The subtrace order and counting first-order logic

Dietrich Kuske

Technische Universität Ilmenau, Germany

Abstract. We study the subtrace relation among Mazurkiewicz traces which generalizes the much-studied subword order. Here, we consider the 2-variable fragment of a counting extension of first-order logic with regular predicates. It is shown that all definable trace languages are effectively recognizable implying that validity of a sentence of this logic is decidable (this problem is known to be undecidable for virtually all stronger logics already for the subword relation).

Keywords: Mazurkiewicz traces · counting logic · subword relation.

1 Introduction

The subword relation is one of the simplest nontrivial examples of a well-quasi ordering [7] and can be used in the verification of infinite state systems [4]. It 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. [11, 8, 21, 19]; this paper initiates its consideration for the class of Mazurkiewicz traces. (The prefix order on the set of traces has been studied extensively before, both order-theoretically (cf. [5]) and under logical aspects (e.g. [15]).)

These traces were first investigated by Cartier and Foata [2] to study the combinatorics of free partially commutative or, equivalently, trace monoids. Later, Mazurkiewicz [16] used them to relate the interleaving and the partial-order semantics of a distributed system (see [3] for surveys on the many results on trace monoids).

Many of the above mentioned papers on the embeddability relation study its logical aspects. Regarding the subword relation, they provide a rather sharp description of the border between decidable and undecidable fragments of first-order logic: For the subword order alone, the $\exists^*$-theory is decidable [12] and the $\exists^*\forall^*$-theory is undecidable [9]. For the subword order together with regular predicates, the two-variable theory is decidable [9] (this holds even for the two-variable fragment of a counting extension of first-order logic [14]) and the three-variable theory [9] as well as the $\exists^*$-theory are undecidable [6] (even if we only consider singleton predicates, i.e., constants). If one restricts the universe from all words to a particular language, an even more diverse picture appears [14].

All the undecidability results hold for the subtrace relation since it generalizes the subword relation. The strongest decidability result for the subword relation is the decidability of the 2-variable fragment of a counting extension of first-order logic [14]. The proof shows that every definable unary relation is an effectively regular language. It proceeds by quantifier elimination and relies crucially on the fact that the downwards closure, the upwards closure, and the "incomparability language" (i.e., the set of words that are incomparable to some element of the language) of a regular language is effectively regular. These three preservation results hold since the subword relation and the incomparability relation are unambiguous rational transductions [9].

Considering the subtrace relation, the main result of this paper shows the decidability of the 2-variable fragment of the extension of first-order logic by threshold-counting quantifiers. This extends results by Karandikar and Schnoebelen [9] and by Kuske and Zetzsche [14] from words to traces. As their proofs for words, we proceed by quantifier elimination and rely on the preservation properties mentioned above, but this time for trace languages. Differently from the study of subwords, here we cannot use rational relations for traces since they do not preserve recognizability (and are not available for other classes of structures at all).

To substitute the use of rational relations, we consider the internal structure of a trace, i.e., we consider a trace not as an element of a monoid, but as a labeled directed graph. Now monadic second order (abbreviated MSO) logic can be used to make statements about such a graph. Generalizing Büchi's result, Thomas [20] showed that a set of traces is recognizable if, and only if, it is the set of models of some MSO-sentence. With this shift of view, we have to prove the preservation results not for recognizable, but for MSO-definable sets of traces. This is rather straightforward for the upwards closure since a trace has a subtrace satisfying some MSO-sentence σ if, and only if, some induced subgraph satisfies σ which is easily expressible in MSO logic. Since we consider also threshold counting quantifiers, we have to express, e.g., that there are two non-isomorphic induced subgraphs satisfying σ. Since isomorphism is not expressible in MSO logic, the solution relies on "leftmost" or "canonical" subgraphs. When talking about the incomparability relation, we are interested in traces (i.e., graphs) that are neither a sub- nor a supergraph. We base the solution on the largest prefix of one trace that is a subtrace of the other trace as well as on the combinatorics of traces and, in particular, on MSO logic.

Methodwise, we derive the decidability without the use of rational relations. Instead, our arguments are based on the rich theory of traces and in particular on the relation between recognizability and MSO-definability in this setting. It remains to be explored whether these ideas can be transfered to other settings where rational relations are not available.

2 Definitions and main result

2.1 Traces and subtraces

A *dependence alphabet* is a pair (Σ, D) where Σ is a finite alphabet and the *dependence relation* $D \subseteq \Sigma^2$ is symmetric and reflexive.

A *trace over* (Σ, D) is (an isomorphism class of) a directed acyclic graph $t = (V, E, \lambda)$ with node-labels from Σ (i.e., $\lambda\colon V \to \Sigma$) such that, for all $x, y \in V$,

- $(x, y) \in E \implies (\lambda(x), \lambda(y)) \in D$ and
- $(\lambda(x), \lambda(y)) \in D \implies (x, y) \in E$ or $x = y$ or $(y, x) \in E$.

The set of all traces is denoted $\mathbb{M}(\Sigma, D)$, 1 is the unique trace with empty set of nodes. For two traces $s = (V_s, E_s, \lambda_s)$ and $t = (V_t, E_t, \lambda_t)$, we define their product $s \cdot t = u = (V_u, E_u, \lambda_u)$ setting $V_u = V_s \uplus V_t$, $\lambda_u = \lambda_s \cup \lambda_t$, and $E_u = E_s \cup E_t \cup \{(x, y) \in V_s \times V_t \mid (\lambda_s(x), \lambda_t(y)) \in D\}$.

This operation is easily seen to be associative with neutral element 1, i.e., $\mathbb{M}(\Sigma, D)$ forms a monoid that we call *trace monoid (induced by* (Σ, D)*)*.

