Words-to-Letters Valuations for Language Kleene Algebras with Variable Complements

Yoshiki Nakamura
Tokyo Institute of Technology, Japan
nakamura.yoshiki.ny@gmail.com

Ryoma Sin'ya
Akita University, Japan
ryoma@math.akita-u.ac.jp

We investigate the equational theory of Kleene algebra terms with *variable complements*—(language) complement where it applies only to variables—w.r.t. languages. While the equational theory w.r.t. languages coincides with the language equivalence (under the standard language valuation) for Kleene algebra terms, this coincidence is broken if we extend the terms with complements. In this paper, we prove the decidability of some fragments of the equational theory: the universality problem is coNP-complete, and the inequational theory $t \leq s$ is coNP-complete when t does not contain Kleene-star. To this end, we introduce *words-to-letters valuations*; they are sufficient valuations for the equational theory and ease us in investigating the equational theory w.r.t. languages. Additionally, we prove that for words with variable complements, the equational theory coincides with the word equivalence.

1 Introduction

Kleene algebra (KA) [3, 6] is an algebraic system for regular expressions consisting of union ($\cup$), composition ($\cdot$), Kleene-star ($\_^*$), emptiness ($\bot$), and identity (I). In this paper, we consider KAs *w.r.t. languages* (a.k.a., language models of KAs, language KAs). Interestingly, the equational theory of KAs w.r.t. languages coincides with the language equivalence under the standard language valuation (see also, e.g., [1, 11]): for all KA terms (i.e., regular expressions) t, s, we have

$$\mathsf{LANG} \models t = s \quad\iff\quad [t] = [s]. \tag{†}$$

Here, we write $\mathsf{LANG} \models t = s$ if the equation $t = s$ holds for all language models (i.e., each variable x maps to not only the singleton language $\{x\}$ but also any languages); we write $[u]$ for the language of a regular expression u (i.e., each variable x maps to the singleton language $\{x\}$). Since the valuation $[\_]$ is an instance of valuations in LANG, the direction $\Longrightarrow$ is trivial (this direction always holds even if we extend KA terms with some extra operators). The direction $\Longleftarrow$ is a consequence of the completeness of KAs (see Appendix A for an alternative proof not relying on the completeness of KAs). However, the direction $\Longleftarrow$ fails in general when we extend KA terms with extra operators. Namely, the equational theory w.r.t. languages does not coincide with the language equivalence in general (see below for complements). The equational theory of KAs with some operators w.r.t. languages was studied, e.g., with reverse [2], with tests [7] (where languages are of guarded strings, not words), with intersection [1], and with universality ($\top$) [11]. Nevertheless, to the best of authors' knowledge, *variable complements* (and even complements) w.r.t. languages has not yet been investigated, while those w.r.t. binary relations were studied, e.g., in [10] (for complements; cf. Tarski's calculus of relations [13]) and [9] (for variable complements).

In this paper, we investigate the equational theory of KA terms with *variable complements* (x^-)—(language) complement, where it applies only to variables (we use x to denote variables)—w.r.t. lan-

Zs. Gazdag, Sz. Iván, G. Kovásznai (Eds.): 16th International Conference on Automata and Formal Languages (AFL 2023)
EPTCS 386, 2023, pp. 185–199, doi:10.4204/EPTCS.386.15

guages. For KA terms with variable complements, (†) fails. The following is an example:

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

(For $\mathsf{LANG} \not\models$, consider a valuation such that x^- maps to the language $\{a\}$.) As the example above (see also Remark 1, for more examples) shows, the equational theory of KAs with variable complements w.r.t. languages significantly differs from the language equivalence under the standard language valuation. While the language equivalence problem of KA terms with variable complements is decidable in PSPACE by a standard automata construction [14] (and hence, PSPACE-complete [5, 8, 12]), it remains whether the equational theory w.r.t. languages is decidable.

