \not\sqsubseteq w$ (iv) $w \not\sqsubseteq x$ (vi) $x \not\sqsubseteq y$

(Note that literals of the form $x = y$ can be written as $x \sqsubseteq y \wedge y \sqsubseteq x$, $x \neq y$ as $x \not\sqsubseteq y \vee y \not\sqsubseteq x$, and similarly $x \neq w$ as $x \not\sqsubseteq w \vee w \not\sqsubseteq x$. Furthermore literals mentioning two words like $u \sqsubseteq v$ can be replaced by $\top = (x \sqsubseteq y \vee x \not\sqsubseteq y)$ or $\bot = (x \sqsubseteq y \wedge x \not\sqsubseteq y)$.)

We first show that also literals of types (i), (ii), and (iv) can be eliminated.

First note that $\{v \in L \mid v \sqsubseteq w\}$ is finite and computable from w. Hence the literal $x \sqsubseteq w$ is equivalent to a finite disjunction of literals of the form $x = w'$.

Next note that $\{v \in L \mid v \not\sqsubseteq w\}$ is upward closed in $(L, \sqsubseteq)$. Since L is regular, we can compute the finite set of minimal elements of this set. Thus, $v \not\sqsubseteq w$ is equivalent to a finite disjunction of literals of the form $w' \sqsubseteq x$.

Finally, by the finiteness claim in Lemma 6.1, also the literal $w \not\sqsubseteq x$ can be replaced by a finite disjunction of literals of the form $x = w'$.

As a result, we get a positive Boolean combination ψ of literals of the form (iii), (v), (vi), and (vii) that is equivalent to φ. We can even assume ψ to be in disjunctive normal form. To verify whether ψ is satisfiable in $(L, \sqsubseteq)$, it therefore suffices to verify satisfiability of conjunctions of literals of the form (iii), (v), (vi), and (vii).

So let γ be such a conjunction. If γ contains literals $x = u$ and $x = v$ for two distinct words u and v, then γ is not satifiable in $(L, \sqsubseteq)$. Otherwise, for all literals $x = w$ in γ, replace all occurrences of x in γ by w, and delete all literals

$x = w$ from γ. This can result in conjuncts of the form $v \sqsubseteq w$; if v is a subword of w, then simply delete this conjuncts, otherwise the conjunction is not satisfiable and can be deleted altogether. The resulting conjunction of literals of the form (iii), (v), and (vi) is equivalent to γ. It can be written as $\gamma_1 \wedge \gamma_2$ where γ_1 is a conjunction of literals of the form (iii) and γ_2 is a conjunction of literals of the form (v) and (vi).

Let $\gamma_1 = \bigwedge_{i \in I} w_i \sqsubseteq x$. Note that

$$L \cap \bigcap_{i \in I} w_i\!\uparrow = \bigcap_{i \in I} L \setminus (L \setminus w_i\!\uparrow) \,.$$

By Lemma 6.1, the sets $L \setminus w_i\!\uparrow$ are finite. Since L is unbounded, it is infinite. Now the finiteness of I implies $L \cap \bigcap_{i \in I} w_i\!\uparrow \neq \emptyset$. Hence there is some word $w \in L$ with $w_i \sqsubseteq w$ for all $i \in I$.

Let n denote the number of variables appearing in γ_2.

If γ is satisfiable in $(L, \sqsubseteq)$, then γ_2 is satisfied by some partial order with at most n elements. Conversely, let γ_2 be satisfied by $(P, \leq)$ where P has at most n elements. By Lemma 6.1, $L \cap w\!\uparrow$ is unbounded and regular. Hence, by Corollary 5.7, $(P, \leq)$ embeds into $(L \cap w\!\uparrow, \sqsubseteq)$. Consequently, $\gamma_1 \wedge \gamma_2$ is satisfiable in $(L \cap w\!\uparrow, \sqsubseteq)$ and therefore in $(L, \sqsubseteq)$. In summary, γ is satifiable in $(L, \sqsubseteq)$ iff γ_2 holds in some finite partial order of size at most n. Since there are only finitely many such finite partial orders, we get that satisfiability of γ in $(L, \sqsubseteq)$ is decidable. $\square$

Open questions

We did not consider complexity issues. In particular, from [18], we know that the C^2-theory (that allows threshold-counting quantifiers, but no modulo-counting quantifiers) of the structure $(\Sigma^*, \sqsubseteq, (w)_{w \in \Sigma^*})$ can be decided in twofold exponential alternating time with linearly many alternations. It follows that the same result holds for the structure $(L, \sqsubseteq, (w)_{w \in L})$ provided L is piecewise-testable. But the techniques from [18] that are based on [12] do not allow to handle the modulo-counting quantifier nor arbitrary regular languages.

We reduced the FO+MOD-theory of the full structure (for L context-free and bounded) to the FO+MOD(tuple)-theory of $(\mathbb{N}, +)$ which is known to be decidable in twofold exponential space [5]. But we did not analyse the size of the FO+MOD(tuple)-formula in terms of the size of the FO+MOD-formula.