Let $a \in \Sigma$. Abusing notation, we denote the singleton trace $(\{x\}, \emptyset, \{(x, a)\})$ by a. Then the monoid $\mathbb{M}(\Sigma, D)$ is generated by the set Σ of singleton traces.

Note that $\mathbb{M}(\Sigma, \{(a, a) \mid a \in \Sigma\}) \cong (\mathbb{N}, +)^{|\Sigma|}$ and $\mathbb{M}(\Sigma, \Sigma \times \Sigma) \cong \Sigma^*$. Further, the direct and the free product of two trace monoids is a trace monoid, again. But there are also trace monoids not arising by free and direct products from free monoids (consider, e.g., the dependence alphabet with $\Sigma = \{a_1, a_2, a_3, a_4\}$ and $(a_i, a_j) \in D \iff |i - j| \leq 1$). See [3] for a collection of surveys on the many results known for trace.

Let $t = (V, E, \lambda)$ be a trace. To simplify notation, we write $X \subseteq t$ for "X is a set of nodes of t", i.e., for $X \subseteq V$.

Now let $X \subseteq t$. Then $t{\restriction_X}$ denotes the subgraph of t induced by X, i.e., $(X, E \cap X^2, \lambda{\restriction_X})$. Note that $s = t{\restriction_X}$ is a trace that we call *subtrace of* t *(induced by* X*)*. We denote this fact by $s \sqsubseteq_{\text{sub}} t$ and call t a *supertrace* of s.

It can be observed that $s \sqsubseteq_{\text{sub}} t$ if, and only if, there are a natural number $n \geq 0$ and traces $s_1, s_2, \ldots, s_n$ and $t_0, t_1, \ldots, t_n$ such that $s = s_1 s_2 \cdots s_n$ and $t = t_0 s_1 t_1 s_2 t_2 \cdots s_n t_n$.

2.2 Recognizable sets

Let $(M, \cdot, 1)$ be some monoid. A set $S \subseteq M$ is *recognizable* if there exists a monoid homomorphism $\eta\colon (M, \cdot, 1) \to M'$ into some finite monoid M' such that $\eta(s) = \eta(t)$ and $s \in S$ imply $t \in S$ for all $s, t \in M$. We call the triple $(M', \eta, \eta(S))$ an *automaton accepting* S.

2.3 The logic $\mathbf{C}^2$ and the main result

Let (Σ, D) be some dependence alphabet and let $\mathcal{R}$ denote the class of recognizable subsets of $\mathbb{M}(\Sigma, D)$. We consider the structure

$$\mathcal{S} = (\mathbb{M}(\Sigma, D), \sqsubseteq_{\text{sub}}, \mathcal{R})$$

whose universe is the set of traces, whose only binary relation is the subtrace relation and that has a unary relation for each recognizable subset of $\mathbb{M}(\Sigma, D)$. We will make statements about this structure using some variant of classical first-order logic. More precisely, the formulas of C^2 are defined by the following syntax:

$$\varphi := x_1 \sqsubseteq_{\text{sub}} x_2 \mid x_1 = x_2 \mid x_1 \in S \mid \varphi \vee \varphi \mid \neg\varphi \mid \exists^{\geq k} x_1\, \varphi$$

where x_1, x_2 are variables from $\{x, y\}$, $S \in \mathcal{R}$ is some recognizable set, and $k \in \mathbb{N}$. Note that we allow only two variables, namely x and y. The semantics of these formulas is as expected with the understanding that $\exists^{\geq k} x_1\, \varphi$ holds if there are at least k mutually distinct traces $t_1, t_2, \cdots, t_k \in \mathbb{M}(\Sigma, D)$ that all make the formula φ true. Note that $\exists^{\geq 1}$ is the usual existential quantifier and that $\exists^{\geq 0} x\, \varphi$ is always true. Now we can formulate the main result of this paper and sketch its proof from results to be demonstrated in later sections:

Theorem 2.1. *If $\varphi(x)$ is a formula from* C^2 *with a single free variable, then the set of traces $S(\varphi) = \{t \in \mathbb{M}(\Sigma, D) \mid \mathcal{S} \models \varphi(t)\}$ is recognizable.*

Even more, from the dependence alphabet (Σ, D) and the formula φ, one can compute an automaton accepting this set. Consequently, the C^2*-theory of $(\mathbb{M}(\Sigma, D), \sqsubseteq_{\text{sub}}, \mathcal{R})$ is decidable uniformly in (Σ, D).*

Proof. The proof proceeds by induction on the construction of the formula φ, the most interesting case being $\varphi = \exists^{\geq k} x\, \psi(x, y)$. Using arguments like de Morgan's laws and basic arithmetic, one can reduce this to the case that $\psi(x, y)$ is a conjunction of possibly negated formulas of the following form:

(a) $x \sqsubseteq_{\text{sub}} y$, $x \sqsubseteq_{\text{sub}} x$, $y \sqsubseteq_{\text{sub}} x$, $y \sqsubseteq_{\text{sub}} y$
(b) $x \in S$ and $y \in S$ for $S \in \mathcal{R}$
(c) $\exists^{\geq \ell} x \colon \alpha(x, y)$ and $\exists^{\geq \ell} y \colon \alpha(x, y)$

Since formulas of the form (c) have at most one free variable, we can apply the induction hypothesis, i.e., replace them by formulas of the form (b). Since $\mathcal{R}$ is closed under Boolean operations, there are $S_i, T_i \in \mathcal{R}$ such that the formula ψ is equivalent to the formula

$$\begin{aligned} y \in T_1 \vee & (x \sqsubseteq_{\text{sub}} y \wedge y \not\sqsubseteq_{\text{sub}} x \wedge x \in S_2 \wedge y \in T_2) \\ \vee & (x \not\sqsubseteq_{\text{sub}} y \wedge y \sqsubseteq_{\text{sub}} x \wedge x \in S_3 \wedge y \in T_3) \\ \vee & (x \not\sqsubseteq_{\text{sub}} y \wedge y \not\sqsubseteq_{\text{sub}} x \wedge x \in S_4 \wedge y \in T_4)\,. \end{aligned}$$