We prove the decidability and complexity of some fragments of the equational theory of KA terms with variable complements w.r.t. languages: the universality problem is coNP-complete (Cor. 18), and the inequational theory $t \le s$ is coNP-complete when t does not contain Kleene-star (Cor. 29). To this end, we introduce *words-to-letters valuations*. Words-to-letters valuations are sufficient for the equational theory of KA terms with variable complements w.r.t. languages (words-to-letters-valuation property; Cor. 30): for all terms t, s, if there is some language valuation such that it refutes $t = s$, there is some words-to-letters valuation such that it refutes $t = s$. This property eases us in investigating the equational theory w.r.t. languages.

Additionally, by using words-to-letters valuations, we prove a completeness theorem for words with variable complements: the equational theory coincides with the word equivalence (Thm. 32). A limitation of words-to-letters valuations is that the number of letters is not bounded; so, they cannot apply to the equational theory over LANG_n (language models over sets of cardinality at most a natural number n). Nevertheless, by giving another valuation, we can show the coincidence for one-variable words (Thm. 36). We leave open for the many-variable words.

Outline

In Sect. 2, we briefly give basic definitions, including the syntax and semantics of KA terms with variable complements. Additionally, we give languages for KA terms with variable complements (Sect. 2.3). In Sects. 3–5, we consider fragments of the equational theory of KA terms with variable complements w.r.t. languages, step-by-step. In Sect. 3, we consider the identity inclusion problem ($\mathsf{LANG} \models \mathsf{I} \le t$?). This problem is relatively easy but contains the coNP-hardness result (Cor. 6). In Sect. 4, we consider the variable inclusion problem ($\mathsf{LANG} \models x \le t$?) and the universality problem ($\mathsf{LANG} \models \top \le t$?). In Sect. 5, we consider the word inclusion problem ($\mathsf{LANG} \models w \le t$?). This section proceeds in the same way as Sect. 4, thanks to words-to-letters valuations (Def. 21). Consequently, the inequational theory $t \le s$ is coNP-complete when t does not contain Kleene-star (Cor. 29). Additionally, we prove the words-to-letters valuation property (Cor. 30) for the equational theory of (full) KA terms with variable complements w.r.t. languages. In Sect. 6, we consider the equational theory of words with variable complements and show a completeness theorem (Thm. 32). Sect. 7 concludes this paper.

2 Preliminaries

We write $\mathbb{N}$ for the set of non-negative integers. For a set X, we write $\#(X)$ for the cardinality of X and $\wp(X)$ for the power set of X.

For a set X (of letters), we write X^* for the set of words over X: finite sequences of elements of 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 composition $L \cdot K$ and the Kleene star L^* is defined by:

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

2.1 Syntax: KA terms with variable complements

Let $\mathbf{V}$ be a set of variables. The set of Kleene algebra (KA) terms with variable complements (x^-) is defined by the following grammar:

$$\mathbf{T} \ni t, s, u \quad ::= \quad x \mid \mathsf{I} \mid \bot \mid t \cdot s \mid t \cup s \mid t^* \mid x^- \qquad (x \in \mathbf{V})$$

We use parentheses in ambiguous situations. We often abbreviate $t \cdot s$ to ts. We write $\top$ for the term $x \cup x^-$, where x is any variable.

An *equation* $t = s$ is a pair of terms. An *inequation* $t \leq s$ is an abbreviation of the equation $t \cup s = s$.

2.2 Semantics: language models

Consider the signature $S \triangleq \{\mathsf{I}_{(0)}, \bot_{(0)}, \cdot_{(2)}, \cup_{(2)}, \_^*{}_{(1)}, \_^-{}_{(1)}\}$. An *$S$-algebra* A is a tuple $\langle |A|, \{f^A\}_{f_{(k)} \in S} \rangle$, where $|A|$ is a non-empty set and $f^A \colon |A|^k \to |A|$ is a k-ary map for each $f_{(k)} \in S$. A *valuation* $\mathfrak{v}$ on an S-algebra A is a map $\mathfrak{v} \colon \mathbf{V} \to |A|$. For a valuation $\mathfrak{v}$, we write $\hat{\mathfrak{v}} \colon \mathbf{T} \to |A|$ for the unique homomorphism extending $\mathfrak{v}$.