Finally, we did not give any new undecidability results. For example, we know that the Σ_1-theory of $(L, \sqsubseteq, (w)_{w \in L})$ is undecidable for $L = \Sigma^*$ [6] and decidable for $L = \{ab, ba\}^*$ (Theorem 6.2). To narrow the gap between decidable and undecidable cases, one should find more undecidable cases.

References

1. J. Berstel. *Transductions and context-free languages.* Teubner Studienbücher, Stuttgart, 1979.

2. M. Droste, W. Kuich, and H. Vogler, editors. *Handbook of Weighted Automata*. EATCS Monographs in Theoretical Computer Science. Springer, 2009.
3. M. Droste and D. Kuske. Weighted automata. In J.-E. Pin, editor, *Handbook of Automata Theory*, volume 1, pages 113–150. EMS Press, 2021.
4. A. Finkel and Ph. Schnoebelen. Well-structured transition systems everywhere! *Theoretical Computer Science*, 256:63–92, 2001.
5. P. Habermehl and D. Kuske. On Presburger arithmetic extended with non-unary counting quantifiers. *Logical Methods in Computer Science*, 19(3), 2023.
6. S. Halfon, Ph. Schnoebelen, and G. Zetzsche. Decidability, complexity, and expressiveness of first-order logic over the subword ordering. In *Proc. of the Thirty-Second Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2017)*, pages 1–12. IEEE Computer Society, 2017.
7. G. Higman. Ordering by divisibility in abstract algebras. *Proc. London Math. Soc.*, 2:326–336, 1952.
8. J. Ježek and R. McKenzie. Definability in substructure orderings. I: Finite semilattices. *Algebra Univers.*, 61(1):59–75, 2009.
9. J. Ježek and R. McKenzie. Definability in substructure orderings. III: Finite distributive lattices. *Algebra Univers.*, 61(3-4):283–300, 2009.
10. J. Ježek and R. McKenzie. Definability in substructure orderings. IV: Finite lattices. *Algebra Univers.*, 61(3-4):301–312, 2009.
11. J. Ježek and R. McKenzie. Definability in substructure orderings. II: Finite ordered sets. *Order*, 27(2):115–145, 2010.
12. P. Karandikar and Ph. Schnoebelen. Decidability in the logic of subsequences and supersequences. In Prahladh Harsha and G. Ramalingam, editors, *Proceedings of the 35th Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'15)*, volume 45 of *Leibniz International Proceedings in Informatics*, pages 84–97. Leibniz-Zentrum für Informatik, 2015.
13. P. Karandikar and Ph. Schnoebelen. The height of piecewise-testable languages and the complexity of the logic of subwords. *Logical Methods in Computer Science*, 15(2), 2019.
14. O. V. Kudinov and V. L. Selivanov. Undecidability in the homomorphic quasiorder of finite labelled forests. *J. Log. Comput.*, 17(6):1135–1151, 2007.
15. O. V. Kudinov, V. L. Selivanov, and L. V. Yartseva. Definability in the subword order. In *Programs, Proofs, Processes, 6th Conference on Computability in Europe, CiE 2010, Ponta Delgada, Azores, Portugal, June 30 - July 4, 2010. Proceedings*, Lecture Notes in Computer Science vol. 6158, pages 246–255. Springer, 2010.
16. O. V. Kudinov, V. L. Selivanov, and A. V. Zhukov. Definability in the h-quasiorder of labeled forests. *Ann. Pure Appl. Logic*, 159(3):318–332, 2009.
17. D. Kuske. Theories of orders on the set of words. *Theoretical Informatics and Applications*, 40:53–74, 2006.
18. D. Kuske and Ch. Schwarz. Alternating complexity of counting first-order logic for the subword order. *Acta Informatica*, 60(1):79–100, 2023.
19. D. Kuske and G. Zetzsche. Languages ordered by the subword order. In *FoSSaCS'19*, Lecture Notes in Comp. Science vol. 11425, pages 348–364. Springer, 2019.
20. R. Parikh. On context-free languages. *Journal of the ACM*, 13(4):570–581, 1966.
21. J. Sakarovitch. *Elements of Automata Theory.* Cambridge University Press, 2009.
22. J. Sakarovitch. Rational and recognizable series. In M. Droste, W. Kuich, and H. Vogler, editors, *Handbook of Weighted Automata*, pages 105–174. Springer, 2009.
23. A. Salomaa and M. Soittola. *Automata-Theoretic Aspects of Formal Power Series.* Springer, 1978.

24. Ramanathan S. Thinniyam. Definability of recursive predicates in the induced subgraph order. In *Logic and Its Applications - 7th Indian Conference, ICLA 2017, Kanpur, India, January 5-7, 2017, Proceedings*, Lecture Notes in Comp. Science vol. 10119, pages 211–223. Springer, 2017.
25. Ramanathan S. Thinniyam. Defining recursive predicates in graph orders. *Logical Methods in Computer Science*, 14(3), 2018.