Since the order relations between x and y in this formula are mutually exclusive, the formule φ is equivalent to a Boolean combination of formulas of the form $y \in T$ and

$$\exists^{\geq \ell} x \colon (x\ \theta_1\ y \wedge y\ \theta_2\ x \wedge x \in S \wedge y \in T)$$

with $\theta_1, \theta_2 \in \{\sqsubseteq_{\text{sub}}, \not\sqsubseteq_{\text{sub}}\}$, $\ell \leq k$ and $S, T \in \mathcal{R}$. Depending on θ_1 and θ_2, this last formula defines a Boolean combination of T and sets of traces t satisfying

> S contains $\geq \ell$ traces s that are a proper subtrace of (a proper supertrace of, are incomparable with, resp.) t.

Theorems 3.4, 4.5, and 5.13 demonstrate that these sets are effectively recognizable which completes this proof. ◻

The proofs of the three results on recognizable trace languages (Theorems 3.4, 4.5, and 5.13) are the content of the remaining paper. But before, we formulate a simple consequence that describes the expressive power of the logic C^2.

Corollary 2.2. *Let $R \subseteq \mathbb{M}(\Sigma, D)^2$. Then the following are equivalent:*

1. *There is some $\varphi(x,y) \in \mathrm{C}^2$ such that $R = \{(s,t) \in \mathbb{M}(\Sigma, D)^2 \mid \mathcal{S} \models \varphi(s,t)\}$.*
2. *R is a finite union of relations of the form $\{(s,t) \in S \times T \mid s\ \theta_1\ t\ \theta_2\ s\}$ where S and T are recognizable subsets of $\mathbb{M}(\Sigma, D)$ and $\theta_1, \theta_2 \in \{\sqsubseteq_{\mathrm{sub}}, \not\sqsubseteq_{\mathrm{sub}}\}$.*

By Mezei's theorem (cf. [1]), this can be reformulated as "R is a Boolean combination of recognizable subsets of the monoid $\mathbb{M}(\Sigma, D)^2$ and the subtrace relation."

2.4 Auxiliary definitions

Let $E \subseteq V^2$ be a binary relation (e.g., a partial order or an acyclic relation). Then $vE = \{w \in V \mid (v,w) \in E\}$ and $Ev = \{w \in V \mid (w,v) \in E\}$ for $v \in V$. A set $X \subseteq V$ is *downwards closed wrt.* E if $Ex \subseteq X$ for all $x \in X$. By $X{\downarrow_E}$, we denote the least downwards closed subset of V containing X. A node $v \in V$ is *maximal in* V if $vE = \emptyset$, $\max(V,E)$ denotes the set of maximal elements of V. Dually, we define upwards closed sets, $X{\uparrow_E}$, and minimal elments of V.

Let $t = (V, E, \lambda) \in \mathbb{M}(\Sigma, D)$ be a trace. Then $|t| = |V|$ denotes the size of t, i.e., its number of nodes. We write $|t|_a$ for the number of nodes of t that are labeled by a (for $a \in \Sigma$). By $\mathrm{alphmin}(t)$, we denote the set of letters $\lambda(v)$ for $v \in \min(t)$.

Let $s, t \in \mathbb{M}(\Sigma, D)$ be traces. We call s a *prefix* of t (denoted $s \sqsubseteq_{\mathrm{pref}} t$) if there exists a trace s' with $s \cdot s' = t$. The set of all prefixes of t forms a finite lattice under the relation $\sqsubseteq_{\mathrm{pref}}$. Even more, any set L of traces that all are prefixes of some trace t have a least upper bound that we denote $\sup(L)$ and call the *supremum of* L.

Let, again, $t = (V, E, \lambda) \in \mathbb{M}(\Sigma, D)$ be a trace and $A \subseteq \Sigma$. Then $X = \lambda^{-1}(A){\downarrow_E} \subseteq V$ is the set of nodes of t that are dominated by some node whose label belongs to A. We denote $t{\restriction_X}$ by $\partial_A(t)$. This is the smallest prefix s of t such that $|s|_a = |t|_a$ for all letters $a \in A$. We write $\partial_b(t)$ for $\partial_{\{b\}}(t)$ for $b \in \Sigma$. In this context, we also need the definition $D(B) = \bigcup_{b \in B} Db$ for $B \subseteq \Sigma$ of letters that are dependent from some letter in B.

3 Downward closure

Definition 3.1. *Let S be a set of traces and $k \in \mathbb{N}$. Then $S{\downarrow^{\geq k}}$ is the set of traces t such that there are $\geq k$ traces $s \in S$ with $t \sqsubseteq_{\mathrm{sub}} s$.*

Note that $S{\downarrow^{\geq 1}}$ is the usual downward closure $S{\downarrow_{\sqsubseteq_{\mathrm{sub}}}}$ of S as defined above. It is our aim to prove that $S{\downarrow^{\geq k}}$ is effectively recognizable if S is recognizable.