The *language model* A over a set X is an S-algebra such that $|A| = \wp(X^*)$ and for all $L, K \subseteq X^*$,

$$\mathsf{I}^A = \{\mathsf{I}\} \qquad L \cdot^A K = L \cdot K \qquad L^{*^A} = L^*$$
$$\bot^A = \emptyset \qquad L \cup^A K = L \cup K \qquad L^{-^A} = X^* \setminus L.$$

We write LANG for the class of all language models. A *language valuation* over a set X is a valuation on some language model over X. For an equation $t = s$, we let

$$\mathsf{LANG} \models t = s \quad \overset{\triangle}{\Longleftrightarrow} \quad \hat{\mathfrak{v}}(t) = \hat{\mathfrak{v}}(s) \text{ holds for all language valuations } \mathfrak{v}.$$

The *equational theory w.r.t. languages* is the set of all equations $t = s$ such that $\mathsf{LANG} \models t = s$.

Additionally, the *language* $[t] \subseteq \mathbf{V}^*$ of a term t is defined by:

$$[x] \triangleq \{x\} \qquad [\mathsf{I}] \triangleq \{\mathsf{I}\} \qquad [t \cdot s] \triangleq [t] \cdot [s] \qquad [t^*] \triangleq [t]^*$$
$$[\bot] \triangleq \emptyset \qquad [t \cup s] \triangleq [t] \cup [s] \qquad [t^-] \triangleq \mathbf{V}^* \setminus [t].$$

By definition, we have $[t] = \hat{\mathfrak{v}}(t)$ if $\mathfrak{v}$ is the valuation on the language model over the set $\mathbf{V}$ defined by $\mathfrak{v}(x) = \{x\}$ for all $x \in \mathbf{V}$. Hence, for all t, s, we have

$$\mathsf{LANG} \models t = s \quad \Longrightarrow \quad [t] = [s]. \qquad (\ddagger)$$

Remark 1. The converse direction fails[1]; for example, when $x \neq y$,

$$\mathsf{LANG} \not\models y \leq x^- \qquad\qquad [y] \subseteq [x^-].$$

[1] The failure can be also shown by that the universality $\top$ can be expressed by $x \cup x^-$; see also [11, Remark. 3.6].

Here $t \leq s$ denotes the equation $t \cup s = s$ (so, indeed, an equation). For $\mathsf{LANG} \not\models y \leq x^-$, consider a language valuation $\mathfrak{v}$ such that $a \in \mathfrak{v}(x)$ and $a \in \mathfrak{v}(y)$; then we have $a \in \hat{\mathfrak{v}}(y) \setminus \hat{\mathfrak{v}}(x^-)$. $[y] \subseteq [x^-]$ is shown by $[y] = \{y\} \subseteq \mathbf{V}^* \setminus \{x\} = [x^-]$. More generally, for any word w over $\mathbf{V}$ such that $w \neq x$,

$$\mathsf{LANG} \not\models w \leq x^- \qquad\qquad [w] \subseteq [x^-].$$

Moreover, for example, there are the following examples (for $\mathsf{LANG} \not\models$, consider a valuation such that both x and y map to the language $X^* \setminus \{a\}$, where X is a set and $a \in X$):

$$\begin{array}{ll} \mathsf{LANG} \not\models x^- = x^- \cdot x^- & \qquad [x^-] = [x^- \cdot x^-] \\ \mathsf{LANG} \not\models \top = x^- \cdot y^- & \qquad [\top] = [x^- \cdot y^-] \\ \mathsf{LANG} \not\models \top = x^- \cup y^- & \qquad [\top] = [x^- \cup y^-]. \end{array}$$

As the examples above show, for KA terms with variable complements, the equational theory w.r.t. languages ($\mathsf{LANG} \models t = s$?) significantly differs from the language equivalence problem ($[t] = [s]$?).

