Language Kleene Algebra with Complement: A Finite Relational Semantics and (Un)decidability

Yoshiki Nakamura

Language Kleene Algebra with Complement: A Finite Relational Semantics and (Un)decidability

Yoshiki Nakamura
Tokyo Institute of Technology, Japan

Abstract

We study the equational theory of Kleene algebra with complement w.r.t. languages. While the equational theory w.r.t. languages coincides with the language equivalence (under the standard language valuation) for terms of Kleene algebra, this coincidence is broken if we extend the terms with complement. In this paper, we present *a finite relational semantics*, which completely characterizes the equational theory above. As applications, we show that the equational theory is (1) undecidable and Π^0_1-complete for full terms; (2) decidable and PSPACE-complete if the complement only applies to variables or constants.

2012 ACM Subject Classification Theory of computation → Equational logic and rewriting; Theory of computation → Logic and verification; Theory of computation → Formal languages and automata theory

Keywords and phrases Kleene algebra, Relational model, Complexity

1 Introduction

Kleene algebra (KA) [12, 7] is an algebraic system for regular expressions consisting of union ($\cup$), composition ($\cdot$), Kleene-star ($\_^*$), empty ($\bot$), and identity (I). In this paper, we consider the equational theory w.r.t. languages for extended KA terms. We write $\mathsf{LANG} \models t = s$ if $t = s$ holds on all models of languages, i.e., for any sets X, when each letter maps to any languages over X, the two expressions, t and s, have the same language. By the completeness of KA, for (non-extended) KA terms, the equational theory w.r.t. languages coincides[1] with the language equivalence under the standard language valuation (i.e., each x maps to the singleton language $\{x\}$): for all KA terms t, s, we have

$$[t] = [s] \quad \Leftrightarrow \quad \mathsf{LANG} \models t = s \qquad (\dagger\text{-KA})$$

where $[u]$ denotes the language of an expression u under the standard language valuation.

The equivalence above is broken in general if we extend KA terms with some operators. The equational theory w.r.t. languages for KA terms with some operators was studied, e.g., with reverse [3], with tests [14] (where languages are of guarded strings, not words), with intersection ($\cap$) [2], with universality ($\top$) [28, 21], with variable complements ($\overline{x}$) [19], and combinations of some of them [4, 5]. For example, for (variable) complements, Equation (†-KA) fails, e.g., by the following counter-example [19]:[2]

$$[\overline{x}] = [\overline{x} \cdot \overline{x}] \qquad \mathsf{LANG} \not\models \overline{x} = \overline{x} \cdot \overline{x}.$$

LANG is more compatible with equational reasoning, in that valid equations in LANG are preserved under substitution (by definition), while those under the language equivalence are not: $[\overline{\overline{x}}] = \{x\} \neq \{xx\} = [\overline{\overline{x}} \cdot \overline{\overline{x}}]$.

Extending with full complement significantly enhances the expressive power. For example, we can define $\top$ and $\cap$ using complement: $\top = \overline{\bot}$ and $t \cap s = \overline{\overline{t} \cup \overline{s}}$. Moreover, we can encode quantifier-free formulas (including hypotheses as Horn formulas) by equations (Sect. 4).

[1] By the completeness theorem of Kleene algebra (e.g., [13]); see also, e.g., [2, 21] and [19, Appendix A].
[2] $[\overline{x}] = \mathbf{V}^* \setminus \{x\} = [\overline{x} \cdot \overline{x}]$ where $\mathbf{V}$ is the alphabet. For $\mathsf{LANG} \not\models \overline{x} = \overline{x} \cdot \overline{x}$: By the valuation $x \mapsto \mathbf{V}^* \setminus \{x\}$.

In this paper, we present *a finite relational semantics* for KA with complement w.r.t. languages: relational subword models RSUB (Sect. 3), which ease us to study the equational theory w.r.t. languages, relationally. Our relational semantics can more robustly characterize the equational theory than known relational semantics, e.g., for KAs (REL) [22, third page] and for KAs with $\top$ (GREL) [28, 21] (Remark 3.11). As applications, by transforming the techniques in [18] from relational models REL to RSUB, we have the following complexity results: the equational theory w.r.t. languages is

1. Π^0_1-complete for KA terms with full complement (Cor. 4.8 and Thms. 4.7 and 4.10);
2. PSPACE-complete for KA terms with variable and constant complements (Thm. 6.10).

The first result gives a natural example that the equational theory w.r.t. languages is undecidable, while the standard language equivalence is decidable (by a standard automata construction). The second result positively settles the open problem posed in [19].

This paper is structured as follows. In Sect. 2, we give basic definitions. In Sect. 3, we give RSUB and show that the quantifier-free theory of LANG coincides with that of RSUB. In Sect. 4, we give a reduction from the quantifier-free theory into the equational theory and show that the equational theory is Π^0_1-complete for KA terms with full complement. In Sect. 5, by using RSUB, we give a graph characterization for KA terms with variable and constant complements. In Sect. 6, we show that the equational theory for KA terms with variable and constant complements is PSPACE-complete. In Sect. 7, we conclude this paper.

2 Preliminaries

We write $\mathbb{N}$ for the set of non-negative integers. For $l, r \in \mathbb{N}$, we write $[l, r]$ for the set $\{i \in \mathbb{N} \mid l \leq i \leq r\}$. For a set X, we write $\wp(X)$ for the power set of X.

For a set X (of *letters*), we write X^* for the set of *words* over X. We write I for the empty word. We write wv for the concatenation of words w and v. A *language* over X is a subset of X^*. We use w, v to denote words and use L, K to denote languages, respectively. For languages $L, K \subseteq X^*$, the *concatenation* $L \cdot K$ and the *Kleene star* L^* is defined by:

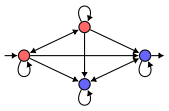
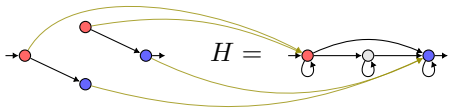
$$L \cdot K \triangleq \{wv \mid w \in L \ \wedge \ w \in K\} \qquad L^* \triangleq \{w_0 \dots w_{n-1} \mid \exists n \in \mathbb{N}, \forall i < n, \ w_i \in L\}.$$

A *(2-pointed) graph* G over a set A is a tuple $\langle |G|, \{a^G\}_{a \in A}, 1^G, 2^G \rangle$, where $|G|$ is a non-empty set (of vertices), each $a^G \subseteq |G|^2$ is a binary relation, and $1^G, 2^G \in |G|$ are vertices. Let G, H be graphs over a set A. For a map $f \colon |G| \to |H|$, we say that f is a *graph homomorphism* from G to H, written $f \colon G \longrightarrow H$ if for all x, y, and a, $\langle x, y \rangle \in a^G$ implies $\langle f(x), f(y) \rangle \in a^H$, $f(1^G) = 1^H$, and $f(2^G) = 2^H$. We say that f is a *graph isomorphism* from G to H if f is bijective and for all x, y, and a, $\langle x, y \rangle \in a^G$ iff $\langle f(x), f(y) \rangle \in a^H$, $f(1^G) = 1^H$, and $h(2^G) = 2^H$. We say that H is an *edge-extension* of G if $|H| = |G|$, $a^H \supseteq a^G$ for all a, $1^H = 1^G$, and $2^H = 2^G$. For a set $\{1^G, 2^G\} \subseteq X \subseteq |G|$, the *induced subgraph* of G on X is the graph $\langle X, \{a^G \cap (X \times X)\}_{a \in A}, 1^G, 2^G \rangle$. For an equivalence relation E, the *quotient graph* of G w.r.t. E is the graph $G/E \triangleq \langle |G|/E, \{\langle X, Y \rangle \mid \exists x \in X, y \in Y, \langle x, y \rangle \in a^G\}_{a \in A}, [1^G]_E, [2^G]_E \rangle$ where X/E denotes the set of equivalence classes of X by E and $[x]_E$ denotes the equivalence class of x. Additionally, we use the following operation:

▶ **Definition 2.1.** *For a graph homomorphism $h \colon G \longrightarrow H$ (G, H over a set A), the* edge-saturation *of G w.r.t. h is the graph* $\mathcal{S}(h) \triangleq \langle |G|, \{\{\langle x, y \rangle \mid \langle h(x), h(y) \rangle \in a^H\}\}_{a \in A}, 1^G, 2^G \rangle$.

▶ **Example 2.2.** Let $h \colon G \longrightarrow H$ be the graph homomorphism indicated by green colored arrows and also by vertices' colors (graphs are written as unlabelled graphs for simplicity).

Then the edge-saturation $\mathcal{S}(h)$ is the following graph, which is an edge-extension of G where the extended edges are derived from edges of H:

$\mathcal{S}(h) =$  $G =$  $H =$

2.1 Syntax: terms of KA with complement

We consider *terms* over the signature $S \triangleq \{\mathsf{I}_{(0)}, \bot_{(0)}, \cdot_{(2)}, \cup_{(2)}, \_^{*}{}_{(1)}, \_^{-}{}_{(1)}\}$. Let $\mathbf{V}$ be a set of *variables*. For a term t over S, let $\bar{t}$ be s if $t = s^-$ for some s and be t^- otherwise. We use the abbreviations:

$$\top \triangleq \bot^- \qquad\qquad t \cap s \triangleq (t^- \cup s^-)^-.$$

For $X \subseteq \{\overline{x}, \bar{\mathsf{I}}, \top, \cap, -\}$, let KA_X be the minimal subset A of the set of terms over S satisfying the following:

$$\frac{y \in \mathbf{V}}{y \in A} \qquad \frac{}{\mathsf{I} \in A} \qquad \frac{}{\bot \in A} \qquad \frac{t \in A \quad s \in A}{t \cdot s \in A} \qquad \frac{t \in A \quad s \in A}{t \cup s \in A} \qquad \frac{t \in A}{t^* \in A}$$

$$\frac{\overline{x} \in X \quad y \in \mathbf{V}}{\overline{y} \in A} \qquad \frac{\bar{\mathsf{I}} \in X}{\bar{\mathsf{I}} \in A} \qquad \frac{\top \in X}{\top \in A} \qquad \frac{\cap \in X \quad t \in A \quad s \in A}{t \cap s \in A} \qquad \frac{- \in X \quad t \in A}{t^- \in A}.$$

We use parentheses in ambiguous situations. We often abbreviate $t \cdot s$ to ts.

An *equation* $t = s$ is a pair of terms. An *inequation* $t \leq s$ abbreviates the equation $t \cup s = s$. The set of *quantifier-free formulas* of KA_X is defined by the following grammar:

$$\varphi, \psi \quad ::= \quad t = s \mid \varphi \wedge \psi \mid \neg\varphi. \qquad\qquad (t, s \in \mathrm{KA}_X)$$

We use the abbreviations: $\varphi \vee \psi \triangleq \neg(\neg\varphi \wedge \neg\psi)$, $\varphi \to \psi \triangleq \neg\varphi \vee \psi$, $\varphi \leftrightarrow \psi \triangleq (\varphi \to \psi) \wedge (\psi \to \varphi)$.

2.2 Semantics: language models

An *S-algebra* $\mathcal{A}$ is a tuple $\langle |\mathcal{A}|, \{f^{\mathcal{A}}\}_{f_{(k)} \in S} \rangle$, where $|\mathcal{A}|$ is a non-empty set and $f^{\mathcal{A}} \colon |\mathcal{A}|^k \to |\mathcal{A}|$ is a k-ary map for each $f_{(k)} \in S$. A *valuation* $\mathfrak{v}$ of an S-algebra $\mathcal{A}$ is a map $\mathfrak{v} \colon \mathbf{V} \to |\mathcal{A}|$. For a valuation $\mathfrak{v}$, we write $\hat{\mathfrak{v}} \colon \mathrm{KA}_{\{-\}} \to |\mathcal{A}|$ for the unique homomorphism extending $\mathfrak{v}$. Moreover, for a quantifier-free formula φ, we define $\hat{\mathfrak{v}}(\varphi) \in \{\mathsf{true}, \mathsf{false}\}$ by:

$$\hat{\mathfrak{v}}(t = s) \Leftrightarrow (\hat{\mathfrak{v}}(t) = \hat{\mathfrak{v}}(s)) \qquad \hat{\mathfrak{v}}(\varphi \wedge \psi) \Leftrightarrow (\hat{\mathfrak{v}}(\varphi) \text{ and } \hat{\mathfrak{v}}(\psi)) \qquad \hat{\mathfrak{v}}(\neg\varphi) \Leftrightarrow (\text{not } \hat{\mathfrak{v}}(\varphi)).$$

For a quantifier-free formula φ and a class of valuations (of S-algebra) $\mathcal{C}$,[3] we write

$$\mathcal{C} \models \varphi \quad \Leftrightarrow \quad \hat{\mathfrak{v}}(\varphi) \text{ holds for all valuations } \mathfrak{v} \in \mathcal{C}.$$

We abbreviate $\{\mathfrak{v}\} \models \varphi$ to $\mathfrak{v} \models \varphi$. The *equational theory of $\mathcal{C}$* is the set of all equations $t = s$ such that $\mathcal{C} \models t = s$. The *quantifier-free theory of $\mathcal{C}$* is the set of all quantifier-free formulas φ such that $\mathcal{C} \models \varphi$.

The *language model* $\mathcal{A}$ over a set X, written lang_X, is the S-algebra defined by $|\mathcal{A}| = \wp(X^*)$, $\mathsf{I}^{\mathcal{A}} = \{\mathsf{I}\}$, $\bot^{\mathcal{A}} = \emptyset$, and for all $L, K \subseteq X^*$,

$$L \cdot^{\mathcal{A}} K = L \cdot K \qquad L \cup^{\mathcal{A}} K = L \cup K \qquad L^{*^{\mathcal{A}}} = L^* \qquad L^{-^{\mathcal{A}}} = X^* \setminus L.$$

[3] This paper considers classes of valuations rather than classes of S-algebras (cf. Remark 3.12).

We write LANG_X for the class of all valuations of lang_X and write LANG for $\bigcup_X \mathsf{LANG}_X$. The *equational theory (resp. quantifier-free theory) w.r.t. languages* means that of LANG.

The *language* $[t] \subseteq \mathbf{V}^*$ of a term t is $\hat{\mathfrak{v}}_{\mathrm{st}}(t)$ where $\mathfrak{v}_{\mathrm{st}}$ is the valuation on the language model over the set $\mathbf{V}$ defined by $\mathfrak{v}_{\mathrm{st}}(x) = \{x\}$ for $x \in \mathbf{V}$. Since $\mathfrak{v}_{\mathrm{st}} \in \mathsf{LANG}$, we have

$$\mathsf{LANG} \models t = s \quad \Rightarrow \quad [t] = [s] \tag{$\dagger$-KA$_{\{-\}}$}$$

The converse direction fails, as in the introduction. In the sequel, we consider the equational theory w.r.t. languages.

2.3 (Generalized) relational models

We write $\triangle_A$ for the identity relation on a set A: $\triangle_A \triangleq \{\langle x, x\rangle \mid x \in A\}$. For binary relations R, S on a set B, the *composition* $R \cdot S$, the *n-th iteration* R^n (where $n \in \mathbb{N}$), and the *reflexive transitive closure* R^* are defined by:

$$R \cdot S \triangleq \{\langle x, z\rangle \mid \exists y, \langle x, y\rangle \in R \wedge \langle y, z\rangle \in S\} \quad R^n \triangleq \begin{cases} R \cdot R^{n-1} (n \geq 1) \\ \triangle_B \qquad (n = 0) \end{cases} \quad R^* \triangleq \bigcup_{n \in \mathbb{N}} R^n.$$

Let U be a binary relation on a non-empty set B. A *generalized relational model*[4] $\mathcal{A}$ on U is an S-algebra such that $|\mathcal{A}| \subseteq \wp(U)$, $\mathsf{I}^{\mathcal{A}} = \triangle_B$, $\bot^{\mathcal{A}} = \emptyset$, and for all $R, S \subseteq U$,

$$R \cdot^{\mathcal{A}} S = R \cdot S \qquad R \cup^{\mathcal{A}} S = R \cup S \qquad R^{*^{\mathcal{A}}} = R^* \qquad R^{-^{\mathcal{A}}} = U \setminus R.$$

$\mathcal{A}$ is a *relational model* if $U = B \times B$ and $|\mathcal{A}| = \wp(B \times B)$. We write GREL (resp. REL) for the class of all valuations of generalized relational models (resp. relational models).[5]

3 Finite relational models for language models

In this section, we present the finite relational semantics RSUB for LANG. Sect. 3.1 introduces the class SUB, which is intuitively an intermediate class between LANG and RSUB, and show that the quantifier-free theory of LANG coincides with that of SUB. Sect. 3.2 introduces RSUB and show that the quantifier-free theory of LANG coincides with that of RSUB.

3.1 SUB: subword models

For the standard language valuation ($\mathfrak{v}_{\mathrm{st}}$), the membership $w \in \hat{\mathfrak{v}}_{\mathrm{st}}(t)$ can determined from $w' \in \hat{\mathfrak{v}}_{\mathrm{st}}(t')$ where w' ranges over *subwords* of w (and t' ranges over subterms of t). This situation is the same also for any valuations $\mathfrak{v} \in \mathsf{LANG}$, To reflect this property to LANG, we define the following class: subword models SUB. These models are language models where the universe is bounded to the set of subwords of a word $\ell_0 \dots \ell_{n-1}$.