Lemma 3.2. *Let $S \subseteq \mathbb{M}(\Sigma, D)$ be a recognizable trace language. Then the trace language $S{\downarrow}^{\geq 1}$ is effectively recognizable.*

Proof. The set S is effectively rational [17, Theorem 2]. By induction on the rational expression denoting S, one can construct a starfree expression denoting $S{\downarrow}^{\geq 1}$. Since $\mathcal{R}$ is effectively closed under Boolean operations and concatenation (cf. [3]), the result follows. ☐

Lemma 3.3. *Let $S \subseteq \mathbb{M}(\Sigma, D)$ be a recognizable set of traces and $k \geq 1$. Then the trace language $S{\downarrow}^{\geq 1} \setminus S{\downarrow}^{\geq k}$ is effectively recognizable.*

Proof. Let n be the size of some automaton accepting S. A pumping argument shows that all traces from $S{\downarrow}^{\geq 1}$ of length $\geq n$ also belong to $S{\downarrow}^{\geq k}$. Consequently, the difference of these two sets is finite and therefore recognizable. ☐

Now $S{\downarrow}^{\geq k}$ is effectively recognizable since it is the difference of the two sets from the two lemmas above. Note that a trace t has $\geq k$ *proper* supertraces in S if, and only if, it belongs to $(S \cap S{\downarrow}^{\geq k+1}) \cup (S{\downarrow}^{\geq k} \setminus S)$. Thus, we showed the following result:

Theorem 3.4. *Let $S \subseteq \mathbb{M}(\Sigma, D)$ be recognizable and $k \in \mathbb{N}$ with $k \geq 1$. Then the set of traces t with at least k distinct* proper *supertraces from S is effectively recognizable.*

4 Upward closure

Definition 4.1. *Let S be a set of traces and $k \in \mathbb{N}$. Then $S{\uparrow}_{\geq k}$ is the set of traces t such that there are $\geq k$ traces $s \in S$ with $s \sqsubseteq_{\text{sub}} t$.*

It is our aim to prove that $S{\uparrow}_{\geq k}$ is effectively recognizable if S is recognizable. The main tool in this section (and also in the following one) is a logic that talks about the internal structure of a trace $t = (V, E, \lambda)$.

The logic C^2 considers traces as elements of the structure $(\mathbb{M}(\Sigma, D), \sqsubseteq_{\text{sub}}, \mathcal{R})$ such that it allows to describe "external" properties of traces (e.g., the existence of at least two subtraces in a recognizable set S). We now shift our point of view and look at traces as relational structures. Then logical formulas describe their "internal" properties (e.g., the existence of two a-labeled nodes).

To define the set of MSO-formulas, we fix a set of first-order and a (disjoint) set of monadic second-order variables (the former are usually denoted by small letters, the latter by capital letters). Then *MSO-formulas* are defined by the following syntax (where x and y are first-order variables, X is a second-order variable, and $a \in \Sigma$):

$$\varphi := (x = y) \mid \lambda(x) = a \mid (x, y) \in E \mid x \in X \mid \varphi \vee \varphi \mid \neg\varphi \mid \exists x\, \varphi \mid \exists X\, \varphi\,.$$

Henceforth, we will speak of "formulas" when we actually mean "MSO-formulas".

The satisfaction relation $\models$ between a trace $t = (V, E, \lambda)$ and a formula φ is defined in the obvious way with the understanding that first-order variables denote single nodes and second-order variables denote sets of nodes of the trace.

Definition 4.2. *Let S be a set of traces. Then S is* definable *if there exists a sentence φ with $S = \{s \in \mathbb{M}(\Sigma, D) \mid s \models \varphi\}$.*

Since the notions "definable" and "recognizable" are effectively equivalent for sets of traces [20], we can reformulate the aim of this section as "if $S \subseteq \mathbb{M}(\Sigma, D)$ is definable, then so is $S\uparrow_{\geq k}$".

Consequently, we have to write down a formula that holds in a trace t if, and only if, it has at least k subtraces from S. The idea is to express that there are k distinct subsets of t that all induce traces from S. The problem we face here is that distinct subsets can induce the same subtrace. This problem is solved by choosing the "minimal", "leftmost" or, as we call it, "canonical" set X.

Definition 4.3. *Let t be some trace and $Z \subseteq t$. Then Z is* canonical in t *if $t \models \mathrm{canon}(Z)$, where $\mathrm{canon}(Z)$ is the formula*

$$\forall x, z\colon \left(\begin{array}{r} \big(\lambda(x) = \lambda(z) \wedge x \notin Z \wedge z \in Z \wedge (x, z) \in E\big) \qquad\qquad \\ \rightarrow \exists y \in Z\colon \big((x, y) \in E \wedge (y, z) \in E\big) \end{array} \right) .$$

Then we can show that every subtrace of t is induced by precisely one set canonical in t:

Theorem 4.4. *Let $s \sqsubseteq_{\mathrm{sub}} t$ be traces. Then there is a unique canonical set $X \subseteq t$ with $s \cong t{\restriction}_X$.*

Theorem 4.4 allows us to obtain the main result of this section:

Theorem 4.5. *Let $S \subseteq \mathbb{M}(\Sigma, D)$ be definable and $k \in \mathbb{N}$ with $k \geq 1$. Then the set $S\uparrow_{\geq k}$ is effectively definable. Similarly, the set of traces with $\geq k$* proper *subtraces from S is effectively definable.*

Proof. Let σ be a sentence defining S and consider the sentence

$$\exists X_1, X_2, \ldots, X_k \left(\bigwedge_{1 \leq i \leq k} \big(\sigma{\restriction}_{X_i} \wedge \mathrm{canon}(X_i)\big) \wedge \bigwedge_{1 \leq i < j \leq k} X_i \neq X_j \right)$$

where $\sigma{\restriction}_X$ arises from σ by restricting all quantifications to elements and subsets of X. By Theorem 4.4, it defines the set $S(\sigma)\downarrow^{\geq k}$. To show the claim about proper subtraces, we require the sets X_i to be different from the set of all nodes. □

5 Incomparable traces

For two traces s and t, we write $s \parallel t$ as abbreviation for $t \not\sqsubseteq_{\mathrm{sub}} s \not\sqsubseteq_{\mathrm{sub}} t$.