In the sequel, we focus on the equational theory w.r.t. languages and investigate its fragments. We prepare a useful tool (Lem. 2), which enables us to decompose terms into languages of words.

2.3 Languages for KA terms with variable complements

Let $\mathbf{V}' = \{x, x^- \mid x \in \mathbf{V}\}$. For a term t, we write $[t]_{\mathbf{V}'}$ for the language of t where t is viewed as the regular expression over $\mathbf{V}'$. Each word over $\mathbf{V}'$ is a term such that both the union ($\cup$) and the Kleene-star ($\_^*$) do not occur. Note that $[x^-]_{\mathbf{V}'} = \{x^-\}$, cf. $[x^-] = \mathbf{V}^* \setminus \{x\}$. For a language valuation $\mathfrak{v}$ and a language L over $\mathbf{V}'$, we define

$$\hat{\mathfrak{v}}(L) \quad \triangleq \quad \bigcup_{w \in L} \hat{\mathfrak{v}}(w).$$

By using the distributive law of $\cdot$ w.r.t. $\cup$, for all languages L, K and language valuations $\mathfrak{v}$, we have:

$$\hat{\mathfrak{v}}(L \cup K) = \hat{\mathfrak{v}}(L) \cup \hat{\mathfrak{v}}(K) \qquad \hat{\mathfrak{v}}(L \cdot K) = \hat{\mathfrak{v}}(L) \cdot \hat{\mathfrak{v}}(K) \qquad \hat{\mathfrak{v}}(L^*) = \hat{\mathfrak{v}}(L)^*.$$

We can decompose each term t to the set $[t]_{\mathbf{V}'}$ of words over $\mathbf{V}'$ as follows:

Lemma 2. *Let $\mathfrak{v}$ be a language valuation. For all terms t, we have*

$$\hat{\mathfrak{v}}(t) \quad = \quad \hat{\mathfrak{v}}([t]_{\mathbf{V}'}).$$

Proof. By easy induction on t using the equations above. Case $t = x, x^-, \mathsf{I}$: Clear, by $[t]_{\mathbf{V}'} = \{t\}$. Case $t = \bot$: By $\hat{\mathfrak{v}}(\bot) = \emptyset = \hat{\mathfrak{v}}([\bot]_{\mathbf{V}'})$. Case $t = s \cup u$, Case $t = s \cdot u$, Case $t = s^*$: By IH with the equations above. For example, when $t = s \cdot u$, we have

$$\begin{aligned} \hat{\mathfrak{v}}(s \cdot u) = \hat{\mathfrak{v}}(s) \cdot \hat{\mathfrak{v}}(u) &= \hat{\mathfrak{v}}([s]_{\mathbf{V}'}) \cdot \hat{\mathfrak{v}}([u]_{\mathbf{V}'}) && \text{(IH)} \\ &= \hat{\mathfrak{v}}([s]_{\mathbf{V}'} \cdot [u]_{\mathbf{V}'}) = \hat{\mathfrak{v}}([s \cdot u]_{\mathbf{V}'}). \end{aligned}$$

□

3 The identity inclusion problem

We first consider the *identity inclusion problem* w.r.t. languages:

Given a term t, does $\mathsf{LANG} \models \mathsf{I} \leq t$?

This problem is relatively easily solvable. Since $\mathsf{LANG} \models \mathsf{I} \leq t$ iff $\mathsf{I} \in \hat{\mathfrak{v}}(t)$ for all language valuation $\mathfrak{v}$, it suffices to consider the membership of the empty word I. We use the following facts.