▶ **Definition 3.1.** *Let $n \in \mathbb{N}$ and let $\ell_0, \dots, \ell_{n-1}$ be pairwise distinct letters. Let $U = \{\ell_i \dots \ell_{j-1} \mid 0 \leq i \leq j \leq n\}$. The* subword language model *$\mathcal{A}$ of length n, written SUB_n, is the S-algebra defined by: $|\mathcal{A}| = \wp(U)$, $\mathsf{I}^{\mathcal{A}} = \{\mathsf{I}\}$, $\bot^{\mathcal{A}} = \emptyset$, and for all $L, K \subseteq U$,*

$$L \cdot^{\mathcal{A}} K = (L \cdot K) \cap U \qquad L \cup^{\mathcal{A}} K = L \cup K \qquad L^{*^{\mathcal{A}}} = L^* \cap U \qquad L^{-^{\mathcal{A}}} = U \setminus L. \qquad \lrcorner$$

[4] By definition, U is a preorder. Reflexivity: By $\triangle_B = \mathsf{I}^{\mathcal{A}} \in |\mathcal{A}| \subseteq \wp(U)$, we have $\triangle_B \subseteq U$. Transitivity: By $\emptyset = \bot^{\mathcal{A}} \in |\mathcal{A}|$, $U = \emptyset^{-^{\mathcal{A}}} \in |\mathcal{A}|$, and $U \cdot U = U \cdot^{\mathcal{A}} U \in |\mathcal{A}| \subseteq \wp(U)$, we have $U \cdot U \subseteq U$.

[5] Generalized relational models and relational models are variants of proper relation algebras and full proper relation algebras (see, e.g., [23]), respectively, where B is non-empty set and the converse operator is not introduced (due to this, U is possibly not symmetric, cf. [23, Lem. 3.4]) here.

We write SUB_n for the class of all valuations of sub_n and write SUB for $\bigcup_{n\in\mathbb{N}} \mathsf{SUB}_n$.

We use the following transformation of valuations from LANG into SUB.

▶ **Definition 3.2** (*letters-to-letters valuations*; cf. words-to-letters valuations [19]). *Let X be a set and let $\mathfrak{v} \in \mathsf{LANG}_X$. Let $n \in \mathbb{N}$ and let $a_0, \ldots, a_{n-1}$ be (possibly not pairwise distinct) letters over X. Then we define $\mathfrak{v}^{\langle a_0,\ldots,a_{n-1}\rangle}$ as the following valuation in SUB_n:*

$$\mathfrak{v}^{\langle a_0,\ldots,a_{n-1}\rangle}(x) \quad \triangleq \quad \{\ell_i \ldots \ell_{j-1} \mid 0 \le i \le j \le n \ \wedge\ a_i \ldots a_{j-1} \in \mathfrak{v}(x)\}. \qquad \lrcorner$$

▶ **Example 3.3.** If $\mathfrak{v}(x) = \{aba,\ a,\ bb\}$, then $\mathfrak{v}^{\langle a,b,a\rangle}(x) = \{\ell_0\ell_1\ell_2,\ \ell_0, \ell_2\}$. Note that $\mathfrak{v}^{\langle a,b,a\rangle}$ only reflects w.r.t. subwords of aba.

Using this, we can embed the membership in LANG into that in SUB.

▶ **Lemma 3.4.** *Let X be a set and let $\mathfrak{v} \in \mathsf{LANG}_X$. Let $n \in \mathbb{N}$ and let $a_0, \ldots, a_{n-1}$ be letters over X. For all terms t and $0 \le i \le j \le n$, $\ell_i \ldots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle a_0,\ldots,a_{n-1}\rangle}(t)$ iff $a_i \ldots a_{j-1} \in \hat{\mathfrak{v}}(t)$.*

Proof. By easy induction on t, cf. [19, Lem. 23] (see Appendix D, for detailed proof).[6] ◀

Moreover, this transformation is a surjective map to SUB, as follows:

▶ **Proposition 3.5.** *Let $n \in \mathbb{N}$ and let $\mathfrak{v}' \in \mathsf{SUB}_n$. Let $\mathfrak{v} \in \mathsf{LANG}_{\{\ell_0,\ldots,\ell_{n-1}\}}$ be the valuation defined by $\mathfrak{v}(x) = \mathfrak{v}'(x)$. Then $\mathfrak{v}^{\langle \ell_0,\ldots,\ell_{n-1}\rangle} = \mathfrak{v}'$.*

Proof. By Def. 3.2, $\ell_i \ldots \ell_{j-1} \in \mathfrak{v}^{\langle \ell_0,\ldots,\ell_{n-1}\rangle}(x)$ iff $\ell_i \ldots \ell_{j-1} \in \mathfrak{v}(x)$ iff $\ell_i \ldots \ell_{j-1} \in \mathfrak{v}'(x)$. ◀

From them, we have that the quantifier-free theory of LANG coincides with that of SUB.

▶ **Lemma 3.6.** *For all quantifier-free formulas φ, we have: $\mathsf{LANG} \models \varphi \ \Leftrightarrow\ \mathsf{SUB} \models \varphi$.*

Proof. By $\mathfrak{v} \models t = s \leftrightarrow (t \le s \wedge s \le t)$ and $\mathfrak{v} \models t \le s \leftrightarrow t \cap s^- \le \bot$ for $\mathfrak{v} \in \mathsf{LANG} \cup \mathsf{SUB}$, without loss of generality, we can assume that each equation in φ is of the form $u \le \bot$. By taking the conjunctive normal form, it suffices to show when φ is of the form $(\bigvee_k t_k \le \bot) \vee (\bigvee_k \neg s_k \le \bot)$. By $\mathfrak{v} \models (t \le \bot \wedge s \le \bot) \leftrightarrow t \cup s \le \bot$ for $\mathfrak{v} \in \mathsf{LANG} \cup \mathsf{SUB}$, it suffices to show when φ is of the form $(\bigvee_{k=1}^m t_k \le \bot) \vee (\neg s \le \bot)$. Then, we have:

$$\mathsf{SUB} \not\models (\bigvee_{k=1}^{m} t_k \le \bot) \vee (\neg s \le \bot)$$

$$\Rightarrow \quad \exists n, \mathfrak{v} \in \mathsf{LANG}_{\{\ell_0,\ldots,\ell_{n-1}\}},\ \mathfrak{v}^{\langle \ell_0,\ldots,\ell_{n-1}\rangle} \not\models (\bigvee_{k=1}^{m} t_k \le \bot) \vee (\neg s \le \bot) \qquad \text{(Prop. 3.5)}$$

$$\Rightarrow \quad \mathsf{LANG} \not\models (\bigvee_{k=1}^{m} t_k \le \bot) \vee (\neg s \le \bot) \qquad \text{(By Lem. 3.4 and } \mathfrak{v} \in \mathsf{LANG})$$

$$\Rightarrow \quad \exists n \in \mathbb{N}, i_1, j_1, \ldots, i_m, j_m \text{ s.t. } 0 \le i_1 \le j_1 \le i_2 \le j_2 \le \cdots \le i_m \le j_m \le n,$$

$$\exists X, \mathfrak{v} \in \mathsf{LANG}_X, a_0, \ldots, a_{n-1} \in X,\ (\bigwedge_{k=1}^{m} a_{i_k} \ldots a_{j_k-1} \in \hat{\mathfrak{v}}(t_k)) \wedge \hat{\mathfrak{v}}(s) = \emptyset$$

$$\Rightarrow \quad \mathsf{SUB} \not\models (\bigvee_{k=1}^{m} t_k \le \bot) \vee (\neg s \le \bot). \qquad \text{(By Lem. 3.4 and } \mathfrak{v}^{\langle a_0,\ldots,a_{n-1}\rangle} \in \mathsf{SUB}_n) \quad ◀$$

[6] This equivalence holds thanks to that $a_i, \ldots, a_{j-1}$ are letters not words. For words-to-letters valuations, the direction $\Rightarrow$ of Lem. 3.4 only holds [19, Lem. 23].

3.2 RSUB: relational subword models

Next, we introduce RSUB, the relational version of SUB.

▶ **Definition 3.7.** *Let $n \in \mathbb{N}$. The relational subword language model $\mathcal{A}$ of length n, written rsub_n, is the generalized relational model on the set $U = \{\langle i, j\rangle \in [0,n] \times [0,n] \mid i \leq j\}$ s.t.*

$$|\mathcal{A}| \quad = \quad \{R \in \wp(U) \mid R \supseteq \triangle_{[0,n]} \ \vee \ U \setminus R \supseteq \triangle_{[0,n]}\}.$$

We write RSUB_n for the class of all valuations of rsub_n and write RSUB for $\bigcup_{n\in\mathbb{N}} \mathsf{RSUB}_n$.

▶ **Proposition 3.8.** *For each $n \in \mathbb{N}$, sub_n is isomorphic to rsub_n.*

Proof. By the bijective[7] map $f\colon L \mapsto \{\langle i, j\rangle \mid \ell_i \dots \ell_{j-1} \in L\}$.[8] Fig. 1 presents illustrative

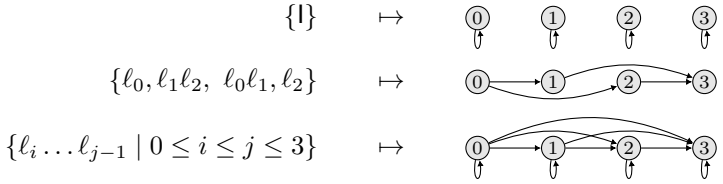


Figure 1 Illustrative instances of the bijective map from sub_n to rsub_n when $n = 3$ (Prop. 3.8).

examples of this map (we display binary relations as graphs in a standard way). ◀

▶ **Theorem 3.9.** *For all quantifier-free formulas φ, we have:* $\mathsf{LANG} \models \varphi \Leftrightarrow \mathsf{RSUB} \models \varphi$.

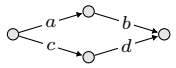
Proof. By Lem. 3.6 and Prop. 3.8. ◀

▶ **Corollary 3.10.** *The quantifier-free theory w.r.t. languages is in Π^0_1 for $\mathrm{KA}_{\{-\}}$ terms.*

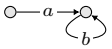
Proof. By the finite model property of RSUB (the universe $|\mathsf{rsub}_n|$ is finite for each n). ◀

▶ Remark 3.11 (RSUB and GREL). For $\mathrm{KA}_{\{\top\}}$, the equational theory of LANG coincides with that of GREL [21, REL′ in Sect. 5][28]. However for $\mathrm{KA}_{\{-\}}$, this coincidence is broken, e.g.,[9]

$\mathsf{LANG} \models ab \cap cd \leq a\top d \cup b\top c$ $\quad$ $\mathsf{GREL} \not\models ab \cap cd \leq a\top d \cup b\top c$



$\mathsf{LANG} \models a \leq \bar{b}a\bar{b} \cup bab$ $\quad$ $\mathsf{(G)REL} \not\models a \leq \bar{b}a\bar{b} \cup bab$



(Each right figure denotes a valuation for (G)REL $\not\models$ _ where some edges are omitted.) ⌟

▶ Remark 3.12 (the subclass of RSUB for the standard language valuation). Let

$$\mathsf{RSUB}_{\mathrm{st}} \triangleq \bigcup_{n\in\mathbb{N}} \left\{ \mathfrak{v} \in \mathsf{RSUB}_n \;\middle|\; \begin{array}{c} \bigcup_{a\in\mathbf{V}} \mathfrak{v}(a) = \{\langle i-1, i\rangle \mid i \in [1,n]\} \\ \mathfrak{v}(a) \text{ (where } a \text{ ranges over } \mathbf{V}) \text{ are disjoint sets} \end{array} \right\}.$$

By an analogy of Thm. 3.9, we have that $[t] = [s]$ iff $\mathsf{RSUB}_{\mathrm{st}} \models t = s$ (Appendix A). ⌟

---

[7] $f(\{\mathsf{I}\}) = \triangle_{[0,n]}$ (cf. Def. 3.7) and $f(\{\ell_i \dots \ell_{j-1}\}) = \{\langle i, j\rangle\}$ is singleton if $i < j$.

[8] This map is the same as Pratt's embedding [22, third page], cf. [28, 21]: $L \mapsto \{\langle w, wv\rangle \mid w \in X^* \wedge v \in L\}$, up to isomorphism, except that the universe is bounded to subwords of $\ell_0 \dots \ell_{n-1}$.

[9] $\mathsf{LANG} \models ab \cap cd \leq a\top d \cup b\top c$: By Levi's inequation [15][5, Example 26]. $\mathsf{LANG} \models a \leq \bar{b}a\bar{b} \cup bab$: for every valuation $\mathfrak{v}$ of LANG, $\mathfrak{v} \models a \leq bab$ if $\mathsf{I} \in \mathfrak{v}(b)$ and $\mathfrak{v} \models a \leq \bar{b}a\bar{b}$ if $\mathsf{I} \notin \mathfrak{v}(b)$.

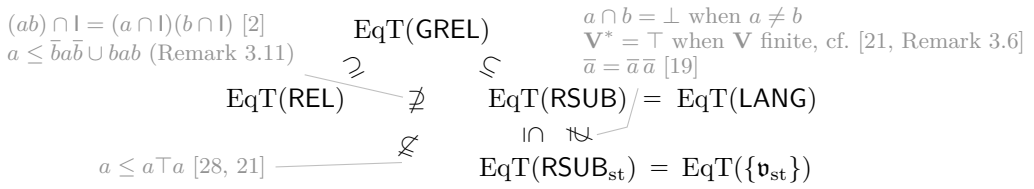


Figure 2 Equational theories for $\mathrm{KA}_{\{-\}}$ under GREL.

In Fig. 2, we summarize the equational theories for $\mathrm{KA}_{\{-\}}$ terms above where the inclusions are shown by $\mathsf{REL} \subseteq \mathsf{GREL} \supseteq \mathsf{RSUB} \supseteq \mathsf{RSUB}_{\mathrm{st}}$ and Thm. 3.9 and the non-inclusions are shown by counter-examples.[10]

4 From quantifier-free formulas to equations

Using SUB/RSUB, we give a reduction from the quantifier-free theory into the equational theory of LANG. For $n \in \mathbb{N}$, $\mathfrak{v} \in \mathsf{SUB}_n$, and i, j s.t. $0 \leq i \leq j \leq n$, we define $\mathfrak{v}, i, j \models \varphi$ as

$$\mathfrak{v}, i, j \models t = s \triangleq (\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}(t) \text{ iff } \ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}(s))$$

$$\mathfrak{v}, i, j \models \varphi \wedge \psi \triangleq (\mathfrak{v}, i, j \models \varphi \text{ and } \mathfrak{v}, i, j \models \psi) \qquad \mathfrak{v}, i, j \models \neg\varphi \triangleq (\text{not } \mathfrak{v}, i, j \models \varphi).$$

The following shows that to check $\mathsf{SUB} \models t = s$, it suffices to check about the pairs of the left-most and right-most vertices.

▶ **Lemma 4.1.** *For all terms* t, s, $\mathsf{SUB} \models t = s \Leftrightarrow \forall n \in \mathbb{N}, \forall \mathfrak{v} \in \mathsf{SUB}_n, \mathfrak{v}, 0, n \models t = s.$

Proof. Because $t = s \leftrightarrow (t \leq s \wedge s \leq t)$ and $t \leq s \leftrightarrow t \cap s^- \leq \bot$ hold for "$\mathsf{RSUB} \models \_$" and "$\mathfrak{v}, 0, n \models \_$", it suffices to show when the equation is of the form $t \leq \bot$. We then have

$$\begin{aligned}
&\mathsf{SUB} \not\models t \leq \bot \\
&\Rightarrow \quad \exists \mathfrak{v} \in \mathsf{LANG}, n \in \mathbb{N}, i, j \text{ s.t. } 0 \leq i \leq j \leq n,\ \ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle \ell_0, \dots, \ell_{n-1} \rangle}(t) && \text{(Prop. 3.5)} \\
&\Rightarrow \quad \exists \mathfrak{v} \in \mathsf{LANG}, n \in \mathbb{N}, i, j \text{ s.t. } 0 \leq i \leq j \leq n,\ \ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}(t) && \text{(Lem. 3.4)} \\
&\Rightarrow \quad \exists \mathfrak{v} \in \mathsf{LANG}, n \in \mathbb{N}, i, j \text{ s.t. } 0 \leq i \leq j \leq n,\ \ell_0 \dots \ell_{j-i-1} \in \hat{\mathfrak{v}}^{\langle \ell_i, \dots, \ell_{j-1} \rangle}(t) && \text{(Lem. 3.4)} \\
&\Rightarrow \quad \exists n \in \mathbb{N}, \mathfrak{v} \in \mathsf{SUB}_n,\ \ell_0 \dots \ell_{n-1} \in \hat{\mathfrak{v}}(t) && \text{(Let } n \triangleq j - i \text{ as } \mathfrak{v}^{\langle \ell_i, \dots, \ell_{j-1} \rangle} \in \mathsf{SUB}_{j-i}) \\
&\Rightarrow \quad \mathsf{SUB} \not\models t \leq \bot. && \text{(Trivial)}
\end{aligned}$$

(The second line from the bottom is equivalent to $\exists n \in \mathbb{N}, \mathfrak{v} \in \mathsf{SUB}_n, \mathfrak{v}, 0, n \not\models t \leq \bot$.) ◀

Lem. 4.1 fails for quantifier-free formulas, e.g., $a \leq \bot \vee \bar{a} \leq \bot$ only holds in the right-hand side. We consider replacing each inequation $u \leq \bot$ with $\top u \top \leq \bot$. Then we have that $\mathfrak{v} \models u \leq \bot$ iff $\mathfrak{v}, 0, n \models \top u \top \leq \bot$, where $n \in \mathbb{N}$ and $\mathfrak{v} \in \mathsf{SUB}_n$ (see Lem. 4.2). More generally, for a quantifier-free formula φ, let $\mathrm{Tr}(\varphi)$ be the term defined by:[11]

$$\mathrm{Tr}(t = s) \triangleq \top((t \cap s^-) \cup (t^- \cap s))\top \quad \mathrm{Tr}(\varphi \wedge \psi) \triangleq \mathrm{Tr}(\varphi) \cup \mathrm{Tr}(\psi) \quad \mathrm{Tr}(\neg\varphi) \triangleq \mathrm{Tr}(\varphi)^-.$$

[10] Note that EqT(LANG) = EqT(GREL) for $\mathrm{KA}_{\{\top\}}$ and EqT($\{\mathfrak{v}_{\mathrm{st}}\}$) = EqT(GREL) for KA [28, 21].
[11] $\mathrm{Tr}(t = s)$ can be simplfied for specific cases, e.g., $\mathrm{Tr}(t \leq s) = \top(t \cap s^-)\top$ and $\mathrm{Tr}(t \leq \bot) = \top t \top$.