Definition 5.1. *Let S be a set of traces and $k \in \mathbb{N}$. Then $S^{\parallel}_{\geq k}$ is the set of traces t such that there are $\geq k$ traces $s \in S$ satisfying $t \parallel s$.*

It is our aim to prove that $S^{\parallel}_{\geq k}$ is effectively definable if S is definable.

Two traces s and t are incomparable if, and only if, either $|s| \leq |t|$ and $s \not\sqsubseteq_{\mathrm{sub}} t$, or $|s| > |t|$ and $t \not\sqsubseteq_{\mathrm{sub}} s$. In the following two subsections, we will consider these two cases separately.

5.1 Short non-subtraces

Definition 5.2. *Let S be a set of traces and $k \in \mathbb{N}$. Then $S^{\text{short}}_{\geq k}$ is the set of traces t such that there are $\geq k$ traces $s \in S$ with $|s| \leq |t|$ and $s \not\sqsubseteq_{\text{sub}} t$.*

Let S be defined by the sentence σ. We have to formulate, as a property of the labeled directed graph $t = (V, E, \lambda)$, the existence of k models s of σ that all are incomparable with t and have length at most $|t|$. The idea is to split a trace s into a its largest prefix s_1 that is a subtrace of t and the complementary suffix (using Theorem 4.4, one first shows that s_1 is uniquely defined for any pair of traces (s, t)). Since s_1 is a subtrace of t, Theorem 4.4 ensures that $t \in S^{\text{short}}_{\geq k}$ if, and only if, there are k pairs (X, s_2) such that

(1) $X \subseteq t$ is canonical and $s_2 \in \mathbb{M}(\Sigma, D)$,
(2) $(t{\upharpoonright}_X) \cdot s_2 \models \sigma$,
(3) $t{\upharpoonright}_X = \sup\{s_1 \sqsubseteq_{\text{pref}} (t{\upharpoonright}_X) \cdot s_2 \mid s_1 \sqsubseteq_{\text{sub}} t\}$, and
(4) $1 \leq |s_2| \leq |t| - |X|$.

From Shelah's decomposition theorem [18, Theorem 2.4], we obtain a finite family $(\tau_j, \nu_j)_{j \in J}$ of sentences such that Condition (2) is equivalent to

(2') there exists $j \in J$ with $t{\upharpoonright}_X \models \tau_j$ (equivalently, $(t, X) \models \tau_j{\upharpoonright}_X$) and $s_2 \models \nu_j$.

Thus, we express Condition (2) as a Boolean combination of properties of (t, X) and of s_2. Our next aim is to also express Condition (3) in such a manner.

To this end, let $t, s_2 \in \mathbb{M}(\Sigma, D)$ and $X \subseteq t$. One first shows that Condition (3) holds if, and only if, for all $a \in \text{alphmin}(s_2)$, the trace $(t{\upharpoonright}_X) \cdot a$ is not a subtrace ot t. Let $\mathbb{U}(t, X)$ denote the set of letters $a \in \Sigma$ that violate this last condition, i.e., that satisfy $(t{\upharpoonright}_X) \cdot a \sqsubseteq_{\text{sub}} t$. If X is canonical, we can express the statement $a \in \mathbb{U}(t, X)$ by a formula:

Lemma 5.3. *Let $t = (V, E, \lambda) \in \mathbb{M}(\Sigma, D)$ be a trace, $X \subseteq t$ be canonical, and $a \in \Sigma$. Then $a \in \mathbb{U}(t, X)$ if, and only if, there exists $y \in V$ with $\lambda(y) = a$ and $yE \cap X = \emptyset$.*

Hence, for $A \subseteq \Sigma$, there are formulas $\alpha_A(X)$ and sentences β_A such that

- $(t, X) \models \alpha_A$ if, and only if, $A \cap \mathbb{U}(t, X) = \emptyset$ for all $t \in \mathbb{M}(\Sigma, D)$ and $X \subseteq t$ canonical and
- $s_2 \models \beta_A$ if, and only if, $A = \text{alphmin}(s_2)$.

In summary, we found a family of formulas $(\alpha_A(X), \beta_A)_{A \subseteq \Sigma}$ such that Condition (3) is equivalent to

(3') there exists $A \subseteq \Sigma$ with $(t, X) \models \alpha_A$ and $s_2 \models \beta_A$.

Thus, $t \in S^{\text{short}}_{\geq k}$ if, and only if, there are k pairs (X, s_2) all satisfying the conditions (1), (2'), (3'), and (4). We group these pairs according to their first component. Then $t \in S^{\text{short}}_{\geq k}$ if, and only if, there exists $\ell \leq k$, a function $f \colon \{1, \ldots, \ell\} \to \{0, 1, \ldots, k\}$ with $\sum_{1 \leq i \leq \ell} f(i) = k$, and sets $A_1, A_2, \ldots, A_\ell \subseteq \Sigma$ such that there are mutually distinct canonical sets $X_i \subseteq t$ satisfying, for all $i \in [\ell]$, the existence of some $j \in J$ with

- $(t, X_i) \models \tau_j \wedge \alpha_{A_i}$,
- there are $f(i)$ many traces s_2 of length $1 \le |s_2| \le |t| - |X_i|$ satisfying $s_2 \models \nu_j \wedge \beta_{A_i}$.
  From $\nu_j \wedge \beta_{A_i}$, one can compute a number N such that this holds if, and only if, $|t| - |X_i| \ge N$. Hence this is a property of (t, X_i) that can be expressed by a formula.