Proposition 3. *For all languages L, K, we have:*

$$\begin{aligned} \mathsf{I} \in L \cup K &\iff \mathsf{I} \in L \vee \mathsf{I} \in K \\ \mathsf{I} \in L \cdot K &\iff \mathsf{I} \in L \wedge \mathsf{I} \in K \\ \mathsf{I} \in L^{*} &\iff \mathsf{True}. \end{aligned}$$

Proof. Clear, by definition. □

Lemma 4. *Let $\mathfrak{v}, \mathfrak{v}'$ be language valuations. Assume that for all variables x, $\mathsf{I} \in \mathfrak{v}(x)$ iff $\mathsf{I} \in \mathfrak{v}'(x)$. For all terms t,*

$$\mathsf{I} \in \hat{\mathfrak{v}}(t) \iff \mathsf{I} \in \hat{\mathfrak{v}}'(t).$$

Proof. By easy induction on t using Prop. 3. Case $t = x, x^{-}$: Clear by the assumption. Case t is a constant: Trivial. For inductive cases, e.g., Case $t = s \cup u$: By using Prop. 3, we have

$$\begin{aligned} \mathsf{I} \in \hat{\mathfrak{v}}(s \cup u) \iff \mathsf{I} \in \hat{\mathfrak{v}}(s) \vee \mathsf{I} \in \hat{\mathfrak{v}}(u) &\iff \mathsf{I} \in \hat{\mathfrak{v}}'(s) \vee \mathsf{I} \in \hat{\mathfrak{v}}'(u) && \text{(IH)} \\ &\iff \mathsf{I} \in \hat{\mathfrak{v}}'(s \cup u). \end{aligned}$$

(Similarly for the other inductive cases.) □

By using Lem. 4, it suffices to consider a finite number of valuations, as follows.

Theorem 5. *For all terms t, the following are equivalent:*

1. $\mathsf{LANG} \models \mathsf{I} \leq t$ *(i.e., $\hat{\mathfrak{v}}(\mathsf{I}) \subseteq \hat{\mathfrak{v}}(t)$, for all language valuations $\mathfrak{v}$);*
2. $\hat{\mathfrak{v}}(\mathsf{I}) \subseteq \hat{\mathfrak{v}}(t)$, *for all language valuations $\mathfrak{v}$ over the empty set s.t. for all x, $\mathfrak{v}(x) \subseteq \{\mathsf{I}\}$.*

Proof. 1⇒2: Trivial. 2⇒1: We prove the contraposition. By $\mathsf{LANG} \not\models \mathsf{I} \leq t$, there is a language valuation $\mathfrak{v}$ s.t. $\hat{\mathfrak{v}}(\mathsf{I}) \not\subseteq \hat{\mathfrak{v}}(t)$ (i.e., $\mathsf{I} \notin \hat{\mathfrak{v}}(t)$). Let $\mathfrak{v}^{\langle\rangle}$ be the language valuation over the empty set defined by:

$$\mathfrak{v}^{\langle\rangle}(x) \quad \triangleq \quad \{\mathsf{I} \mid \mathsf{I} \in \mathfrak{v}(x)\}.$$

Then by Lem. 4, $\mathsf{I} \notin \hat{\mathfrak{v}}^{\langle\rangle}(t)$ holds; thus, we have $\hat{\mathfrak{v}}^{\langle\rangle}(\mathsf{I}) \not\subseteq \hat{\mathfrak{v}}^{\langle\rangle}(t)$. □

Corollary 6. *The identity inclusion problem (given a term t, does $\mathsf{LANG} \models \mathsf{I} \leq t$?) is decidable and coNP-complete for KA terms with variable complements.*

Proof. (in coNP): Thm. 5 induces the following non-deterministic polynomial algorithm:

1. Pick up a language valuation $\mathfrak{v}$ over the empty set s.t. for all x, $\mathfrak{v}(x) \subseteq \{\mathsf{I}\}$, non-deterministically.
2. If $\hat{\mathfrak{v}}(\mathsf{I}) \not\subseteq \hat{\mathfrak{v}}(t)$, then return True; otherwise return False.