▶ **Lemma 4.2.** *Let $n \in \mathbb{N}$ and $\mathfrak{v} \in \mathsf{SUB}_n$. Let φ be a quantifier-free formula. Then,*

$$\mathfrak{v} \models \varphi \quad \Leftrightarrow \quad \mathfrak{v}, 0, n \models \mathrm{Tr}(\varphi) \leq \bot.$$

Proof. By easy induction on φ. Case $\varphi = (t = s)$: Let $u = (t \cap s^-) \cup (t^- \cap s)$. Then $\mathfrak{v} \models t = s$ iff $\hat{\mathfrak{v}}(u) = \emptyset$ iff $\ell_0 \dots \ell_{n-1} \notin \hat{\mathfrak{v}}(\top u \top)$ iff $\mathfrak{v}, 0, n \models \top u \top \leq \bot$ iff $\mathfrak{v}, 0, n \models \mathrm{Tr}(t = s)$. For $\wedge$: By $\mathfrak{v}, 0, n \models ((t \leq \bot) \wedge (s \leq \bot)) \leftrightarrow (t \cup s \leq \bot)$. For $\neg$: By $\mathfrak{v}, 0, n \models (\neg(t \leq \bot)) \leftrightarrow (t^- \leq \bot)$. ◀

▶ **Theorem 4.3** (cf. Schröder-Tarski translation [27])**.** *There is a polynomial-time reduction from the quantifier-free theory of* LANG *into the equational theory of* LANG*.*

Proof. By Lems. 4.1 and 4.2, $\mathsf{SUB} \models \varphi$ iff $\mathsf{SUB} \models \mathrm{Tr}(\varphi) \leq \bot$. Thus by Thm. 3.9. ◀

▶ **Corollary 4.4** (Hoare hypotheses elimination)**.** *For all terms t, s, u,*

$$\mathsf{LANG} \models u \leq \bot \to t \leq s \quad \Leftrightarrow \quad \mathsf{LANG} \models t \leq s \cup \top u \top.$$

Proof. By set-theoretic equivalences, we have

$$\begin{aligned}
\mathsf{SUB} \models u \leq \bot \to t \leq s \quad &\Leftrightarrow \quad \mathsf{SUB} \models \top(t \cap s^-)\top \leq \top u \top && \text{(By Thm. 4.3)} \\
&\Leftrightarrow \quad \mathsf{SUB} \models t \cap s^- \leq \top u \top && (\Rightarrow\text{: By } \mathsf{I} \leq \top \Leftarrow\text{: By } \top\top \leq \top) \\
&\Leftrightarrow \quad \mathsf{SUB} \models t \leq s \cup \top u \top && \blacktriangleleft
\end{aligned}$$

▶ Remark 4.5. Note that $\mathsf{LANG} \not\models (\neg u \leq \bot) \leftrightarrow \top \leq \top u \top$, cf. Schröder-Tarski translation [27, XXXII.] for REL. If $\hat{\mathfrak{v}}(u) = \{w\}$ where $\mathfrak{v} \in \mathsf{LANG}_X$ and $w \neq \mathsf{I}$, then $\hat{\mathfrak{v}}(\top) = [X^*] \not\subseteq [X^* w X^*] = \hat{\mathfrak{v}}(\top u \top)$. This is why we go via "$\mathfrak{v}, 0, n \models \_$" (Lems. 4.1 and 4.2). Similarly for Cor. 4.4, $\mathsf{LANG} \not\models (u \leq \bot \to t \leq s) \leftrightarrow (t \leq s \cup \top u \top)$, e.g., when $t = \top$ and $s = \bot$. ⌟

4.1 Undecidability via the Hoare hypotheses elimination

Thm. 4.3/Cor. 4.4 implies the undecidability of the equational theory of LANG for full terms.

A *context-free grammar* (CFG) $\mathfrak{C}$ over a finite set A is a tuple $\langle X, \mathcal{R}, \mathsf{s} \rangle$, where

- X is a finite set of non-terminal labels s.t. $A \cap X = \emptyset$;
- $\mathcal{R}$ is a finite set of rewriting rules $x \leftarrow w$ of $x \in X$ and $w \in (A \cup X)^*$;
- $\mathsf{s} \in X$ is the start label.

We write $[\mathfrak{C}]$ for the *language* of $\mathfrak{C}$. It is well-known that the *universality problem* for CFGs—given a CFG $\mathfrak{C}$, does $[\mathfrak{C}] = A^*$ hold?—is Π^0_1-complete. We can naturally encode this problem by the quantifier-free theory of LANG, as follows (cf. [18, Lem. 47]).

▶ **Lemma 4.6.** *Let $\mathfrak{C} = \langle X, \mathcal{R}, \mathsf{s} \rangle$ be a CFG over a finite set A. Then,*

$$[\mathfrak{C}] = A^* \quad \Leftrightarrow \quad \mathsf{LANG} \models \Big(\bigwedge_{(x \leftarrow w) \in \mathcal{R}} w \leq x \Big) \to (A^* \leq \mathsf{s}).$$

Here, A denotes the term $(a_1 \cup \dots \cup a_n)$ in the right-hand side where $A = \{a_1, \dots, a_n\}$.

Proof Sketch. By an analogy of [18, Lem. 47] (for REL). The relational models used in [18] are of the form of RSUB. See Appendix E for a detail. ◀

▶ **Theorem 4.7.** *The equational theory w.r.t. languages is Π^0_1-complete for $\mathrm{KA}_{\{\overline{x}, \cap\}}$.*

Proof. (in Π^0_1): By Cor. 3.10. (Hardness): Let $\mathfrak{C} = \langle X, \{x_i \leftarrow w_i \mid i \in [1, n]\}, \mathsf{s} \rangle$ be a CFG over a finite set A. By Lem. 4.6 and by Cor. 4.4 with $\mathsf{LANG} \models (\bigwedge_{i \in [1,n]} w_i \leq x_i) \leftrightarrow (\bigcup_{i \in [1,n]} w_i \cap \overline{x}_i) \leq \bot$, we have: $[\mathfrak{C}] = A^*$ iff $\mathsf{LANG} \models A^* \leq \mathsf{s} \cup \top (\bigcup_{i \in [1,n]} w_i \cap \overline{x}_i) \top$. Hence, we can give a reduction from the universality problem of CFGs. ◀

▶ **Corollary 4.8.** *The equational theory w.r.t. languages is* Π^0_1*-complete for* $\mathrm{KA}_{\{-\}}$.

Proof. By Cor. 3.10 and Thm. 4.7. ◀

Moreover, we also have the following, as $\mathsf{LANG} \models \bar{\mathsf{I}} = A\top \to A^* = \top$ holds (see Appendix E).

▶ **Lemma 4.9.** *Let* $\mathfrak{C} = \langle X, \mathcal{R}, \mathsf{s}\rangle$ *be a CFG over a finite set* A*. Then,*

$$[\mathfrak{C}] = A^* \quad \Leftrightarrow \quad \mathsf{LANG} \models (\bar{\mathsf{I}} = A\top \wedge \bigwedge_{(x \leftarrow w) \in \mathcal{R}} w \le x) \to (\top \le \mathsf{s}).$$

Hence, the undecidability still holds even without Kleene star.

▶ **Theorem 4.10.** *The equational theory w.r.t. languages is* Π^0_1*-complete for* $\mathrm{KA}_{\{-\}}$ *without Kleene star.*

Proof. In the same way as Thm. 4.7 (by using Lem. 4.9 instead of Lem. 4.6). ◀

5 Graph characterization for $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{I}},\top,\cap\}}$ terms

In Sects. 5 and 6, we consider the decidability of $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{I}},\top\}}$. We recall Sect. 2 for graphs. In this section, we give a graph characterization of the equational theory of RSUB for $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{I}},\top,\cap\}}$, by generalizing the graph characterization of REL [18, Thm. 18].

5.1 Graph languages for $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{I}},\top,\cap\}}$

Let $\tilde{\mathbf{V}} \triangleq \{x, \overline{x} \mid x \in \mathbf{V}\} \cup \{\bar{\mathsf{I}}, \top\}$ and $\tilde{\mathbf{V}}_{\mathsf{I}} \triangleq \tilde{\mathbf{V}} \cup \{\mathsf{I}\}$. For a $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{I}},\top,\cap\}}$ term t, the graph language $\mathcal{G}(t)$ is a set of graphs over $\tilde{\mathbf{V}}_{\mathsf{I}}$ defined by:[12]

$$\mathcal{G}(x) \triangleq \{ \ \bullet\!\!-\!\!\overset{x}{\rightarrow}\!\!\bullet \ \} \text{ where } x \in \tilde{\mathbf{V}} \qquad \mathcal{G}(\bot) \triangleq \emptyset \qquad \mathcal{G}(\mathsf{I}) \triangleq \{ \ \bullet\!\!\rightarrow\!\!\bullet \ \}$$

$$\mathcal{G}(t \cap s) \triangleq \{ \ \bullet \overset{G}{\underset{H}{\rightrightarrows}} \bullet \ \mid G \in \mathcal{G}(t) \wedge H \in \mathcal{G}(s)\} \qquad \mathcal{G}(t \cup s) \triangleq \mathcal{G}(t) \cup \mathcal{G}(s)$$

$$\mathcal{G}(t \cdot s) \triangleq \{ \ \bullet\!\!-\!\!G\!\!\rightarrow\!\!\bullet\!\!-\!\!H\!\!\rightarrow\!\!\bullet \ \mid G \in \mathcal{G}(t) \wedge H \in \mathcal{G}(s)\} \qquad \mathcal{G}(t^*) \triangleq \bigcup_{n \in \mathbb{N}} \mathcal{G}(t^n).$$

(We use *series-composition* for $(\cdot)$ and *parallel-composition* for $(\cap)$.)

For a valuation $\mathfrak{v} \in \mathsf{GREL}$ on a binary relation on a set B and $\langle x, y\rangle \in \hat{\mathfrak{v}}(\top)$, let $\mathsf{G}(\mathfrak{v}, x, y)$ be the graph defined by:

$$\mathsf{G}(\mathfrak{v}, x, y) \quad \triangleq \quad \langle B, \{\hat{\mathfrak{v}}(a)\}_{a \in \tilde{\mathbf{V}}_{\mathsf{I}}}, x, y\rangle.$$

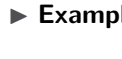
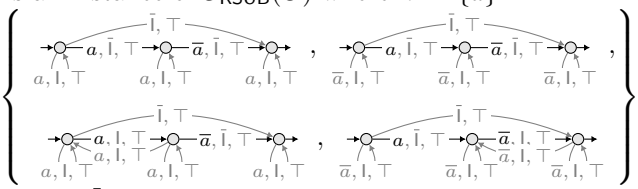
For a class $\mathcal{C} \subseteq \mathsf{GREL}$, let $\mathrm{GR}_{\mathcal{C}}$ be the graph language $\{\mathsf{G}(\mathfrak{v}, x, y) \mid \mathfrak{v} \in \mathcal{C} \text{ and } \langle x, y\rangle \in \hat{\mathfrak{v}}(\top)\}$. We say that a graph language $\mathcal{G}$ is *induced subgraph-closed* if every induced subgraph of every $G \in \mathcal{G}$ is isomorphic to a member of $\mathcal{G}$. We say that a class $\mathcal{C} \subseteq \mathsf{GREL}$ is induced subgraph-closed if $\mathrm{GR}_{\mathcal{C}}$ is induced subgraph-closed. By the form of rsub_n, RSUB is induced subgraph-closed.

We recall edge-saturations $\mathcal{S}(h)$ of Def. 2.1. For a graph G and graph language $\mathcal{G}$, let

$$\mathcal{S}_{\mathcal{C}}(G) \quad \triangleq \quad \{\mathcal{S}(h) \mid \exists H \in \mathrm{GR}_{\mathcal{C}}, h \colon G \longrightarrow H\} \qquad \mathcal{S}_{\mathcal{C}}(\mathcal{G}) \quad \triangleq \quad \bigcup_{H \in \mathcal{G}} \mathcal{S}_{\mathcal{C}}(H).$$

[12] We introduce $\top$-labelled edges, cf. [18, Def. 6], because $\top$ is not fixed to the full relation.

▶ **Example 5.1.** The following is an instance of $\mathcal{S}_{\mathsf{RSUB}}(G)$ where $\mathbf{V} = \{a\}$:

$\mathcal{S}_{\mathsf{RSUB}}($  $) =$ 

(In the sequel, we often omit $\top$- or $\bar{\mathsf{I}}$-labelled edges for simplicity.)

By the form of $\mathrm{GR}_{\mathsf{RSUB}}$, each H, where $H \in \mathcal{S}_{\mathsf{RSUB}}(H')$ for some H', satisfies

- $\top^H$ is a total preorder;
- $a^H \supseteq \mathsf{I}^H$ or $\overline{a}^H \supseteq \mathsf{I}^H$ holds for each $a \in \mathbf{V}$. ⌟

Additionally, let $H^{\mathcal{Q}} \triangleq H/(\mathsf{I}^H)^=$ and $\mathcal{G}^{\mathcal{Q}} \triangleq \{H^{\mathcal{Q}} \mid H \in \mathcal{G}\}$ where $R^=$ denotes the equivalence closure of R. We then have the following graph language characterization. This is an analogy of [18, Theorem 18], but is generalized for including RSUB.

▶ **Theorem 5.2.** *Let $\mathcal{C} \subseteq \mathsf{GREL}$ be induced subgraph-closed. For all $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{I}},\top,\cap\}}$ terms t, s,*

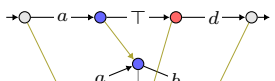
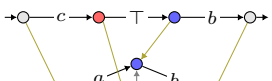
$$\mathcal{C} \models t \leq s \quad \Leftrightarrow \quad \forall H \in \mathcal{S}_{\mathcal{C}}(\mathcal{G}(t))^{\mathcal{Q}}, \exists G \in \mathcal{G}(s), G \longrightarrow H.$$

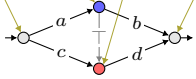
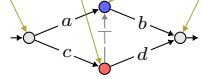
Proof. As with [18, Theorem 18]. See Sect. 5.3 for a detailed proof. ◀

By Thm. 5.2, we can show inequations by graph homomorphisms, e.g., as follows.

▶ **Example 5.3.** We recall the inequations in Remark 3.11. Gray-colored edges are extended by edge-saturations $\mathcal{S}_{\mathsf{RSUB}}$. We omit unimportant edges.

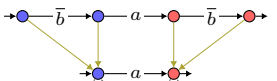
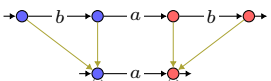
- $\mathsf{LANG} \models ab \cap cd \leq a\top d \cup c\top b$: For each graph $H \in \mathcal{S}_{\mathsf{RSUB}}(\mathcal{G}(ab \cap cd))^{\mathcal{Q}}$, we can give a graph homomorphism from some graph in $\mathcal{G}(a\top d \cup c\top b)$ as follows:

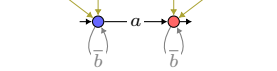
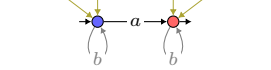
$\mathcal{G}(a\top d \cup c\top b) = \{$  ,  $\}$

$\mathcal{S}_{\mathsf{RSUB}}(\mathcal{G}(ab \cap cd))^{\mathcal{Q}} \ni H$:  

(Case $\langle \bullet, \bullet \rangle \in \top^H$) (Case $\langle \bullet, \bullet \rangle \in \top^H$)

- $\mathsf{LANG} \models a \leq \overline{b}a\overline{b} \cup bab$:

$\mathcal{G}(\overline{b}a\overline{b} \cup bab) = \{$  ,  $\}$

$\mathcal{S}_{\mathsf{RSUB}}(\mathcal{G}(a))^{\mathcal{Q}} \ni H$:  

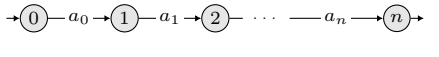
(Case $\overline{b}^H \supseteq \mathsf{I}^H$) (Case $b^H \supseteq \mathsf{I}^H$)

(Additionally, $\_^{\mathcal{Q}}$ is necessary, e.g., for $\top \leq \mathsf{I} \cup \bar{\mathsf{I}}$ [18, Remark 19].) ⌟

5.2 Word languages for $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{I}},\top\}}$

Particularly for $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{I}},\top\}}$, Thm. 5.2 can be rephrased by word languages.