All this can be translated into a sentence that only talks about the trace t. Consequently, we obtain

Proposition 5.4. *Let $S \subseteq \mathbb{M}(\Sigma, D)$ be definable and $k \in \mathbb{N}$ with $k \ge 1$. Then $S^{\text{short}}_{\ge k}$ is effectively definable.*

5.2 Long non-supertraces

Definition 5.5. *Let T be a set of traces and $k \in \mathbb{N}$. Then $T^{\text{long}}_{\ge k}$ is the set of traces s such that there are $\ge k$ traces $t \in T$ with $|s| < |t|$ and $s \not\sqsubseteq_{\text{sub}} t$.*

We have to formulate, as a property of the labeled directed graph $s = (V, E, \lambda)$, the existence of k traces $t \in T$ that all are incomparable with s and have length at least $|s| + 1$. The first idea is, again, to split the trace s into its largest prefix s_1 that is a subtrace of t and the complementary suffix. Since this time, we have to formulate properties of s, we would then have to "fill" the prefix s_1 with arbitrarily many nodes to obtain the trace t (more precisely: the minimal prefix of t that contains s_1 as a subtrace). Since this cannot be done with logical formulas, we have to bound this number of "missing pieces". The central notion here is the following:

Definition 5.6. *Let $t = (V, E, \lambda)$ be a trace and $X \subseteq t$. The* number of holes of X in t *equals* $\text{nh}(X, t) = |X{\downarrow_E} \setminus X|$.

Now let s be a trace. If $s \sqsubseteq_{\text{sub}} t$, then $\text{nh}(s, t) = \text{nh}(X, t)$ *where $X \subseteq t$ is canonical with $s = t{\restriction_X}$. If s is not a subtrace of t, then* $\text{nh}(s, t) = \infty$.

The following lemma describes, in terms of the number of holes and the length difference, when a trace is a subtrace of a longer trace:

Lemma 5.7. *Let s, t be traces with $|s| < |t|$. Then $s \parallel t$ if, and only if, $s \neq$* $\sup\{s' \sqsubseteq_{\text{pref}} s \mid \text{nh}(s', t) \le |t| - |s|\}$.

Recall that we have to express, as a property of the labeled directed graph $s = (V, E, \lambda)$, the existence of k properly longer traces $t \in T$ with $s \not\sqsubseteq_{\text{sub}} t$. In doing so, the previous characterisation is particularly useful if the length difference of t and s is fixed. The following lemma, whose proof uses a straightforward pumping argument, allows to do precisely this:

Lemma 5.8. *One can compute a number $n \in \mathbb{N}$ such that the following holds for all $k \in \mathbb{N}$ and $s \in T^{\text{long}}_{\ge k}$: There exist k traces $t \in T$ such that $|s| < |t|$, $s \not\sqsubseteq_{\text{sub}} t$, and $|t| \le |s| + k \cdot (n + 1)$.*

Thus, it suffices to characterize, for all $k \geq 0$ and all length differences $N > 0$, those traces s that allow $\geq k$ traces $t \in T$ with $|t| = |s| + N$ and $s \neq \sup\{s' \sqsubseteq_{\text{pref}} s \mid \text{nh}(s', t) \leq N\}$.

Grouping these traces t according to $\sup\{s' \sqsubseteq_{\text{pref}} s \mid \text{nh}(s', t) \leq N\}$, it suffices to characterise those pairs (s_1, s_2) with $s_2 \neq 1$ (where we think of $s_1 s_2$ as a factorisation of s) that allow $\geq k$ pairs (t_1, t_2) of traces such that

(a) $t_1 t_2 \in T$ and t_1 is the minimal prefix of $t_1 t_2$ with $s_1 \sqsubseteq_{\text{sub}} t_1$,
(b) $|s_1 s_2| + N = |t_1 t_2|$, and
(c) $s_1 = \sup\{s' \sqsubseteq_{\text{pref}} s_1 s_2 \mid \text{nh}(s', t_1 t_2) \leq N\}$.

Note that t_1 is the minimal prefix of $t_1 t_2$ with $s_1 \sqsubseteq_{\text{sub}} t_1$ if, and only if, $s_1 \sqsubseteq_{\text{sub}} t_1$ and, for all prefixes $t' \sqsubseteq_{\text{pref}} t_1$ with $s_1 \sqsubseteq_{\text{sub}} t'$, we have $t' = t_1$. This allows to reformulate the second half of Condition (a) as a condition on the pair (s_1, t_1), only. Since T is definable, Shelah's decomposition theorem allows us to compute a finite family $(\mu_j, \nu_j)_{j \in J}$ of pairs of sentences such that $t_1 t_2 \in T$ if, and only if, there exists $j \in J$ with $t_1 \models \mu_j$ and $t_2 \models \nu_j$.

Consequently, for $k \geq 0$, $N > 1$, and a fixed index $j \in J$, it suffices to characterise those pairs (s_1, s_2) with $s_2 \neq 1$ that allow $\geq k$ pairs (t_1, t_2) of traces such that, besides Conditions (b) and (c), also the following holds:

(a_j) $t_1 \models \mu_j$, $s_1 \sqsubseteq_{\text{sub}} t_1$, and $s_1 \sqsubseteq_{\text{sub}} t' \Rightarrow t' = t_1$ for all $t' \sqsubseteq_{\text{pref}} t_1$ and $s_2 \models \nu_j$.

Let (t_1, t_2) be a pair of traces with these properties. At this point, it comes in handy that $\text{nh}(s_1, t_1 t_2) \leq N \cdot |\Sigma|$ (this holds for any traces s_1, t_1, and t_2). Further, since t_1 is the smallest prefix of $t_1 t_2$ with $s_1 \sqsubseteq_{\text{sub}} t_1$, we get $\text{nh}(s_1, t_1 t_2) = \text{nh}(s_1, t_1) = |t_1| - |s_1|$.