Then $\mathsf{LANG} \not\models \mathsf{I} \leq t$ if some execution returns True; and $\mathsf{LANG} \models \mathsf{I} \leq t$ otherwise. Hence, the identity inclusion problem is decidable in coNP (as its complemented problem is in NP).

(coNP-hard): Because this problem subsumes the validity problem of propositional formulas in disjunctive normal form, which is a well-known coNP-complete problem [4]. More precisely, given a propositional formula φ in disjunctive normal form, let t be the term obtained from φ by replacing each conjunction $\wedge$ with $\cdot$ and each disjunction $\vee$ with $\cup$ (where we map each positive literal x to the variable

x and each negative literal x^- to the complemented variable x^-); for example, if $\varphi = (x \wedge y^-) \vee (y \vee x^-)$, then $t = (x \cdot y^-) \cup (y \cup x^-)$. Then, for all language valuations $\mathfrak{v}$ (over the empty set s.t. for all $x, \mathfrak{v}(x) \subseteq \{\mathsf{I}\}$), we have: $\hat{\mathfrak{v}}(\mathsf{I}) \subseteq \hat{\mathfrak{v}}(t)$ holds iff φ is True on the valuation $\mathfrak{v}'$, where $\mathfrak{v}'$ is the map mapping each x to True if $\mathsf{I} \in \mathfrak{v}(x)$ and False otherwise. Thus by Thm. 5, $\mathsf{LANG} \models \mathsf{I} \leq t$ iff φ is valid. Hence, the identity inclusion problem is coNP-hard. $\square$

Remark 7. Under the standard language valuation, the identity inclusion problem—given a term t, does $[\mathsf{I}] \subseteq [t]$? (i.e., does $\mathsf{I} \in [t]$?)—is decidable in P (because we can compute "$\mathsf{I} \in [t]$?" by induction on t, as $\mathsf{I} \notin [x]$ and $\mathsf{I} \in [x^-]$ for every variable x). Hence, for KA terms with variable complements, the identity inclusion problem w.r.t. languages is strictly harder than that under the standard language valuation unless P = NP.

4 The variable inclusion problem and the universality problem

Next, we consider the *variable inclusion problem*:

Given a variable x and a term t, does $\mathsf{LANG} \models x \leq t$?

In the identity inclusion problem, if $w \in \hat{\mathfrak{v}}(\mathsf{I}) \setminus \hat{\mathfrak{v}}(t)$, then $w = \mathsf{I}$ should hold; so it suffices to consider the membership of the empty word I. However, in the variable inclusion problem, this situation changes: if $w \in \hat{\mathfrak{v}}(x) \setminus \hat{\mathfrak{v}}(t)$, then w is possibly any word.

Nevertheless, we can overcome the problem above for KA terms with variable complements; more precisely, from a language valuation $\mathfrak{v}$ s.t. $w \in \hat{\mathfrak{v}}(x) \setminus \hat{\mathfrak{v}}(t)$ for some word w, we can construct a language valuation $\mathfrak{v}'$ s.t. $\ell \in \hat{\mathfrak{v}}'(x) \setminus \hat{\mathfrak{v}}'(t)$ for some *letter* ℓ. If such $\mathfrak{v}'$ can be constructed from $\mathfrak{v}$, then considering the membership of letters suffices because we have

$$\begin{aligned}
\mathsf{LANG} \not\models x \leq t &\iff w \in \hat{\mathfrak{v}}(x) \setminus \hat{\mathfrak{v}}(t) \text{ for some language valuation } \mathfrak{v} \text{ and word } w \qquad \text{(By definition)}\\
&\iff \ell \in \hat{\mathfrak{v}}'(x) \setminus \hat{\mathfrak{v}}'(t) \text{ for some language valuation } \mathfrak{v}' \text{ and } \textit{letter } \ell\\
&\qquad (\Longrightarrow\text{: By the condition of } \mathfrak{v}'.\ \Longleftarrow\text{: Trivial by letting } \mathfrak{v} = \mathfrak{v}'.)
\end{aligned}$$