For a word $w = a_1 \dots a_n$ over $\tilde{\mathbf{V}}$, let $\mathsf{G}(w)$ be the following graph where $|\mathsf{G}(w)| = [0, n]$:



$\mathsf{G}(w)$ is the unique graph in $\mathcal{G}(w)$ up to graph isomorphisms.

For a $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{I}},\top\}}$ term t, we write $[t]_{\tilde{\mathbf{V}}}$ for the word language $[t]$ over $\tilde{\mathbf{V}}$ where $\overline{x}, \bar{\mathsf{I}}, \top$ are also viewed as letters; e.g., $[\overline{x}]_{\tilde{\mathbf{V}}} = \{\overline{x}\}$ and $[\overline{x}] = \mathbf{V}^* \setminus \{x\}$ for $x \in \mathbf{V}$. Note that $\mathcal{G}(t) = \{\mathsf{G}(w) \mid w \in [t]_{\tilde{\mathbf{V}}}\}$. Hence, graph languages are expressible by word languages for $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{I}},\top\}}$ terms.

Additionally, we introduce nondeterministic finite word automata with epsilon transitions (*NFAs*). NFAs are (2-pointed) graphs over $\tilde{\mathbf{V}}_{\mathsf{I}}$ where the source and target vertices denote the initial and (single) accepting states, respectively, and I-labelled edges denote epsilon transitions. For a graph H and a word $w = a_1 \dots a_n$, we write δ_w^H for the binary relation $(\mathsf{I}^H)^* \cdot a_1^H \cdot (\mathsf{I}^H)^* \cdot \dots \cdot a_n^H \cdot (\mathsf{I}^H)^*$. For $q \in |H|$, we let $\delta_w^H(q) \triangleq \{q' \mid \langle q, q' \rangle \in \delta_w^H\}$. For $Q \subseteq |H|$, we let $\delta_w^H(Q) \triangleq \bigcup_{q \in Q} \delta_w^H(q)$. The word language $[H]_{\tilde{\mathbf{V}}}$ is defined as $\{w \in \tilde{\mathbf{V}}^* \mid \langle 1^H, 2^H \rangle \in \delta_w^H\}$. Note that $[H]_{\tilde{\mathbf{V}}} = \{w \in \tilde{\mathbf{V}}^* \mid \mathsf{G}(w) \longrightarrow H^{\mathcal{Q}}\}$ if I^H is an equivalence relation. Then we have the following:

▶ **Corollary 5.4.** *Let $\mathcal{C} \subseteq \mathsf{GREL}$ be induced subgraph-closed. For all $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{I}},\top\}}$ terms t, s,*

$$\mathcal{C} \models t \leq s \quad \Leftrightarrow \quad [t]_{\tilde{\mathbf{V}}} \subseteq \{w \in \tilde{\mathbf{V}}^* \mid \forall H \in \mathcal{S}_{\mathcal{C}}(\mathsf{G}(w)), [s]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} \neq \emptyset\}.$$

Proof. By Thm. 5.2. See Sect. 5.3 for a detailed proof. ◀

5.3 Proof of Thm. 5.2 and Cor. 5.4

Let $\mathfrak{v} \in \mathsf{GREL}$. For a graph H and a graph language $\mathcal{G}$, we write

$$\hat{\mathfrak{v}}(H) \quad \triangleq \quad \{\langle x, y \rangle \mid H \longrightarrow \mathsf{G}(\mathfrak{v}, x, y)\} \qquad\qquad \hat{\mathfrak{v}}(\mathcal{G}) \quad \triangleq \quad \bigcup_{H \in \mathcal{G}} \hat{\mathfrak{v}}(H).$$

▶ **Lemma 5.5.** *Let $\mathfrak{v} \in \mathsf{GREL}$. For all $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{I}},\top,\cap\}}$ terms t, we have $\hat{\mathfrak{v}}(t) = \hat{\mathfrak{v}}(\mathcal{G}(t))$.*

Proof (cf. [18, Prop. 11]). By easy induction on t. See Appendix F for a detail. ◀

▶ **Lemma 5.6.** *Let $\mathcal{C} \subseteq \mathsf{GREL}$, $\mathfrak{v} \in \mathcal{C}$, and H be a graph. Then $\hat{\mathfrak{v}}(\mathcal{S}_{\mathcal{C}}(H)^{\mathcal{Q}}) = \hat{\mathfrak{v}}(H)$.*

Proof. ($\subseteq$): Because, for any $J \in \mathcal{S}_{\mathcal{C}}(H)$, we have $J \longrightarrow J^{\mathcal{Q}}$ by $H \longrightarrow J$ and $J \longrightarrow J^{\mathcal{Q}}$. ($\supseteq$): Let $h \colon H \longrightarrow \mathsf{G}(\mathfrak{v}, x, y)$. Then $\mathcal{S}(h) \longrightarrow \mathsf{G}(\mathfrak{v}, x, y)$. Because $\mathsf{I}^{\mathsf{G}(\mathfrak{v},x,y)}$ is the identity relation, $\mathcal{S}(h)^{\mathcal{Q}} \longrightarrow \mathsf{G}(\mathfrak{v}, x, y)$. Hence $\hat{\mathfrak{v}}(\mathcal{S}_{\mathcal{C}}(H)^{\mathcal{Q}}) \supseteq \hat{\mathfrak{v}}(H)$. ◀

▶ **Proposition 5.7.** *Let $h \colon H \longrightarrow \mathsf{G}(\mathfrak{v}, x, y)$. Then the graph $\mathcal{S}(h)^{\mathcal{Q}}$ is isomorphic to the induced subgraph of $\mathsf{G}(\mathfrak{v}, x, y)$ on the range of h.*

Proof. Easy, by construction. Note that since $\mathsf{I}^{\mathsf{G}(\mathfrak{v},x,y)}$ is the identity relation, $\mathsf{I}^{\mathcal{S}(h)}$ is an equivalence relation (see, e.g., Example 2.2). ◀

Proof of Thm. 5.2. We have

$$\begin{aligned}
&\mathcal{C} \models t \leq s \;\Leftrightarrow\; \forall \mathfrak{v} \in \mathcal{C}, \hat{\mathfrak{v}}(t) \subseteq \hat{\mathfrak{v}}(s) \;\Leftrightarrow\; \forall \mathfrak{v} \in \mathcal{C}, \hat{\mathfrak{v}}(\mathcal{S}_{\mathcal{C}}(\mathcal{G}(t))^{\mathcal{Q}}) \subseteq \hat{\mathfrak{v}}(\mathcal{G}(s)) && \text{(Lems. 5.5 and 5.6)}\\
&\Leftrightarrow \forall G \in \mathcal{S}_{\mathcal{C}}(\mathcal{G}(t))^{\mathcal{Q}}, \forall J \in \mathrm{GR}_{\mathcal{C}},\ (G \longrightarrow J) \text{ implies } (\exists H \in \mathcal{G}(s),\ H \longrightarrow J) && \text{(Def. of } \hat{\mathfrak{v}})\\
&\Leftrightarrow \forall G \in \mathcal{S}_{\mathcal{C}}(\mathcal{G}(t))^{\mathcal{Q}}, \exists H \in \mathcal{G}(s),\ H \longrightarrow G. && (\heartsuit)
\end{aligned}$$

Here, for ($\heartsuit$), $\Leftarrow$: Let $H \in \mathcal{G}(s)$ be s.t. $H \longrightarrow G$. Then for all J s.t. $G \longrightarrow J$, we have $H \longrightarrow J$ by transitivity of $\longrightarrow$. $\Rightarrow$: By Prop. 5.7 and that $\mathrm{GR}_{\mathcal{C}}$ is induced subgraph-closed, we have $\mathcal{S}_{\mathcal{C}}(\mathcal{G}(t))^{\mathcal{Q}} \subseteq \mathrm{GR}_{\mathcal{C}}$. Thus by letting $J = G$, this completes the proof. ◀

Proof of Cor. 5.4. We have

$$\begin{aligned}
&\mathcal{C} \models t \leq s \;\Leftrightarrow\; \forall w \in [t]_{\tilde{\mathbf{V}}}, \forall H \in \mathcal{S}_{\mathcal{C}}(\mathsf{G}(w)), \exists v \in [s]_{\tilde{\mathbf{V}}}, \mathsf{G}(v) \longrightarrow H^{\mathcal{Q}}\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(Thm. 5.2 and } \mathcal{G}(s) = \{\mathsf{G}(v) \mid v \in [s]_{\tilde{\mathbf{V}}}\})\\
&\Leftrightarrow\quad \forall w \in [t]_{\tilde{\mathbf{V}}}, \forall H \in \mathcal{S}_{\mathcal{C}}(\mathsf{G}(w)), [s]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} \neq \emptyset \qquad ([H]_{\tilde{\mathbf{V}}} = \{v \in \tilde{\mathbf{V}}^* \mid \mathsf{G}(v) \longrightarrow H^{\mathcal{Q}}\})\\
&\Leftrightarrow\quad [t]_{\tilde{\mathbf{V}}} \subseteq \{w \in \tilde{\mathbf{V}}^* \mid \forall H \in \mathcal{S}_{\mathcal{C}}(\mathsf{G}(w)), [s]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} \neq \emptyset\}.
\end{aligned}$$

◀

6 PSPACE decidability for $\mathrm{KA}_{\{\overline{x},\overline{1},\top\}}$ terms

In this section, we show that the equational theory w.r.t. languages for $\mathrm{KA}_{\{\overline{x},\overline{1},\top\}}$ is decidable and PSPACE-complete. We consider graphs over $\tilde{\mathbf{V}}_{\mathsf{I}}$ / NFAs, instead of $\mathrm{KA}_{\{\overline{x},\overline{1},\top\}}$ terms / regular expressions over the alphabet $\tilde{\mathbf{V}}$. Relying on the graph characterization (Cor. 5.4), we consider that, given an NFA J (corresponding to the term s in Cor. 5.4), we give an NFA recognizing the following word language:

$$\mathsf{L}_J \triangleq \{w \in \tilde{\mathbf{V}}^* \mid \exists H \in \mathcal{S}_{\mathsf{RSUB}}(\mathsf{G}(w)), [J]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} = \emptyset\}.$$

Note that $\mathsf{RSUB} \models t \leq s \Leftrightarrow [t]_{\tilde{\mathbf{V}}} \cap \mathsf{L}_J = \emptyset$ when $[s]_{\tilde{\mathbf{V}}} = [J]_{\tilde{\mathbf{V}}}$. We first give an equivalent notion of "$w \in \mathsf{L}_J$" in Sect. 6.1, and then we give an NFA construction in Sect. 6.2.

6.1 Saturable paths for RSUB

We first give an equivalent notion of $[J]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} = \emptyset$, compatible with edge-saturations.

▶ **Definition 6.1.** *Let J and H be graphs. A map $U\colon |H| \to \wp(|J|)$ is an* emptiness-witness *for $[J]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} = \emptyset$ if the following hold where $U_x \triangleq U(x)$:*

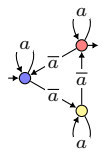
- $1^J \in U_{1^H}$ *and* $\forall a \in \tilde{\mathbf{V}}_{\mathsf{I}}, \forall \langle x, y\rangle \in a^H$, $\delta^J_a(U_x) \subseteq U_y$;
- $2^J \notin U_{2^H}$. ⌟

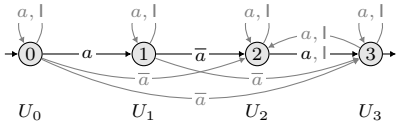
Intuitively, the first condition denotes that U is a cover of the reachable states from the pair "$1^J \in U_{1^H}$". If the second condition holds, we can see that the pair "$2^J \in U_{2^H}$" is unreachable. As expected, we have the following (see Appendix B, for a proof).

▶ **Proposition 6.2.** *Let J and H be graphs where I^H is reflexive. Then*

$$[J]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} = \emptyset \quad \Leftrightarrow \quad \exists U\colon |H| \to \wp(|J|),\ U \text{ is an emptiness-witness for } [J]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} = \emptyset.$$

▶ **Example 6.3.** We consider the following graphs. J denotes the NFA s.t. $[J]_{\tilde{\mathbf{V}}} = \{w \in \{a, \overline{a}\}^* \mid \exists n \in \mathbb{N}, \overline{a} \text{ occurs } 3n+2 \text{ times in } w\}$. H is a graph in $\mathcal{S}_{\mathsf{RSUB}}(\mathsf{G}(a\overline{a}a))$ where $\top$- or $\overline{\mathsf{I}}$-labelled edges are omitted. By the form of H, one can see that $[J]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} = \emptyset$.

$J =$ 

$H =$ 

If $U_0 = U_1 = \{\bullet\}$ and $U_2 = U_3 = \{\circ, \bullet\}$, then this U is an emptiness-witness; e.g., for $\langle 1, 2\rangle \in \overline{a}^H$, $\delta^J_{\overline{a}}(U_1) = \delta^J_{\overline{a}}(\{\bullet\}) = \{\circ\} \subseteq U_2$. Besides this, if $U_0 = U_1 = \{\bullet\}$ and $U_2 = U_3 = \{\circ\}$, this U is also an emptiness-witness. These witnesses also show $[J]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} = \emptyset$. ⌟

Next, we give an equivalent notion of "$w \in \mathsf{L}_J$". In the following, we abstract saturated edges (gray-colored edges in Example 6.3) using U, as U implies the existence of edge-saturations.

▶ **Definition 6.4.** *Let J be a graph and w be a word. A pair $P = \langle H, U\rangle$ is a* saturable path *for $w \in \mathsf{L}_J$ if the following hold:*

(P-Ext) *H is an edge-extension of $\mathsf{G}(w)$ such that*[13]

[13] In this definition, $\top^H, \mathsf{I}^H, \overline{\mathsf{I}}^H$ and edges in I^H are already edge-saturated, for preserving (P-Con) easily.

- $\top^H$ *is a total preorder and* $\top^H \supseteq \{\langle i-1, i\rangle \mid i \in [1,n]\}$ *where* $w = a_1 \dots a_n$*;*
- $\mathsf{I}^H = \top^H \cap \{\langle j,i\rangle \mid \langle i,j\rangle \in \top^H\}$ *and* $\bar{\mathsf{I}}^H = \top^H \setminus \mathsf{I}^H$*;*
- $\forall a \in \mathbf{V}$*,* $\langle a^H, \overline{a}^H\rangle$ *is either* $\langle a^{\mathsf{G}(w)} \cup \mathsf{I}^H, \overline{a}^{\mathsf{G}(w)}\rangle$ *or* $\langle a^{\mathsf{G}(w)}, \overline{a}^{\mathsf{G}(w)} \cup \mathsf{I}^H\rangle$*.*

(P-Con) *H is* consistent*:* $\forall a \in \mathbf{V},\ a^{H^{\circlearrowright}} \cap \overline{a}^{H^{\circlearrowright}} = \emptyset$*.*

(P-Wit) *$U\colon |H| \to \wp(|J|)$ is an emptiness-witness for* $[J]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} = \emptyset$*.*

(P-Sat) *H is* saturable*:* $\forall a \in \mathbf{V}, \forall \langle i,j\rangle \in \bar{\mathsf{I}}^H$*,* $\delta^J_a(U_i) \subseteq U_j$ *or* $\delta^J_{\overline{a}}(U_i) \subseteq U_j$*.* ⌟

▶ **Lemma 6.5.** *Let J be a graph and w be a word. Then,*

$$w \in \mathsf{L}_J \quad \Leftrightarrow \quad \text{there is a saturable path for } w \in \mathsf{L}_J.$$

Proof. ($\Rightarrow$): By Prop. 6.2, let $H' \in \mathcal{S}_{\mathsf{RSUB}}(\mathsf{G}(w))$ and let U be an emptiness-witness for $[J]_{\tilde{\mathbf{V}}} \cap [H']_{\tilde{\mathbf{V}}} = \emptyset$. We define the graph H as follows:

- $|H| = |H'|$;
- $a^H = a^{H'}$ for $a \in \{\top, \mathsf{I}, \bar{\mathsf{I}}\}$;
- $a^H = a^{\mathsf{G}(w)} \cup (a^{H'} \cap \mathsf{I}^{H'})$ for $a \in \tilde{\mathbf{V}}_{\mathsf{I}} \setminus \{\top, \mathsf{I}, \bar{\mathsf{I}}\}$.

We then have that the pair $P \triangleq \langle H, U\rangle$ is a saturable path for $w \in \mathsf{L}_J$, as follows:

- (P-Ext): By that H' is an edge-saturation w.r.t. RSUB.
- (P-Con): Because H' is consistent by $H' \in \mathcal{S}_{\mathsf{RSUB}}(\mathsf{G}(w))$.
- (P-Wit): Because U is an emptiness-witness for $[J]_{\tilde{\mathbf{V}}} \cap [H']_{\tilde{\mathbf{V}}} = \emptyset$.
- (P-Sat): Because $a^{H'} \cup \overline{a}^{H'} = \top^{H'}$ and U is an emptiness-witness for $[J]_{\tilde{\mathbf{V}}} \cap [H']_{\tilde{\mathbf{V}}} = \emptyset$.