Consequently, we can group these pairs (t_1, t_2) according to the length difference $|t_1| - |s_1|$ (which can be bounded by $N \cdot |\Sigma|$ by the above). Hence, it suffices to characterize, for $k \geq 0$, $N > 0$, $j \in J$ and for a fixed length difference ℓ, those pairs (s_1, s_2) of traces with $s_2 \neq 1$ that allow $\geq k$ pairs (t_1, t_2) of traces such that, besides (a_j) and (c), the following holds:

(b_ℓ) $|s_1| + \ell = |t_1|$ and $|t_2| = |s_2| + N - \ell$.

Note that Conditions (a_j) and (b_ℓ) form a Boolean combination of properties of the pairs (s_1, t_1) and (s_2, t_2), respectively. Our next aim is to ensure that this also holds for Condition (c) which forms the main work in this section.

Lemma 5.9. *Let s_1, s_2, t_1, and t_2 be traces such that t_1 is the minimal prefix of $t_1 t_2$ with $s_1 \sqsubseteq_{\text{sub}} t_1$. Then Condition (c) is equivalent to*

(c_1) For all $a \in \Sigma$, there exists a trace s' with $\partial_a(s_1) \sqsubseteq_{\text{pref}} s' \sqsubseteq_{\text{pref}} s_1$ and $\text{nh}(s', t_1) \leq N$ and
(c_2) For all $b \in \text{alphmin}(s_2)$ and all s' with $\partial_{D(b)}(s_1) \sqsubseteq_{\text{pref}} s' \sqsubseteq_{\text{pref}} s_1$, we have $\text{nh}(s' b, t_1 t_2) > N$.

Condition (c_1) only depends on the pair (s_1, t_1). Since Σ is finite, Condition (c_2) is a Boolean combination of properties of s_2 and of properties of the triple (s_1, t_1, t_2). We now reformulate this last condition using the following lemma.

Lemma 5.10. *Let $b \in \Sigma$, let s_1, t_1, t_2, and s' be traces such that t_1 is the minimal prefix of t_1t_2 with $s_1 \sqsubseteq_{\text{sub}} t_1$ and $\partial_b(s_1) \sqsubseteq_{\text{pref}} s' \sqsubseteq_{\text{pref}} s_1$. Then $\text{nh}(s'b, t_1t_2) > N$ if, and only if, one of the following hold:*

- $N < \text{nh}(s'b, t_1) < \infty$ *or*
- $\text{nh}(s'b, t_1) = \text{nh}(b, t_2) = \infty$ *or*
- $\text{nh}(s'b, t_1) = \infty$, $b \sqsubseteq_{\text{sub}} t_2$, *and* $\text{nh}\big(\sup(s', \partial_{D(B)}(s_1), t_1)\big) + \text{nh}(b, t_2) > N$ *where $B \subseteq \Sigma$ is the set of letters appearing before the first b in t_2.*

Replacing $\text{nh}(s'b, t_1t_2) > N$ in Condition (c_2) by the properties from the above lemma, it turns into a Boolean combination (c_2') of statements

(i) $\text{nh}(b, t_2) < h$ for $h \in \mathbb{N} \cup \{\infty\}$,
(ii) "A is the set of letters appearing in t_2 before the first b" for $A \subseteq \Sigma$,
(iii) "for all traces s' with $\partial_b(s_1) \sqsubseteq_{\text{pref}} s' \sqsubseteq_{\text{pref}} s_1$" followed by a Boolean combination of statements of the form
   - $N \geq \text{nh}(s'b, t_1)$ and $\text{nh}(s'b, t_1) = \infty$ for $b \in \Sigma$, and
   - $\text{nh}\big(\sup\big(\sup(s', \partial_{D(B)}(s_1)), t_1\big)\big) < N_1$ for $B \subseteq \Sigma$ and $N_1 \in \mathbb{N}$.

To finish, let $k \in \mathbb{N}$, $N > 0$, $j \in J$, and $\ell \in \mathbb{N}$. Further, let H denote the set of pairs (s_1, s_2) of traces such that conditions (a_j), (b_ℓ), (c_1), and (c_2') hold for at least k pairs of traces (t_1, t_2). The conjunction of these four properties, that talks about the quadruple (s_1, s_2, t_1, t_2), forms a Boolean combination of properties that talk about the pairs (s_1, t_1) and (s_2, t_2), respectively.

Now we can transform the statement "$(s_1, s_2) \in H$" into a Boolean combination of statements of the following form:

(A) there are $\geq k_1$ traces t_1 satisfying a Boolean combination of statements of the form
   - $t_1 \models \mu_j$,
   - $s_1 \sqsubseteq_{\text{sub}} t_1$ and $s_1 \sqsubseteq_{\text{sub}} t' \;\Rightarrow\; t' = t_1$ for all $t' \sqsubseteq_{\text{pref}} t_1$,
   - $|s_1| + \ell = |t_1|$,
   - $\exists s' \colon \partial_a(s_1) \sqsubseteq_{\text{pref}} s' \sqsubseteq_{\text{pref}} s_1 \wedge \text{nh}(s', t_1) \leq N$, and
   - statements of the form (iii).

(B) there are $\geq k_2$ traces t_2 satisfying a Boolean combination of statements of the following forms:
   - $t_2 \models \nu_j$,
   - $|t_2| = |s_2| + N - \ell$,
   - $b \in \text{alphmin}(s_2)$,
   - $b \sqsubseteq_{\text{sub}} t_2$,
   - $\text{nh}(b, t_2) < h$,
   - "A is the set of letters in t_2 before the first b"

Regarding (B), we can formulate all the properties that do not mention s_2 as a formula. This turns all of (B) into a condition on the minimal letters in s_2 and the length of s_2 which both can be expressed by a formula.