($\Leftarrow$): Let $P = \langle H, U\rangle$ be the saturable path. By (P-Ext), I^H is an equivalence relation. We define the graph H' as follows:

- $|H'| = |H|$;
- $a^{H'} = a^H$ for $a \in \{\top, \mathsf{I}, \bar{\mathsf{I}}\}$;
- for $a \in \mathbf{V}$ and $\langle x, y\rangle \in \top^H$,
  - if $\langle [x]_{\mathsf{I}^H}, [y]_{\mathsf{I}^H}\rangle \in a^{H^{\circlearrowright}}$, then $\langle x,y\rangle \in a^{H'} \setminus \overline{a}^{H'}$;
  - else if $\langle [x]_{\mathsf{I}^H}, [y]_{\mathsf{I}^H}\rangle \in \overline{a}^{H^{\circlearrowright}}$, then $\langle x,y\rangle \in \overline{a}^{H'} \setminus a^{H'}$;
  - else if $U_y \subseteq \delta^J_a(U_x)$, then $\langle x,y\rangle \in a^{H'} \setminus \overline{a}^{H'}$;
  - else $\langle x,y\rangle \in \overline{a}^{H'} \setminus a^{H'}$.

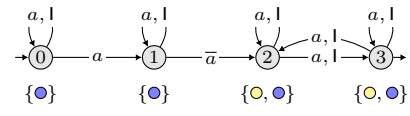
By the construction of H', we have the following:

- H' is consistent: If $[x]_{\mathsf{I}^H} = [y]_{\mathsf{I}^H}$ then $U_x = U_y$, because $U_x \subseteq \delta^J_{\mathsf{I}}(U_x) \subseteq U_y \subseteq \delta^J_{\mathsf{I}}(U_y) \subseteq U_x$ by (P-Wit); thus, if $[x]_{\mathsf{I}^H} = [x']_{\mathsf{I}^H}$ and $[y]_{\mathsf{I}^H} = [y']_{\mathsf{I}^H}$, then $\langle x,y\rangle \in a^{H'}$ iff $\langle x',y'\rangle \in a^{H'}$.
- for $a \in \mathbf{V}$, $\overline{a}^{H'} = \top^{H'} \setminus a^{H'}$: Because $a^{H'} \cup \overline{a}^{H'} = \top^{H'}$ and H' is consistent.

Combining (P-Ext) and them yields $H' \in \mathcal{S}_{\mathsf{RSUB}}(\mathsf{G}(w))$. Also, U is an emptiness-witness for $[J]_{\tilde{\mathbf{V}}} \cap [H']_{\tilde{\mathbf{V}}} = \emptyset$ as follows. For edges already in H, it is shown by (P-Wit). For extended edges from H, it is shown by the construction of H' (for the last case of the four cases above, by $U_y \not\subseteq \delta^J_a(U_x)$ and (P-Sat), we have $U_y \subseteq \delta^J_{\overline{a}}(U_x)$). Hence, this completes the proof. ◀

▶ **Example 6.6.** We recall the graphs J and $H \in \mathcal{S}_{\mathsf{RSUB}}(\mathsf{G}(a\overline{a}a))$ in Example 6.3. The following P is a saturable path for $a\overline{a}a \in \mathsf{L}_J$ where $\top$- or $\bar{\mathsf{I}}$-labelled edges are omitted:

$$P \quad = \quad$$



(P is of the form of a *path graph* by taking the quotient graph w.r.t. I-labelled edges.) P is an abstraction of edge-saturated graphs. From P, we can construct a graph $H \in \mathcal{S}_{\mathsf{RSUB}}(\mathsf{G}(a\overline{a}a))$ s.t. $[J]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} = \emptyset$. Because both $\delta_a^J(\{\bullet\}) \subseteq \{\circ, \bullet\}$ and $\delta_{\overline{a}}^J(\{\bullet\}) \subseteq \{\circ, \bullet\}$ hold, in addition to the graph H in Example 6.3, for instance, the following are also possible:

Thanks to saturable paths, we can replace the existence of such gray-colored edges connecting distant vertices with a locally defined witness U. ⌟

Moreover, we can replace (P-Sat)—a condition between distant vertices—with a local condition. Let $\varphi^J(\mathcal{U}, U) \triangleq \forall a \in \mathbf{V}, \forall \langle u, u' \rangle \in \mathcal{U}, \delta_a^J(u) \subseteq U \vee \delta_{\overline{a}}^J(u') \subseteq U$.

▶ **Proposition 6.7.** *Let J and H be graphs. Let $i \in |H|$. Then we have:*

$$(\forall a \in \mathbf{V}, \forall j \text{ s.t. } \langle j, i \rangle \in \bar{\mathsf{I}}^H, \delta_a^J(U_j) \subseteq U_i \vee \delta_{\overline{a}}^J(U_j) \subseteq U_i) \quad \Leftrightarrow \quad \varphi^J\Big(\bigcup_{j;\langle j,i\rangle \in \bar{\mathsf{I}}^H} U_j \times U_j, U_i\Big).$$

Proof. $(\forall a \in \mathbf{V}, \delta_a^J(U_j) \subseteq U_i \vee \delta_{\overline{a}}^J(U_j) \subseteq U_i)$ iff $(\forall a \in \mathbf{V}, (\forall u \in U_j, \delta_a^J(u) \subseteq U_i) \vee (\forall u' \in U_j, \delta_{\overline{a}}^J(u') \subseteq U_i))$ iff $\varphi^J(U_j \times U_j, U_i)$ (by taking the prenex normal form). Thus by $(\forall j \text{ s.t. } \langle j, i \rangle \in \bar{\mathsf{I}}^H, \varphi^J(U_j \times U_j, U_i))$ iff $\varphi^J(\bigcup_{j;\langle j,i\rangle \in \bar{\mathsf{I}}^H} U_j \times U_j, U_i)$, this completes the proof. ◀

6.2 Automata from saturable paths

Let $\mathcal{X} \triangleq \{X \in \wp(\tilde{\mathbf{V}}_{\mathsf{I}}) \mid \mathsf{I}, \top \in X, \bar{\mathsf{I}} \notin X, \text{ and } \forall x \in \mathbf{V}, x \in X \leftrightarrow \overline{x} \notin X\}$. (This set is equivalent to the set $\{\{x \in \tilde{\mathbf{V}}_{\mathsf{I}} \mid \mathsf{I}^H \subseteq x^H\} \mid H \in \mathrm{GR}_{\mathsf{RSUB}}\}$.)

▶ **Definition 6.8** (NFA construction). *Let ▶ and ◀ be two fresh symbols. For a graph J and a set $X \in \mathcal{X}$, let $J^{\mathcal{S}X}$ be the graph G defined as follows:*

- $|G| = \{\blacktriangleright, \blacktriangleleft\} \cup \{\langle \mathcal{U}, U \rangle \in \wp(|J| \times |J|) \times \wp(|J|) \mid \varphi^J(\mathcal{U}, U) \wedge \forall x \in X, \delta_x^J(U) \subseteq U\}$;
- $\mathsf{I}^G = \{\blacktriangleright\} \times \{\langle \mathcal{U}, U \rangle \in |G| \mid 1^J \in U \wedge \mathcal{U} = \emptyset\} \cup \{\langle \mathcal{U}, U \rangle \in |G| \mid 2^J \notin U\} \times \{\blacktriangleleft\}$;
- $x^G = \{\langle \langle \mathcal{U}, U \rangle, \langle \mathcal{U}', U' \rangle \rangle \in |G| \times |G| \mid \psi_{x,\mathsf{I}}^X(\mathcal{U}, U, \mathcal{U}', U') \vee \psi_{x,\bar{\mathsf{I}}}^X(\mathcal{U}, U, \mathcal{U}', U')\}$ *for* $x \in \tilde{\mathbf{V}}$;
- $1^G = \blacktriangleright$;
- $2^G = \blacktriangleleft$.

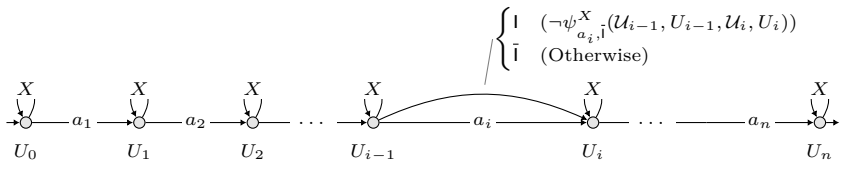
Here, $\psi_{x,\bar{\mathsf{I}}}^X(\mathcal{U}, U, \mathcal{U}', U')$ and $\psi_{x,\mathsf{I}}^X(\mathcal{U}, U, \mathcal{U}', U')$ are as follows:

- $\psi_{x,\bar{\mathsf{I}}}^X(\mathcal{U}, U, \mathcal{U}', U') \quad \triangleq \quad \left(\mathcal{U}' = \mathcal{U} \cup (U \times U) \wedge \bigwedge \left\{ \begin{array}{r} \delta_x^J(U) \subseteq U' \\ \delta_\top^J(\{u \mid \langle u, u \rangle \in \mathcal{U}'\}) \subseteq U' \\ \delta_{\bar{\mathsf{I}}}^J(\{u \mid \langle u, u \rangle \in \mathcal{U}'\}) \subseteq U' \end{array} \right\}\right)$;
- $\psi_{x,\mathsf{I}}^X(\mathcal{U}, U, \mathcal{U}', U') \quad \triangleq \quad (\mathcal{U}' = \mathcal{U} \wedge U' = U \wedge x \in X)$. ⌟

By the form of $J^{\mathcal{S}X}$, if $a_1 \dots a_n \in [J^{\mathcal{S}X}]_{\tilde{\mathbf{V}}}$, then its run is of the following form:

$$\to \blacktriangleright \xrightarrow{\mathsf{I}} \langle \mathcal{U}_0, U_0 \rangle \xrightarrow{a_1} \langle \mathcal{U}_1, U_1 \rangle \xrightarrow{a_2} \langle \mathcal{U}_2, U_2 \rangle \longrightarrow \cdots \longrightarrow \xrightarrow{a_n} \langle \mathcal{U}_n, U_n \rangle \xrightarrow{\mathsf{I}} \blacktriangleleft \to .$$

Intuitively, this run corresponds to the following saturable path where some $\top$-, $\bar{\mathsf{I}}$-, or I-labelled edges are omitted and $\overset{X}{\bigcirc}$ denotes that an x-labelled edge exists for each $x \in X$:



Here, $\mathcal{U}_i$ is used to denote the set $\bigcup_{j;\langle j,i\rangle\in\bar{\mathsf{l}}^H} U_j \times U_j$ (cf. Prop. 6.7) where H is the graph of the saturable path above.

According to the correspondence above, from a word $w \in \bigcup_{X\in\mathcal{X}}[J^{\mathcal{S}_X}]_{\tilde{\mathbf{V}}}$, we can construct a saturable path for $w \in \mathsf{L}_J$. Conversely, from a saturable path for $w \in \mathsf{L}_J$, we can show $w \in \bigcup_{X\in\mathcal{X}}[J^{\mathcal{S}_X}]_{\tilde{\mathbf{V}}}$. Thus, we have the following. See Appendix C for the detailed proof.

▶ **Lemma 6.9** (Appendix C). *Let J be a graph. We have $\mathsf{L}_J = \bigcup_{X\in\mathcal{X}}[J^{\mathcal{S}_X}]_{\tilde{\mathbf{V}}}$.*

▶ **Theorem 6.10.** *The equational theory of $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{l}},\top\}}$ w.r.t. languages is PSPACE-complete.*

Proof. (in PSPACE): Let t and s be $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{l}},\top\}}$ terms. Let G and J be graphs s.t. $[G]_{\tilde{\mathbf{V}}} = [t]_{\tilde{\mathbf{V}}}$ and $[J]_{\tilde{\mathbf{V}}} = [s]_{\tilde{\mathbf{V}}}$. Then we have

$$\begin{aligned} \mathsf{RSUB} \models t \le s \quad &\Leftrightarrow \quad [G]_{\tilde{\mathbf{V}}} \cap \mathsf{L}_J = \emptyset && \text{(Cor. 5.4)} \\ &\Leftrightarrow \quad [G]_{\tilde{\mathbf{V}}} \cap \Big(\bigcup_{X\in\mathcal{X}} [J^{\mathcal{S}_X}]_{\tilde{\mathbf{V}}}\Big) = \emptyset. && \text{(Lem. 6.9)} \end{aligned}$$

Hence, we can reduce the equational theory into the emptiness problem of NFAs of size exponential to the size of the input equation. Therefore, by using a standard on-the-fly algorithm for the non-emptiness problem of NFAs (essentially the graph reachability problem), we can give a non-deterministic polynomial space algorithm. (Hardness): The equational theory of KA coincides with the language equivalence problem of regular expressions, which is PSPACE-complete [16]. Hence, the equational theory of $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{l}},\top\}}$ is PSPACE-hard. ◀

▶ Remark 6.11. Cf. [18, Remark 45]: the equational theory of REL for $\mathrm{KA}_{\{\overline{x},\bar{\mathsf{l}},\top\}}$ is decidable in coNEXP, but its PSPACE-decidability is open. In RSUB, each equivalence class induced from $\bar{\mathsf{l}}$-labelled edges is always an interval. Thus, the problematic case of [18, Remark 45] does not occur in RSUB, fortunately. ⌟

7 Future directions

We leave open the axiomatizability of the equational theory of $\mathrm{KA}_{\{\overline{x}\}}/\mathrm{KA}_{\{\bar{\mathsf{l}}\}}/\mathrm{KA}_{\{\overline{x},\bar{\mathsf{l}}\}}$ w.r.t. languages. Note that the equational theory of $\mathrm{KA}_{\{-\}}$ (without Kleene star) w.r.t. languages is not finitely axiomatizable because it is Π^0_1-complete (Thms. 4.7 and 4.10).

A natural interest is to consider variants or fragments of $\mathrm{KA}_{\{-\}}$, e.g., with reverse [3], with tests [14] (by considering guarded strings) or with (anti-)domain [8]. Also, the combination of variables and letters (variables under standard valuations) would be interesting in the context of regular expressions.

Additionally, this paper did not so deal with the expressive power w.r.t. languages. A weak tool is (†-$\mathrm{KA}_{\{-\}}$): $[t] \neq [s] \Rightarrow \mathsf{LANG} \not\models t = s$; we can separate the expressive powers w.r.t. languages of given two term sets if their language classes are not equivalent. However, this cannot show e.g., the expressive power difference between $\mathrm{KA}_{\{\overline{x}\}}$ and KA, while this difference is an immediate consequence of the counter-example in the introduction and the completeness theorem of KA. To obtain a more useful tool, it is possibly interesting to modify existing arguments about the standard relational structures (REL), e.g., [25, 24, 26, 11, 17] (including finite model theoretical studies such as Ehrenfeucht-Fraïssé games [9, 10]) for RSUB.

References

1 Hajnal Andréka and D. A. Bredikhin. The equational theory of union-free algebras of relations. *Algebra Universalis*, 33(4):516–532, 1995. doi:10.1007/BF01225472.

2 Hajnal Andréka, Szabolcs Mikulás, and István Németi. The equational theory of Kleene lattices. *Theoretical Computer Science*, 412(52):7099–7108, 2011. doi:10.1016/J.TCS.2011.09.024.

3 S. L. Bloom, Z. Ésik, and Gh. Stefanescu. Notes on equational theories of relations. *algebra universalis*, 33(1):98–126, 1995. doi:10.1007/BF01190768.

4 Paul Brunet. Reversible Kleene lattices. In *MFCS*, volume 83 of *LIPIcs*, pages 66:1–66:14. Schloss Dagstuhl, 2017. doi:10.4230/LIPICS.MFCS.2017.66.

5 Paul Brunet. A complete axiomatisation of a fragment of language algebra. In *CSL*, volume 152 of *LIPIcs*, page 11:1–11:15. Schloss Dagstuhl, 2020. doi:10.4230/LIPIcs.CSL.2020.11.

6 Paul Brunet and Damien Pous. Petri automata. *Logical Methods in Computer Science*, 13(3), 2017. doi:10.23638/LMCS-13(3:33)2017.

7 John H. Conway. *Regular Algebra and Finite Machines*. Chapman and Hall, 1971.

8 Jules Desharnais, Bernhard Möller, and Georg Struth. Kleene algebra with domain. *ACM Transactions on Computational Logic*, 7(4):798–833, 2006. doi:10.1145/1183278.1183285.

9 Andrzej Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundamenta Mathematicae*, 49(2):129–141, 1961. doi:10.4064/fm-49-2-129-141.

10 Roland Fraïssé. Sur les classifications des systems de relations. *Publ. Sci. Univ. Alger I.*, 1954.

11 Jelle Hellings. *On Tarski's Relation Algebra: querying trees and chains and the semi-join algebra*. PhD thesis, Hasselt University and Transnational University of Limburg, 2018. URL: http://hdl.handle.net/1942/25932.

12 Stephen C. Kleene. Representation of events in nerve nets and finite automata. In *Automata Studies. (AM-34)*, pages 3–42. Princeton University Press, 1956. doi:10.1515/9781400882618-002.

13 Dexter Kozen. A completeness theorem for Kleene algebras and the algebra of regular events. In *LICS*, pages 214–225. IEEE Comput. Sco. Press, 1991. doi:10.1109/LICS.1991.151646.

14 Dexter Kozen and Frederick Smith. Kleene algebra with tests: Completeness and decidability. In *CSL*, volume 1258 of *LNCS*, pages 244–259. Springer, 1996. doi:10.1007/3-540-63172-0_43.

15 F. W. Levi. On semigroups. *Bulletin of the Calcutta Mathematical Society*, (36):141–146, 1944.

16 A. R. Meyer and L. J. Stockmeyer. The equivalence problem for regular expressions with squaring requires exponential space. In *SWAT*, pages 125–129. IEEE, 1972. doi:10.1109/SWAT.1972.29.

17 Yoshiki Nakamura. Expressive power and succinctness of the positive calculus of binary relations. *Journal of Logical and Algebraic Methods in Programming*, 127:100760, 2022. doi:10.1016/j.jlamp.2022.100760.

18 Yoshiki Nakamura. Existential calculi of relations with transitive closure: Complexity and edge saturations. In *LICS*, pages 1–13. IEEE, 2023. doi:10.1109/LICS56636.2023.10175811.

19 Yoshiki Nakamura and Ryoma Sin'ya. Words-to-letters valuations for language Kleene algebras with variable complements. In *AFL*, volume 386, pages 185–199. EPTCS, 2023. doi:10.4204/EPTCS.386.15.

20 Damien Pous. On the positive calculus of relations with transitive closure. In *STACS*, volume 96 of *LIPIcs*, pages 3:1–3:16. Schloss Dagstuhl, 2018. doi:10.4230/LIPICS.STACS.2018.3.

21 Damien Pous and Jana Wagemaker. Completeness theorems for kleene algebra with top. In *CONCUR*, volume 243 of *LIPIcs*, pages 26:1–26:18. Schloss Dagstuhl, 2022. doi:10.4230/LIPICS.CONCUR.2022.26.

22 V. R. Pratt. Dynamic algebras and the nature of induction. In *STOC*, page 22–28. ACM, 1980. doi:10.1145/800141.804649.

23 Robin Hirsch, Ian Hodkinson. *Relation Algebras by Games*, volume 147 of *Studies in logic and the foundations of mathematics*. Elsevier, 1 edition, 2002.

24 Dimitri Surinx. A framework for comparing query languages in their ability to express boolean queries. 2017. URL: http://hdl.handle.net/1942/25215.

25 Dimitri Surinx, George H. L. Fletcher, Marc Gyssens, Dirk Leinders, Jan Van den Bussche, Dirk Van Gucht, Stijn Vansummeren, and Yuqing Wu. Relative expressive power of navigational

querying on graphs using transitive closure. *Logic Journal of the IGPL*, 23(5):759–788, 2015. doi:10.1093/jigpal/jzv028.

26 Dimitri Surinx, Jan Van den Bussche, and Dirk Van Gucht. The primitivity of operators in the algebra of binary relations under conjunctions of containments. In *LICS*, pages 1–10. IEEE, 2017. doi:10.1109/LICS.2017.8005122.

27 Alfred Tarski. On the calculus of relations. *The Journal of Symbolic Logic*, 6(3):73–89, 1941. doi:10.2307/2268577.

28 Cheng Zhang, Arthur Azevedo de Amorim, and Marco Gaboardi. On incorrectness logic and Kleene algebra with top and tests. *Proceedings of the ACM on Programming Languages*, 6(POPL):29:1–29:30, 2022. doi:10.1145/3498690.

A Proof for Remark 3.12

▶ **Definition A.1.** *Let*

$$\mathsf{SUB}_{\mathrm{st}} \triangleq \bigcup_{n\in\mathbb{N}} \left\{ \mathfrak{v} \in \mathsf{SUB}_n \;\middle|\; \begin{array}{c} \bigcup_{a\in\mathbf{V}} \mathfrak{v}(a) = \{\ell_i \mid i \in [0,n-1]\} \\ \mathfrak{v}(a) \text{ (where } a \in \mathbf{V}\text{) are disjoint sets} \end{array} \right\}.$$

▶ **Proposition A.2.** *Let $n \in \mathbb{N}$ and let $\mathfrak{v}' \in \mathsf{SUB}_n \cap \mathsf{SUB}_{\mathrm{st}}$. For $i \in [0,n-1]$, let a_i be the unique one s.t. $\mathfrak{v}'(a_i) = \{\ell_i\}$. Then $\mathfrak{v}_{\mathrm{st}}^{\langle a_0,\dots,a_{n-1}\rangle} = \mathfrak{v}'$.*

Proof. By Def. 3.2, $\ell_i \in \mathfrak{v}_{\mathrm{st}}^{\langle a_0,\dots,a_{n-1}\rangle}(x)$ iff $a_i \in \mathfrak{v}_{\mathrm{st}}(x)$ iff $a_i = x$ iff $\ell_i \in \mathfrak{v}'(x)$. Also, both $\mathfrak{v}_{\mathrm{st}}^{\langle a_0,\dots,a_{n-1}\rangle}(x)$ and $\mathfrak{v}'(x)$ are subsets of $\{\ell_0,\dots,\ell_{n-1}\}$. Thus, $\mathfrak{v}_{\mathrm{st}}^{\langle a_0,\dots,a_{n-1}\rangle} = \mathfrak{v}'$. ◀

▶ **Lemma A.3** (cf. Lem. 3.6)**.** *For all quantifier-free formulas φ, we have:*

$$\mathfrak{v}_{\mathrm{st}} \models \varphi \quad\Leftrightarrow\quad \mathsf{SUB}_{\mathrm{st}} \models \varphi.$$

Proof. As with Lem. 3.6, it suffices to prove when φ is of the form $(\bigvee_{k=1}^{m} t_k \le \bot) \vee (\neg s \le \bot)$.

$$\mathsf{SUB}_{\mathrm{st}} \not\models (\bigvee_{k=1}^{m} t_k \le \bot) \vee (\neg s \le \bot)$$

$$\Rightarrow\quad \exists n, a_0,\dots,a_{n-1} \in \mathbf{V},\ \mathfrak{v}_{\mathrm{st}}^{\langle a_0,\dots,a_{n-1}\rangle} \not\models (\bigvee_{k=1}^{m} t_k \le \bot) \vee (\neg s \le \bot) \qquad \text{(Prop. A.2)}$$

$$\Rightarrow\quad \mathfrak{v}_{\mathrm{st}} \not\models (\bigvee_{k=1}^{m} t_k \le \bot) \vee (\neg s \le \bot) \qquad \text{(By Lem. 3.4)}$$

$$\Rightarrow\quad \exists n \in \mathbb{N}, i_1, j_1, \dots, i_m, j_m \text{ s.t. } 0 \le i_1 \le j_1 \le i_2 \le j_2 \le \cdots \le i_m \le j_m \le n,$$

$$\exists a_0,\dots,a_{n-1} \in \mathbf{V},\ (\bigwedge_{k=1}^{m} a_{i_k} \dots a_{j_k-1} \in \hat{\mathfrak{v}}_{\mathrm{st}}(t_k)) \wedge \hat{\mathfrak{v}}_{\mathrm{st}}(s) = \emptyset$$

$$\Rightarrow\quad \mathsf{SUB}_{\mathrm{st}} \not\models (\bigvee_{k=1}^{m} t_k \le \bot) \vee (\neg s \le \bot).$$

(By Lem. 3.4 and $\mathfrak{v}_{\mathrm{st}}^{\langle a_0,\dots,a_{n-1}\rangle} \in \mathsf{SUB}_n \cap \mathsf{SUB}_{\mathrm{st}}$) ◀

▶ **Proposition A.4** (Cor. of Prop. 3.8)**.** *For all quantifier-free formulas φ, we have:*

$$\mathsf{SUB}_{\mathrm{st}} \models \varphi \quad\Leftrightarrow\quad \mathsf{RSUB}_{\mathrm{st}} \models \varphi.$$

Proof. By the same bijective map $f\colon L \mapsto \{\langle i,j\rangle \mid \ell_i \dots \ell_{j-1} \in L\}$ as Prop. 3.8. ◀

Combining Lem. A.3 and Prop. A.4 yields

$$\mathfrak{v}_{\mathrm{st}} \models \varphi \quad \Leftrightarrow \quad \mathsf{RSUB}_{\mathrm{st}} \models \varphi.$$

Particularly, we have

$$[t] = [s] \quad \Leftrightarrow \quad \mathsf{RSUB}_{\mathrm{st}} \models t = s.$$

This matches the following correspondence between words and binary relations (cf. Sect. 5).

$$a_0 a_1 \dots a_{n-1} \qquad | \qquad \rightarrow\!\circ\!\xrightarrow{a_0}\!\circ\!\xrightarrow{a_1}\!\circ\!\cdots\!\xrightarrow{a_{n-1}}\!\circ\!\rightarrow \ .$$

B Proof of Prop. 6.2

Let $R' \subseteq |H| \times |J|$ be the minimal set such that

- $\langle 1^H, 1^J \rangle \in R'$;
- $\forall a \in \tilde{\mathbf{V}}_{\mathsf{I}}, x, y, x', y',\ \langle x, y \rangle \in R' \wedge \langle x, x' \rangle \in \delta_a^H \wedge \langle y, y' \rangle \in \delta_a^J \Rightarrow \langle x', y' \rangle \in R'$.

▷ Claim B.1. $[J]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} \neq \emptyset \quad \Leftrightarrow \quad \langle 2^H, 2^J \rangle \in R'$.

Proof Sketch. R' is the set of all reachable states of the product NFA of H and J. Hence this completes the proof. ◁

Let $R \subseteq |H| \times |J|$ be the minimal set such that

- $\langle 1^H, 1^J \rangle \in R$;
- $\forall a \in \tilde{\mathbf{V}}_{\mathsf{I}}, x, y, x', y',\ \langle x, y \rangle \in R \wedge \langle x, x' \rangle \in a^H \wedge \langle y, y' \rangle \in \delta_a^J \Rightarrow \langle x', y' \rangle \in R$.

▷ Claim B.2. $R = R'$.

Proof. ($\subseteq$): Clear, by $a^H \subseteq \delta_a^H$. ($\supseteq$): By induction on derivations of R'.

- Case $\langle 1^H, 1^J \rangle \in R'$: Trivial by $\langle 1^H, 1^J \rangle \in R$.
- Case $(\langle x, y \rangle \in R' \wedge \langle x, x' \rangle \in \delta_a^H \wedge \langle y, y' \rangle \in \delta_a^J) \Rightarrow \langle x', y' \rangle \in R'$: By IH, $\langle x, y \rangle \in R$.
  - Sub-Case $a \neq \mathsf{I}$: Let $x_0, \dots, x_{n-1}, x_n, \dots, x_m$ be s.t. $\langle x, x' \rangle = \langle x_0, x_m \rangle$ and
    - for all $i \in [1, n-1]$, $\langle x_{i-1}, x_i \rangle \in \mathsf{I}^H$;
    - $\langle x_{n-1}, x_n \rangle \in a^H$;
    - for all $i \in [n+1, m]$, $\langle x_{i-1}, x_i \rangle \in \mathsf{I}^H$.

    Let $y_0 = \dots = y_{n-1} = y$ and $y_n = \dots = y_m = y'$. Then by applying the second rule multiply, we have $\langle x', y' \rangle \in R$.
  - Sub-Case $a = \mathsf{I}$: By reflexivity of I^H, $\langle x, x' \rangle \in (\mathsf{I}^H)^+$. Let $x_0, \dots, x_m$ $(m > 0)$ be s.t. $\langle x, x' \rangle = \langle x_0, x_m \rangle$ and
    - for all $i \in [1, m]$, $\langle x_{i-1}, x_i \rangle \in \mathsf{I}^H$.

    Let $y_0 = y$ and $y_1 = \dots = y_m = y'$. Then by applying the second rule multiply, we have $\langle x', y' \rangle \in R$. ◁

Proof of Prop. 6.2. ($\Rightarrow$): By letting U as the map defined by $U(x) \triangleq \{y \mid \langle x, y \rangle \in R\}$. Here, $2^J \notin U_{2^H}$ is shown by $[J]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} = \emptyset$ with Claim B.1 and B.2. ($\Leftarrow$): Let $R'' \triangleq \{\langle x, y \rangle \mid y \in U(x)\}$. By the minimality of R, we have $R \subseteq R''$. By $\langle 2^H, 2^J \rangle \notin R''$, we have $\langle 2^H, 2^J \rangle \notin R$. Hence by Claim B.1 and B.2, we have $[J]_{\tilde{\mathbf{V}}} \cap [H]_{\tilde{\mathbf{V}}} = \emptyset$. ◀

C Proof of Lem. 6.9

Proof of Lem. 6.9. ($\subseteq$): Let $w = a_1 \dots a_n \in \mathsf{L}_J$. Let $P = \langle H, U\rangle$ be a saturable path for $w \in \mathsf{L}_J$. Let $X \triangleq \{a \in \tilde{\mathbf{V}}_{\mathsf{I}} \mid a^H \supseteq \mathsf{I}^H\}$ (note that $X \in \mathcal{X}$). For each i, let $\mathcal{U}_i \triangleq \bigcup_{j;\langle j,i\rangle\in\bar{\mathsf{I}}^H} U_j \times U_j$. Then we have:

- $\varphi^J(\mathcal{U}_i, U_i)$: By (P-Sat) and Prop. 6.7.
- $\forall a \in X, \delta^J_a(U_i) \subseteq U_i$: By $a^H \supseteq \mathsf{I}^H \supseteq \Delta_{|H|}$ and (P-Wit).

Thus $\langle \mathcal{U}_i, U_i\rangle \in |J^{\mathcal{S}_X}|$. We consider the following run of the NFA $J^{\mathcal{S}_X}$ on w:

$$\to \blacktriangleright \xrightarrow{\mathsf{I}} \langle \mathcal{U}_0, U_0\rangle \xrightarrow{a_1} \langle \mathcal{U}_1, U_1\rangle \xrightarrow{a_2} \langle \mathcal{U}_2, U_2\rangle \longrightarrow \cdots \xrightarrow{a_n} \langle \mathcal{U}_n, U_n\rangle \xrightarrow{\mathsf{I}} \blacktriangleleft \to .$$

This is indeed a run of the NFA $J^{\mathcal{S}_X}$ as follows:

- $\langle \blacktriangleright, \langle \mathcal{U}_0, U_0\rangle\rangle \in \mathsf{I}^{J^{\mathcal{S}_X}}$: By $1^J \in U_0$ (P-Wit) and $\mathcal{U}_0 = \emptyset$.
- $\langle \langle \mathcal{U}_n, U_n\rangle, \blacktriangleleft\rangle \in \mathsf{I}^{J^{\mathcal{S}_X}}$: By $2^J \notin U_n$ (P-Wit).
- $\forall i \in [1, n], \langle \langle \mathcal{U}_{i-1}, U_{i-1}\rangle, \langle \mathcal{U}_i, U_i\rangle\rangle \in a_i^{J^{\mathcal{S}_X}}$: We distinguish the following cases:
  - Case $\langle i-1, i\rangle \in \mathsf{I}^H$:
    - $\mathcal{U}_i = \mathcal{U}_{i-1}$: By $\langle j, i\rangle \in \bar{\mathsf{I}}^H$ iff $\langle j, i-1\rangle \in \bar{\mathsf{I}}^H$, for all j.
    - $U_i = U_{i-1}$: By (P-Wit), we have $U_{i-1} \subseteq \delta^J_{\mathsf{I}}(U_{i-1}) \subseteq U_i \subseteq \delta^J_{\mathsf{I}}(U_i) \subseteq U_{i-1}$.
    - $a_i \in X$ ($a_i^H \supseteq \mathsf{I}^H$): By $a_i^H \cap \mathsf{I}^H \neq \emptyset$ and (P-Ext), we have $a_i^H = a_i^{\mathsf{G}(w)} \cup \mathsf{I}^H$ (if not, this contradicts to (P-Con)).

    Thus by $\psi^X_{a_i,\mathsf{I}}(\mathcal{U}_{i-1}, U_{i-1}, \mathcal{U}_i, U_i)$, we have $\langle \langle \mathcal{U}_{i-1}, U_{i-1}\rangle, \langle \mathcal{U}_i, U_i\rangle\rangle \in a_i^{J^{\mathcal{S}_X}}$.
  - Case $\langle i-1, i\rangle \in \bar{\mathsf{I}}^H$:
    - $\mathcal{U}_i = \mathcal{U}_{i-1} \cup (U_{i-1} \times U_{i-1})$: By $\langle j, i\rangle \in \bar{\mathsf{I}}^H$ iff $\langle j, i-1\rangle \in \bar{\mathsf{I}}^H \vee j = i-1$, for all j.
    - $\delta^J_{a_i}(U_{i-1}) \subseteq U_i$: By (P-Wit).
    - $\delta^J_{\top}(\{u \mid \langle u, u\rangle \in \mathcal{U}_i\}) \subseteq U_i$: We have $\delta^J_{\top}(\{u \mid \langle u, u\rangle \in \mathcal{U}_i\}) = \delta^J_{\top}(\bigcup_{j;\langle j,i\rangle\in\bar{\mathsf{I}}^H} U_j) = \bigcup_{j<i} \delta^J_{\top}(U_j) \subseteq U_i$ by (P-Wit).
    - $\delta^J_{\mathsf{I}}(\{u \mid \langle u, u\rangle \in \mathcal{U}_i\}) \subseteq U_i$: We have $\delta^J_{\mathsf{I}}(\{u \mid \langle u, u\rangle \in \mathcal{U}_i\}) = \delta^J_{\mathsf{I}}(\bigcup_{j;\langle j,i\rangle\in\bar{\mathsf{I}}^H} U_j) = \bigcup_{j<i} \delta^J_{\mathsf{I}}(U_j) \subseteq U_i$ by (P-Wit).

    Thus by $\psi^X_{a_i,\bar{\mathsf{I}}}(\mathcal{U}_{i-1}, U_{i-1}, \mathcal{U}_i, U_i)$, we have $\langle \langle \mathcal{U}_{i-1}, U_{i-1}\rangle, \langle \mathcal{U}_i, U_i\rangle\rangle \in a_i^{J^{\mathcal{S}_X}}$.

Hence, $w \in [J^{\mathcal{S}_X}]$.