It remains to also express (A) as a property of the trace s_1. As a first step, we replace, in the Boolean combination, all references to s_1 by $t{\restriction_X}$ and add that X is canonical. This gives a formula $\varphi(X)$ talking about (t_1, X) and we have to express that there are $\geq k_1$ models (t_1, X) of $\varphi(X)$ all satisfying $s_1 = t_1{\restriction_X}$. This is achieved by the following result:

Lemma 5.11. *Let $\varphi(X)$ be a formula and $k_1, n \in \mathbb{N}$ such that $(t_1, X) \models \varphi$ implies $|t| - |X| = n$. Then one can construct (from φ, k_1, and n) a sentence ψ such that, for all traces s_1 we have $s_1 \models \psi$ if, and only if, there are $\geq k_1$ pairs (t_1, X) with $(t_1, X) \models \varphi$ and $s_1 = t_1{\restriction}_X$.*

In summary, we obtained the following:

Proposition 5.12. *Let $T \subseteq \mathbb{M}(\Sigma, D)$ be definable and $k \in \mathbb{N}$ with $k \geq 1$. Then the set $T^{\mathrm{long}}_{\geq k}$ is effectively definable.*

Now the following result follows easily from Propositions 5.4 and 5.12:

Theorem 5.13. *Let $S \subseteq \mathbb{M}(\Sigma, D)$ be definable and $k \in \mathbb{N}$ with $k \geq 1$. Then the set $S^{\parallel}_{\geq k}$ of traces t with at least k distinct traces $s \in S$ with $s \not\sqsubseteq_{\mathrm{sub}} t$ and $t \not\sqsubseteq_{\mathrm{sub}} s$ is effectively definable.*

Thus, we demonstrated how to prove Theorems 3.4, 4.5, and 5.13. This closes the gaps left open in our proof of the main result (Theorem 2.1).

6 Concluding remarks

The C+MOD^2-theory of $(\Sigma^*, \sqsubseteq_{\mathrm{sub}}, \mathcal{R})$ is decidable [14]. This logic has, in addition to the logic C^2, modulo-counting quantifiers $\exists^{q,r}$. It seems that the only obstacle in proving the analogous result for the subtrace order is the use of Lemma 5.8 in the proof of Proposition 5.12. Whether this lemma has an analogue in the modulo-counting setting is not clear.

The decision algorithms in this paper (as well as those in [9, 14] for the subword order) are nonelementary. Karandikar and Schnoebelen [10] prove that the FO^2-theory of the subword order can be decided in triply exponential space if we only allow unary languages (instead of all languages from $\mathcal{R}$), current research improves the upper bound to doubly exponential space and extends the result to the C^2-theory [13]. It is not clear whether such an elementary upper bound also holds for the subtrace relation.

Finally, it remains to be explored whether the methods developed in this paper can be applied in other settings where rational relations are not available.

References

1. J. Berstel. *Transductions and context-free languages.* Teubner Studienbücher, Stuttgart, 1979.
2. P. Cartier and D. Foata. *Problemes combinatoires de commutation et rearrangements.* Lecture Notes in Mathematics vol. 85. Springer, Berlin - Heidelberg - New York, 1969.
3. V. Diekert and G. Rozenberg. *The Book of Traces.* World Scientific Publ. Co., 1995.
4. A. Finkel and Ph. Schnoebelen. Well-structured transition systems everywhere! *Theoretical Computer Science*, 256:63–92, 2001.

5. P. Gastin and A. Petit. Infinite traces. In *[3]*, pages 393–486. 1995.
6. S. Halfon, Ph. Schnoebelen, and G. Zetzsche. Decidability, complexity, and expressiveness of first-order logic over the subword ordering. In *LICS'17*, 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. P. Karandikar and Ph. Schnoebelen. Decidability in the logic of subsequences and supersequences. In *FSTTCS'15*, volume 45 of *Leibniz International Proceedings in Informatics*, pages 84–97. Leibniz-Zentrum für Informatik, 2015.
10. 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.
11. O. V. Kudinov, V. L. Selivanov, and L. V. Yartseva. Definability in the subword order. In *CiE'10*, Lecture Notes in Comp. Science vol. 6158, pages 246–255. Springer, 2010.
12. D. Kuske. Theories of orders on the set of words. *Theoretical Informatics and Applications*, 40:53–74, 2006.
13. D. Kuske and Ch. Schwarz. Complexity of counting first-order logic for the subword order. In preparation, 2020.
14. 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.
15. P. Madhusudan. Model-checking trace event structures. In *LICS'03*, pages 371–380. IEEE Computer Society Press, 2003.
16. A. Mazurkiewicz. Concurrent program schemes and their interpretation. Technical report, DAIMI Report PB-78, Aarhus University, 1977.
17. J. McKnight. Kleene's quotient theorems. *Pacific J. Math.*, XIV:1343–1352, 1964.
18. S. Shelah. The monadic theory of order. *Annals of Mathematics*, 102:379–419, 1975.
19. Ramanathan S. Thinniyam. Defining recursive predicates in graph orders. *Logical Methods in Computer Science*, 14(3), 2018.
20. W. Thomas. On logical definability of trace languages. In V. Diekert, editor, *Proceedings of a workshop of the ESPRIT BRA No 3166: Algebraic and Syntactic Methods in Computer Science (ASMICS) 1989*, Report TUM-I9002, Technical University of Munich, pages 172–182, 1990.
21. A. Wires. Definability in the substructure ordering of simple graphs. *Annals of Combinatorics*, 20(1):139–176, 2016.