($\supseteq$): Let $X \subseteq \mathcal{X}$ and $w = a_1 \dots a_n \in [J^{\mathcal{S}_X}]_{\tilde{\mathbf{V}}}$. Let the run of $J^{\mathcal{S}_X}$ on w be as follows:

$$\to \blacktriangleright \xrightarrow{\mathsf{I}} \langle \mathcal{U}_0, U_0\rangle \xrightarrow{a_1} \langle \mathcal{U}_1, U_1\rangle \xrightarrow{a_2} \langle \mathcal{U}_2, U_2\rangle \longrightarrow \cdots \xrightarrow{a_n} \langle \mathcal{U}_n, U_n\rangle \xrightarrow{\mathsf{I}} \blacktriangleleft \to .$$

Let H be the edge-extension of $\mathsf{G}(w)$ defined as follows:

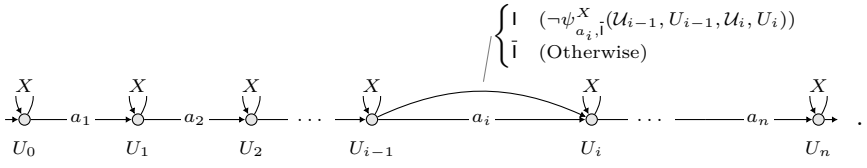
- $\top^H = \{\langle x, y\rangle \in [0, n] \times [0, n] \mid \forall i \in [y+1, x],\ \neg\psi^X_{a_i,\bar{\mathsf{I}}}(\mathcal{U}_{i-1}, U_{i-1}, \mathcal{U}_i, U_i)\}$;
- $\mathsf{I}^H = \top^H \cap \{\langle x, y\rangle \mid \langle y, x\rangle \in \top^H\}$ and $\bar{\mathsf{I}}^H = \top^H \setminus \mathsf{I}^H$;
- $\forall a \in \mathbf{V} \cap X$, $\langle a^H, \overline{a}^H\rangle = \langle a^{\mathsf{G}(w)} \cup \mathsf{I}^H, \overline{a}^{\mathsf{G}(w)}\rangle$.

Note that by definition of $\top^H$, we have

- $\top^H \supseteq \{\langle x, y\rangle \mid x \leq y\}$;
- $\top^H$ is transitive by case analysis.

Hence, $\top^H$ is a total preorder and each equivalence class w.r.t. I^H is an interval $[l, r]$.

Let $P \triangleq \langle H, U\rangle$ where U is defined as $i \mapsto U_i$ for $i \in [0, n]$. The following depicts P.



Then P is a saturable path for $w \in \mathsf{L}_J$ as follows:

- (P-Ext): By the definition of H.
- (P-Con): Assume that $a^{H^{\mathcal{Q}}} \cap \overline{a}^{H^{\mathcal{Q}}} \neq \emptyset$. Let x, x', y, y' be s.t. $[x]_{\mathsf{I}^H} = [x']_{\mathsf{I}^H}$, $[y]_{\mathsf{I}^H} = [y']_{\mathsf{I}^H}$, $\langle x, y\rangle \in a^H$, and $\langle x', y'\rangle \in \overline{a}^H$. WLOG, we can assume that $a \in X$ and $\overline{a} \notin X$. Then
  - $\langle x', y'\rangle \in \overline{a}^{\mathsf{G}(w)}$ (so, $x' = y' - 1$ and $a_{y'} = \overline{a}$): By $\overline{a}^H = \overline{a}^{\mathsf{G}(w)}$ (since $\overline{a} \notin X$).
  - $\langle x, y\rangle \in a^{\mathsf{G}(w)}$ (so, $x = y - 1$ and $a_y = a$): If not, then by $a^H = a^{\mathsf{G}} \cup \mathsf{I}^H$, we have $[x]_{\mathsf{I}^H} = [y]_{\mathsf{I}^H}$. Thus we have $[y' - 1]_{\mathsf{I}^H} = [y']_{\mathsf{I}^H}$. Then by the definition of $\top^H$, we have $\neg\psi^X_{\overline{a},\bar{\mathsf{I}}}(\mathcal{U}_{y'-1}, U_{y'-1}, \mathcal{U}_{y'}, U_{y'})$. By the definition of $a^{J^{\mathcal{S}X}}$, we have $\psi^X_{\overline{a},\mathsf{I}}(\mathcal{U}_{y'-1}, U_{y'-1}, \mathcal{U}_{y'}, U_{y'})$, so $\overline{a} \in X$. This contradicts $\overline{a} \notin X$.
  - $([x, x'] \cup [x', x]) \cap ([y, y'] \cup [y', y]) = \emptyset$ (so, $x = x'$ and $y = y'$): If not, then because the interval between x and x' and that between y and y' have an intersection, we have $[x]_{\mathsf{I}^H} = [y]_{\mathsf{I}^H}$. Then, in the same manner as above, we have $\overline{a} \in X$. This contradicts $\overline{a} \notin X$.

  Thus, we reach a contradiction, because $a = a_y = a_{y'} = \overline{a}$. Hence, $a^{H^{\mathcal{Q}}} \cap \overline{a}^{H^{\mathcal{Q}}} = \emptyset$.
- (P-Sat): By the the form of $J^{\mathcal{S}X}$, we have $\mathcal{U}_x = \begin{cases} \mathcal{U}_{x-1} & (\langle x-1, x\rangle \in \mathsf{I}^H) \\ \mathcal{U}_{x-1} \cup (U_{x-1} \times U_{x-1}) & (\langle x-1, x\rangle \in \bar{\mathsf{I}}^H) \end{cases}$.
  Thus, $\mathcal{U}_y = \bigcup_{x;\langle x,y\rangle \in \bar{\mathsf{I}}^H} U_x \times U_x$ ($\bigstar$). Thus by Prop. 6.7, this completes the proof.
- (P-Wit): For $\mathsf{1}^J \in U_0$ and $\mathsf{2}^J \notin U_n$, they are shown by the form of $J^{\mathcal{S}X}$. For $\forall a \in \tilde{\mathbf{V}}_{\mathsf{I}}, \forall \langle x, y\rangle \in a^H, \delta^J_a(U_x) \subseteq U_y$, we distinguish the following cases:
  - Case $a = \mathsf{I}$: Then we have
    - $U_x = U_y$: By $\langle x, y\rangle \in \mathsf{I}^H$ and the definition of $J^{\mathcal{S}X}$, we have that $\forall i \in [y + 1, x]$, $\psi^X_{a_i,\mathsf{I}}(\mathcal{U}_{i-1}, U_{i-1}, \mathcal{U}_i, U_i)$. Thus, $U_y = U_{y+1} = \cdots = U_x$.
    - $\delta^J_{\mathsf{I}}(U_x) \subseteq U_x$: By $\langle \mathcal{U}_x, U_x\rangle \in |J^{\mathcal{S}X}|$.

    Hence, $\delta^J_{\mathsf{I}}(U_x) \subseteq U_y$.
  - Case $a = \bar{\mathsf{I}}$: Let $z \in [x + 1, y]$ be such that $\psi^X_{a_z,\bar{\mathsf{I}}}(\mathcal{U}_{z-1}, U_{z-1}, \mathcal{U}_z, U_z)$ and $\forall z' \in [z + 1, y], \neg\psi^X_{a_{z'},\bar{\mathsf{I}}}(\mathcal{U}_{z'-1}, U_{z'-1}, \mathcal{U}_{z'}, U_{z'})$. Then we have

$$\begin{aligned} \delta^J_{\bar{\mathsf{I}}}(U_x) &\subseteq \delta^J_{\bar{\mathsf{I}}}(\{u \mid \langle u, u\rangle \in \mathcal{U}_z\}) && \text{(by } (\bigstar) \text{ and } \langle x, z\rangle \in \bar{\mathsf{I}}^H \text{ (by } \langle z-1, z\rangle \in \bar{\mathsf{I}}^H)) \\ &\subseteq U_z && \text{(by } \psi^X_{a_z,\bar{\mathsf{I}}}(\mathcal{U}_{z-1}, U_{z-1}, \mathcal{U}_z, U_z)) \\ &\subseteq U_{z+1} = \cdots = U_y. && \text{(by the form of } J^{\mathcal{S}X}, \psi^X_{a_{z'},\mathsf{I}}(\mathcal{U}_{z'-1}, U_{z'-1}, \mathcal{U}_{z'}, U_{z'})) \end{aligned}$$

  - Case $a = \top$: We distinguish the following two sub-cases:
    - Case $\langle x, y\rangle \in \bar{\mathsf{I}}^H$: By the similar argument as Case $a = \bar{\mathsf{I}}$.
    - Case $\langle x, y\rangle \in \mathsf{I}^H$: By the similar argument as Case $a = \mathsf{I}$, we have $U_x = U_y$ and $\delta^J_{\mathsf{I}}(U_x) \subseteq U_x$, and thus $\delta^J_{\mathsf{I}}(U_x) \subseteq U_y$.
  - Case $a \in \{a, \overline{a} \mid a \in \mathbf{V}\}$: We distinguish the following sub-cases:
    - Case $\langle x, y\rangle \in \bar{\mathsf{I}}^H$: By $\langle x, y\rangle \in a^H \cap \bar{\mathsf{I}}^H = a^{\mathsf{G}(w)}$, we have $x = y - 1$ and $a_y = a$. Thus by $\psi^X_{a_y,\bar{\mathsf{I}}}(\mathcal{U}_{y-1}, U_{y-1}, \mathcal{U}_y, U_y)$, we have $\delta^J_a(U_x) \subseteq U_y$.
    - Case $a \notin X$: By $a^H = a^{\mathsf{G}(w)}$, we have $x = y - 1$ and $a_y = a$. By the form of $J^{\mathcal{S}X}$ with $\neg\psi^X_{a_y,\mathsf{I}}(\mathcal{U}_{y-1}, U_{y-1}, \mathcal{U}_y, U_y)$ (since $a_y \notin X$), we have $\psi^X_{a_y,\bar{\mathsf{I}}}(\mathcal{U}_{y-1}, U_{y-1}, \mathcal{U}_y, U_y)$. Hence, $\delta^J_a(U_x) \subseteq U_y$.
    - Case $\langle x, y\rangle \in \mathsf{I}^H$ and $a \in X$: By the similar argument as Case $a = \mathsf{I}$, we have $U_x = U_y$ (by $\langle x, y\rangle \in \mathsf{I}^H$) and $\delta^J_a(U_x) \subseteq U_x$ (by $a \in X$). Thus, $\delta^J_a(U_x) \subseteq U_y$. ◀

D Explicit proof of Lem. 3.4

(The following proof is almost the same as [19, Lem. 23] except that we consider letters-to-letters valuations instead of words-to-letters valuations.)

Proof of Lem. 3.4. By easy induction on t.

Case $t = x$: By the definition of $\mathfrak{v}^{\langle a_0,\dots,a_{n-1}\rangle}$, we have $\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(x) \Leftrightarrow a_i \dots a_{j-1} \in \hat{\mathfrak{v}}(x)$.

Case $t = \bot$: By $\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(\bot) \Leftrightarrow \mathsf{False} \Leftrightarrow a_i \dots a_{j-1} \in \hat{\mathfrak{v}}(\bot)$.

Case $t = \mathsf{I}$: By $\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(\mathsf{I}) \Leftrightarrow i = j \Leftrightarrow a_i \dots a_{j-1} \in \hat{\mathfrak{v}}(\mathsf{I})$.

Case $t = s \cup u$: By IH, we have $\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(s \cup u) \Leftrightarrow \ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(s) \vee \ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(u) \Leftrightarrow a_i \dots a_{j-1} \in \hat{\mathfrak{v}}(s) \vee a_i \dots a_{j-1} \in \hat{\mathfrak{v}}(u) \Leftrightarrow a_i \dots a_{j-1} \in \hat{\mathfrak{v}}(s \cup u)$.

Case $t = s^-$: By IH, we have $\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(s^-) \Leftrightarrow \ell_i \dots \ell_{j-1} \notin \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(s) \Leftrightarrow a_i \dots a_{j-1} \notin \hat{\mathfrak{v}}(s) \Leftrightarrow a_i \dots a_{j-1} \in \hat{\mathfrak{v}}(s^-)$.

Case $t = s \cdot u$: We have

$$
\begin{aligned}
&\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(s \cdot u) \\
&\Leftrightarrow \ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(s) \cdot \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(u) \\
&\qquad \text{(By the definition of } \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle} \text{ and that } \ell_i \dots \ell_{j-1} \text{ is a subword of } \ell_0 \dots \ell_{n-1}) \\
&\Leftrightarrow \bigvee_{i \le k \le j} (\ell_i \dots \ell_{k-1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(s) \wedge \ell_k \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(u)) \quad \text{(By concatenation)} \\
&\Leftrightarrow \bigvee_{i \le k \le j} (a_i \dots a_{k-1} \in \hat{\mathfrak{v}}(s) \wedge a_k \dots a_{j-1} \in \hat{\mathfrak{v}}(u)) \quad \text{(IH)} \\
&\Leftrightarrow a_i \dots a_{j-1} \in \hat{\mathfrak{v}}(s \cdot u). \quad \text{(By concatenation)}
\end{aligned}
$$

(Note that the last equivalence holds thanks to that $a_i, \dots, a_{j-1}$ are letters not words, cf. [19, Lem. 23].)

Case $t = s^*$: We have

$$
\begin{aligned}
&\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(s^*) \\
&\Leftrightarrow \ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(s)^* \\
&\qquad \text{(By the definition of } \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle} \text{ and that } \ell_i \dots \ell_{j-1} \text{ is a subword of } \ell_0 \dots \ell_{n-1}) \\
&\Leftrightarrow \exists m \in \mathbb{N}, \bigvee_{i=k_0 \le k_1 \le \dots \le k_m = j} \bigwedge_{l=1}^{m} (\ell_{k_{(l-1)}} \dots \ell_{k_l - 1} \in \hat{\mathfrak{v}}^{\langle a_0,\dots,a_{n-1}\rangle}(s)) \quad \text{(By Kleene star)} \\
&\Leftrightarrow \exists m \in \mathbb{N}, \bigvee_{i=k_0 \le k_1 \le \dots \le k_m = j} \bigwedge_{l=1}^{m} (a_{k_{(l-1)}} \dots a_{k_l - 1} \in \hat{\mathfrak{v}}(s)) \quad \text{(IH)} \\
&\Leftrightarrow a_i \dots a_{j-1} \in \hat{\mathfrak{v}}(s^*). \quad \text{(By Kleene star)} \quad \blacktriangleleft
\end{aligned}
$$

E Explicit proof for Sect. 4.1

(The following proof is almost the same as [18, Section VI] except that we consider RSUB instead of REL.)

A context-free grammar (CFG) $\mathfrak{C}$ over a finite set A is a tuple $\langle X, \mathcal{R}, \mathsf{s}\rangle$, where

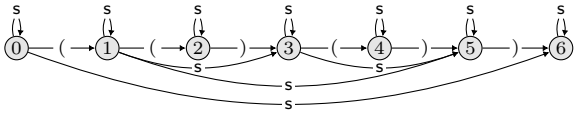
- X is a finite set of non-terminal labels s.t. $A \cap X = \emptyset$;
- $\mathcal{R}$ is a finite set of rewriting rules $x \leftarrow w$ of $x \in X$ and $w \in (A \cup X)^*$;
- $\mathsf{s} \in X$ is the start label.

The relation $x \vdash_{\mathfrak{C}} w$, where $x \in X$ and $w \in A^*$, is defined as the minimal relation closed under the following rule: for all $n \in \mathbb{N}$, $x, x_1, \dots, x_n \in X$ and $w_0, \dots, w_n, v_1, \dots, v_n \in A^*$, if $x \leftarrow w_0 x_1 w_1 \dots x_n w_n \in \mathcal{R}$, then $\dfrac{x_1 \vdash_{\mathfrak{C}} v_1 \quad \dots \quad x_n \vdash_{\mathfrak{C}} v_n}{x \vdash_{\mathfrak{C}} w_0 v_1 w_1 \dots v_n w_n}$. The *language* $[\mathfrak{C}]$ is defined by $[\mathfrak{C}] \triangleq \{w \in A^* \mid \mathsf{s} \vdash_{\mathfrak{C}} w\}$.

For a CFG $\mathfrak{C} = \langle X, \mathcal{R}, \mathsf{s}\rangle$ and a word $w = a_1 \dots a_n$, we write $\mathfrak{v}_{\mathfrak{C},w}$ for the following valuation on RSUB_n:

- $\mathfrak{v}_{\mathfrak{C},w}(x) = \{\langle i-1, i\rangle \mid i \in [1,n] \wedge x = a_i\}$ for $x \in A$;
- $\{\mathfrak{v}_{\mathfrak{C},w}(x)\}_{x\in X}$ are the minimal sets such that for all $x \in X$, $v \in (X \cup A)^*$, and $i, j \in [0, n]$, if $x \leftarrow v \in \mathcal{R}$, then $\dfrac{\langle i, j\rangle \in \hat{\mathfrak{v}}_{\mathfrak{C},w}(v)}{\langle i, j\rangle \in \mathfrak{v}_{\mathfrak{C},w}(x)}$.

For instance, if $A = \{(,)\}$, $\mathfrak{C} = \langle\{\mathsf{s}\}, \{\mathsf{s} \leftarrow (\mathsf{s})\mathsf{s},\ \mathsf{s} \leftarrow \mathsf{I}\}, \mathsf{s}\rangle$ (i.e., $[\mathfrak{C}]$ is the Dyck-1 language), and $w = (()())$, then $\mathfrak{v}_{\mathfrak{C},w}$ is of the following form:

$\mathfrak{v}_{\mathfrak{C},w} \quad =$ 

By construction, we have the following:

▶ **Lemma E.1** (cf. [18, Lem. 47]). *Let $\mathfrak{C} = \langle X, \mathcal{R}, \mathsf{s}\rangle$ be a CFG. Let $\mathcal{C}$ be s.t. $\{\mathfrak{v}_{\mathfrak{C},w}\} \subseteq \mathcal{C} \subseteq \mathsf{GREL}$. For all $x \in X$ and $w = a_0 \dots a_{n-1} \in A^*$, the following are equivalent:*

1. $x \vdash_{\mathfrak{C}} w$;
2. $\mathcal{C} \models (\bigwedge_{(x' \leftarrow w') \in \mathcal{R}} w' \le x') \to w \le x$;
3. $\langle 0, n\rangle \in \mathfrak{v}_{\mathfrak{C},w}(x)$.

Proof. 1⇒2: By induction on the derivation tree of $\vdash_{\mathfrak{C}}$. Let

$$\frac{x_1 \vdash_{\mathfrak{C}} v_1 \quad \dots \quad x_n \vdash_{\mathfrak{C}} v_m}{x \vdash_{\mathfrak{C}} w_0 v_1 w_1 \dots v_m w_m}$$

where $x \leftarrow w_0 x_1 w_1 \dots x_m w_m \in \mathcal{R}$. Let $\mathfrak{v} \in \mathcal{C}$ be any s.t. $\mathfrak{v} \models \bigwedge_{(x' \leftarrow w') \in \mathcal{R}} w' \le x'$. Particularly, we have $\mathfrak{v} \models w_0 x_1 w_1 \dots x_m w_m \le x$. By IH, we have $\mathfrak{v} \models v_k \le x_k$. Combining them yields $\mathfrak{v} \models w_0 v_1 w_1 \dots v_m w_m \le x$. (We only need axioms of monoids, so this holds for any Kleene algebras beyond GREL.)

2⇒3: By $\mathfrak{v}_{\mathfrak{C},w} \in \mathcal{C}$, $\mathfrak{v}_{\mathfrak{C},w} \models \bigwedge_{(x' \leftarrow w') \in \mathcal{R}} w' \le x'$, and $\langle 0, n\rangle \in \hat{\mathfrak{v}}_{\mathfrak{C},w}(w)$, we have $\langle 0, n\rangle \in \mathfrak{v}_{\mathfrak{C},w}(x)$ by 2.

3⇒1: By induction on the derivation tree induced from the definition of $\mathfrak{v}_{\mathfrak{C},w}$. Let

$$\frac{\langle i, j\rangle \in \hat{\mathfrak{v}}_{\mathfrak{C},w}(w_0 x_1 w_1 \dots x_m w_m)}{\langle i, j\rangle \in \mathfrak{v}_{\mathfrak{C},w}(x)}$$

where $x \leftarrow w_0 x_1 w_1 \dots x_n w_m \in \mathcal{R}$. Let $i = i_0 \le j_0 \le \dots \le i_m \le j_m = j - 1$ be s.t. $\langle i_k, j_k\rangle \in \hat{\mathfrak{v}}_{\mathfrak{C},w}(w_k)$ and $\langle j_{k-1}, i_k\rangle \in \mathfrak{v}_{\mathfrak{C},w}(x_k)$. By construction of $\mathfrak{v}_{\mathfrak{C},w}$, we have $w_k = a_{i_k} \dots a_{j_k}$. By IH, $x_k \vdash_{\mathfrak{C}} a_{j_{k-1}} \dots a_{i_k}$. Hence by definition of $\vdash_{\mathfrak{C}}$, we have $x \vdash_{\mathfrak{C}} a_i \dots a_{j-1}$. ◀

▶ **Lemma E.2** (cf. [18, Lem. 48]). *Let $\mathfrak{C} = \langle X, \mathcal{R}, \mathsf{s}\rangle$ be a CFG. Let $\mathcal{C}$ be s.t. $\{\mathfrak{v}_{\mathfrak{C},w} \mid w \in A^*\} \subseteq \mathcal{C} \subseteq \mathsf{GREL}$. Then we have*

$$[\mathfrak{C}] = A^* \quad \Leftrightarrow \quad \mathcal{C} \models \Big(\bigwedge_{(x \leftarrow w) \in \mathcal{R}} w \le x\Big) \to A^* \le \mathsf{s}.$$

Here, A^* denotes the term $(a_1 \cup \dots \cup a_n)^*$ in the right-hand side, where $A = \{a_1, \dots, a_n\}$.

Proof. We have:

$$
\begin{aligned}
&[\mathfrak{C}] = A^* \\
&\Leftrightarrow \quad \forall w' \in A^*, \mathsf{s} \vdash_{\mathfrak{C}} w' \\
&\Leftrightarrow \quad \forall w' \in A^*, \mathcal{C} \models (\bigwedge_{(x \leftarrow w) \in \mathcal{R}} w \leq x) \to w' \leq \mathsf{s} && \text{(By Lem. E.1)} \\
&\Leftrightarrow \quad \forall i \in \mathbb{N}, \mathcal{C} \models (\bigwedge_{(x \leftarrow w) \in \mathcal{R}} w \leq x) \to A^i \leq \mathsf{s} \\
&\Leftrightarrow \quad \mathcal{C} \models (\bigwedge_{(x \leftarrow w) \in \mathcal{R}} w \leq x) \to A^* \leq \mathsf{s}. \\
&\qquad\qquad (\Leftarrow\text{: By } \mathsf{GREL} \models A^i \leq A^* \Rightarrow\text{: Because } \mathsf{GREL} \text{ is } *\text{-continuous}) \quad \blacktriangleleft
\end{aligned}
$$

Proof of Lem. 4.6. By Lem. E.2. ◀

Furthermore, we give a tool to eliminate Kleene star.

▶ **Proposition E.3.** $\mathsf{RSUB} \models \bar{\mathsf{I}} = A\top \to A^* = \top$.

Proof. Let $n \in \mathbb{N}$ and $\mathfrak{v} \in \mathsf{RSUB}_n$. Let $i \in [1, n]$. By $\langle i-1, i-1\rangle \notin \hat{\mathfrak{v}}(\bar{\mathsf{I}}) = \hat{\mathfrak{v}}(A\top)$, we have $\langle i-1, i-1\rangle \notin \hat{\mathfrak{v}}(A)$. By $\langle i-1, i\rangle \in \hat{\mathfrak{v}}(\bar{\mathsf{I}}) = \hat{\mathfrak{v}}(A\top)$, we have $\langle i-1, i\rangle \in \hat{\mathfrak{v}}(A)$ (by $\langle i-1, i-1\rangle \notin \hat{\mathfrak{v}}(A)$). Thus $\hat{\mathfrak{v}}(A^*) = \{\langle i, j\rangle \mid 0 \leq i \leq j \leq n\} = \hat{\mathfrak{v}}(\top)$. ◀

Proof of Lem. 4.9. Because $\mathfrak{v}_{\mathfrak{C},w} \models \bar{\mathsf{I}} = A\top$, Lem. E.1 holds even if we can replace 2 with

$$\mathcal{C} \models (\bar{\mathsf{I}} = A\top \wedge \bigwedge_{(x \leftarrow w) \in \mathcal{R}} w \leq x) \to w' \leq x'.$$

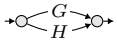
Thus for $\mathcal{C} \subseteq \mathsf{RSUB}$, by the same argument as Lem. E.2 with Prop. E.3, we have

$$[\mathfrak{C}] = A^* \quad \Leftrightarrow \quad \mathcal{C} \models (\bar{\mathsf{I}} = A\top \wedge \bigwedge_{(x \leftarrow w) \in \mathcal{R}} w \leq x) \to \top \leq \mathsf{s}. \quad \blacktriangleleft$$

F Explicit proof of Lem. 5.5

(The following proof is almost the same as [18, Prop. 8, 11] except that we consider GREL instead of REL. Similar arguments can also be found, e.g., in [1, 6, 20].)

We use the two notations for graphs, series-composition $(\cdot)$ and parallel-composition $(\cap)$:

$G \cdot H \triangleq$ $G \cap H \triangleq$  .

▶ **Proposition F.1.** *Let* $\mathfrak{v} \in \mathsf{GREL}$ *and* G, H *be graphs.*

$$\hat{\mathfrak{v}}(G \cap H) = \hat{\mathfrak{v}}(G) \cap \hat{\mathfrak{v}}(H) \qquad \text{(Prop. F.1}\cap\text{)}$$

$$\hat{\mathfrak{v}}(G \cdot H) = \hat{\mathfrak{v}}(G) \cdot \hat{\mathfrak{v}}(H) \qquad \text{(Prop. F.1}\cdot\text{)}$$

Proof. (Prop. F.1$\cap$): It suffices to prove that for every x, y,

$$\exists f, f\colon (G \cap H) \longrightarrow \mathsf{G}(\mathfrak{v}, x, y) \;\Leftrightarrow\; \exists f_G, f_H,\; f_G\colon G \longrightarrow \mathsf{G}(\mathfrak{v}, x, y) \wedge f_H\colon H \longrightarrow \mathsf{G}(\mathfrak{v}, x, y).$$

$\Rightarrow$: By letting $f_G = \{\langle x', f(x')\rangle \mid x' \in |G|\}$ and $f_H = \{\langle x', f(x')\rangle \mid x' \in |H|\}$. $\Leftarrow$: By letting $f = f_G \cup f_H$. Note that $f_G(1^G) = x = f_H(1^H)$ and $f_G(2^G) = y = f_H(2^H)$; so f is indeed a map.

(Prop. F.1·): It suffices to prove that for every x, y,

$$\exists f, f\colon (G \cdot H) \longrightarrow \mathsf{G}(\mathfrak{v}, x, y) \;\Leftrightarrow\; \exists z, \exists f_G, f_H,\ f_G\colon G \longrightarrow \mathsf{G}(\mathfrak{v}, x, z) \wedge f_H\colon H \longrightarrow \mathsf{G}(\mathfrak{v}, z, y).$$

$\Rightarrow$: By letting $z = f(2^G)$, $f_G = \{\langle x', f(x')\rangle \mid x' \in |G|\}$, and $f_H = \{\langle x', f(x')\rangle \mid x' \in |H|\}$. $\Leftarrow$: By letting $f = f_G \cup f_H$. Note that $f_G(2^G) = z = f_H(1^H)$; so f is indeed a map. ◀

Proof of Lem. 5.5. By easy induction on t.

Case $t = x$ where $x \in \tilde{\mathbf{V}} = \{a, \overline{a} \mid a \in \mathbf{V}\} \cup \{\bar{\mathsf{I}}, \top\}$: For every $\langle x, y\rangle \in \hat{\mathfrak{v}}(\top)$, we have

$$\begin{aligned}
\langle x, y\rangle \in \hat{\mathfrak{v}}(a) \quad &\Leftrightarrow \quad \to\!\circ\!\xrightarrow{\;a\;}\!\circ\!\to \;\longrightarrow \mathsf{G}(\mathfrak{v}, x, y) \\
&\Leftrightarrow \quad \langle x, y\rangle \in \hat{\mathfrak{v}}(\,\to\!\circ\!\xrightarrow{\;a\;}\!\circ\!\to\,) && \text{(Def. of } \hat{\mathfrak{v}}) \\
&\Leftrightarrow \quad \langle x, y\rangle \in \hat{\mathfrak{v}}(\mathcal{G}(a)). && \text{(Def. of } \mathcal{G})
\end{aligned}$$

Case $t = \mathsf{I}$: For every $\langle x, y\rangle \in \hat{\mathfrak{v}}(\top)$, we have

$$\begin{aligned}
\langle x, y\rangle \in \hat{\mathfrak{v}}(\mathsf{I}) \quad &\Leftrightarrow \quad \to\!\circ\!\xrightarrow{\;\mathsf{I}\;}\!\circ\!\to \;\longrightarrow \mathsf{G}(\mathfrak{v}, x, y) \\
&\Leftrightarrow \quad \langle x, y\rangle \in \hat{\mathfrak{v}}(\,\to\!\circ\!\xrightarrow{\;\mathsf{I}\;}\!\circ\!\to\,) && \text{(Def. of } \hat{\mathfrak{v}}) \\
&\Leftrightarrow \quad \langle x, y\rangle \in \hat{\mathfrak{v}}(\,\to\!\circ\!\to\,) && (\mathsf{I}^{\mathsf{G}(\mathfrak{v}, x, y)} \text{ is the identity relation}) \\
&\Leftrightarrow \quad \langle x, y\rangle \in \hat{\mathfrak{v}}(\mathcal{G}(\mathsf{I})). && \text{(Def. of } \mathcal{G})
\end{aligned}$$

Case $t = \bot$: For every $\langle x, y\rangle \in \hat{\mathfrak{v}}(\top)$, we have

$$\langle x, y\rangle \in \hat{\mathfrak{v}}(\bot) \Leftrightarrow \mathsf{false} \Leftrightarrow \langle x, y\rangle \in \hat{\mathfrak{v}}(\emptyset) \Leftrightarrow \langle x, y\rangle \in \hat{\mathfrak{v}}(\mathcal{G}(\bot)). \qquad \text{(Def. of } \hat{\mathfrak{v}} \text{ and } \mathcal{G})$$

Case $t = s \cdot u$:

$$\begin{aligned}
\hat{\mathfrak{v}}(s \cdot u) \quad &= \quad \hat{\mathfrak{v}}(s) \cdot \hat{\mathfrak{v}}(u) && \text{(Def. of } \hat{\mathfrak{v}}) \\
&= \quad \hat{\mathfrak{v}}(\mathcal{G}(s)) \cdot \hat{\mathfrak{v}}(\mathcal{G}(u)) && \text{(IH)} \\
&= \quad \bigcup_{G \in \mathcal{G}(s)} \bigcup_{H \in \mathcal{G}(u)} \hat{\mathfrak{v}}(G) \cdot \hat{\mathfrak{v}}(H) && (\cdot \text{ is distributive w.r.t. } \cup) \\
&= \quad \bigcup_{G \in \mathcal{G}(s)} \bigcup_{H \in \mathcal{G}(u)} \hat{\mathfrak{v}}(G \cdot H) && \text{(Equation (Prop. F.1}\cdot\text{))} \\
&= \quad \hat{\mathfrak{v}}(\mathcal{G}(s \cdot u)). && \text{(Def. of } \mathcal{G})
\end{aligned}$$

Case $t = s \cap u$:

$$\begin{aligned}
\hat{\mathfrak{v}}(s \cap u) \quad &= \quad \hat{\mathfrak{v}}(s) \cap \hat{\mathfrak{v}}(u) && \text{(Def. of } \hat{\mathfrak{v}}) \\
&= \quad \hat{\mathfrak{v}}(\mathcal{G}(s)) \cap \hat{\mathfrak{v}}(\mathcal{G}(u)) && \text{(IH)} \\
&= \quad \bigcup_{G \in \mathcal{G}(s)} \bigcup_{H \in \mathcal{G}(u)} (\hat{\mathfrak{v}}(G) \cap \hat{\mathfrak{v}}(H)) && (\cap \text{ is distributive w.r.t. } \cup) \\
&= \quad \bigcup_{G \in \mathcal{G}(s)} \bigcup_{H \in \mathcal{G}(u)} \hat{\mathfrak{v}}(G \cap H) && \text{(Equation (Prop. F.1}\cap\text{))} \\
&= \quad \hat{\mathfrak{v}}(\mathcal{G}(s \cap u)). && \text{(Def. of } \mathcal{G})
\end{aligned}$$

Case $t = s \cup u$:

$$\begin{aligned}
\hat{\mathfrak{v}}(s \cup u) \quad &= \quad \hat{\mathfrak{v}}(s) \cup \hat{\mathfrak{v}}(u) && \text{(Def. of } \hat{\mathfrak{v}}) \\
&= \quad \hat{\mathfrak{v}}(\mathcal{G}(s)) \cup \hat{\mathfrak{v}}(\mathcal{G}(u)) && \text{(IH)} \\
&= \quad \hat{\mathfrak{v}}(\mathcal{G}(s) \cup \mathcal{G}(u)) = \hat{\mathfrak{v}}(\mathcal{G}(s \cup u)). && \text{(Def. of } \mathcal{G})
\end{aligned}$$

867 Case $t = s^*$:

868 $$\hat{\mathfrak{v}}(s^*) = \bigcup_{n\in\mathbb{N}} \hat{\mathfrak{v}}(s^n) = \bigcup_{n\in\mathbb{N}} \hat{\mathfrak{v}}(s)^n \qquad \text{(Def. of } \hat{\mathfrak{v}}\text{)}$$

869 $$= \bigcup_{n\in\mathbb{N}} \hat{\mathfrak{v}}(\mathcal{G}(s))^n \qquad \text{(IH)}$$

870 $$= \bigcup_{n\in\mathbb{N}} \bigcup_{G_1,\ldots,G_n\in\mathcal{G}(s)} \hat{\mathfrak{v}}(G_1)\cdot\ldots\cdot\hat{\mathfrak{v}}(G_n) \qquad (\cdot \text{ is distributive w.r.t. } \cup)$$

871 $$= \bigcup_{n\in\mathbb{N}} \bigcup_{G_1,\ldots,G_n\in\mathcal{G}(s)} \hat{\mathfrak{v}}(G_1\cdot\ldots\cdot G_n) \qquad \text{(Equation (Prop. F.1}\cdot\text{))}$$

872 $$= \bigcup_{n\in\mathbb{N}} \hat{\mathfrak{v}}(\mathcal{G}(s^n) = \hat{\mathfrak{v}}(\mathcal{G}(s^*)). \qquad \text{(Def. of } \mathcal{G}\text{)}$$

873

874 ◀