Such a language valuation $\mathfrak{v}'$ can be defined as follows:

Definition 8. For a language valuation $\mathfrak{v}$ over a set X and a word w over X, the language valuation $\mathfrak{v}^w$ over the set $\{\ell\}$ (where ℓ is a letter) is defined as follows:

$$\mathfrak{v}^w(x) \quad \triangleq \quad \{\mathsf{I} \mid \mathsf{I} \in \mathfrak{v}(x)\} \cup \{\ell \mid w \in \mathfrak{v}(x)\}.$$

In the following, we prove that $\mathfrak{v}^w$ satisfies the condition of $\mathfrak{v}'$ above, that is, the following conditions:

- $w \in \hat{\mathfrak{v}}(x) \Longrightarrow \ell \in \hat{\mathfrak{v}}^w(x)$;
- $w \notin \hat{\mathfrak{v}}(t) \Longrightarrow \ell \notin \hat{\mathfrak{v}}^w(t)$.

The first condition is clear by the definition of $\mathfrak{v}^w$. We prove the second condition in Lem. 10. We prepare the following fact:

Proposition 9 (cf. Prop. 3)**.** *For all languages L, K and letters a, we have:*

$$\begin{aligned}
a \in L \cup K &\iff a \in L \vee a \in K\\
a \in L \cdot K &\iff (a \in L \wedge \mathsf{I} \in K) \vee (\mathsf{I} \in L \wedge a \in K)\\
a \in L^* &\iff a \in L.
\end{aligned}$$

Proof. Clear, by definition. □

Lemma 10 (cf. Lem. 4)**.** *Let $\mathfrak{v}$ be a language valuation and w be a word. For all terms t, we have:*

$$\ell \in \hat{\mathfrak{v}}^{w}(t) \quad \Longrightarrow \quad w \in \hat{\mathfrak{v}}(t).$$

Proof. By induction on t.

Case $t = x, x^{-}$: By the construction of $\mathfrak{v}^{w}$, $\ell \in \hat{\mathfrak{v}}^{w}(x)$ iff $w \in \hat{\mathfrak{v}}(x)$. (Hence, we also have $\ell \in \hat{\mathfrak{v}}^{w}(x^{-})$ iff $w \in \hat{\mathfrak{v}}(x^{-})$.)

Case $t = \bot, \mathsf{I}$: By $\ell \notin \hat{\mathfrak{v}}^{w}(t)$.

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

$$\begin{aligned} \ell \in \hat{\mathfrak{v}}^{w}(s) \cup \hat{\mathfrak{v}}^{w}(u) &\Longleftrightarrow \ell \in \hat{\mathfrak{v}}^{w}(s) \vee \ell \in \hat{\mathfrak{v}}^{w}(u) && \text{(Prop. 9)} \\ &\Longrightarrow w \in \hat{\mathfrak{v}}(s) \vee w \in \hat{\mathfrak{v}}(u) && \text{(IH)} \\ &\Longrightarrow w \in \hat{\mathfrak{v}}(s) \cup \hat{\mathfrak{v}}(u). \end{aligned}$$

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

$$\begin{aligned} \ell \in \hat{\mathfrak{v}}^{w}(s) \cdot \hat{\mathfrak{v}}^{w}(u) &\Longleftrightarrow (\ell \in \hat{\mathfrak{v}}^{w}(s) \wedge \mathsf{I} \in \hat{\mathfrak{v}}^{w}(u)) \vee (\mathsf{I} \in \hat{\mathfrak{v}}^{w}(s) \wedge \ell \in \hat{\mathfrak{v}}^{w}(u)) && \text{(Prop. 9)} \\ &\Longrightarrow (w \in \hat{\mathfrak{v}}(s) \wedge \mathsf{I} \in \hat{\mathfrak{v}}(u)) \vee (\mathsf{I} \in \hat{\mathfrak{v}}(s) \wedge w \in \hat{\mathfrak{v}}(u)) && \text{(IH with Lem. 4)} \\ &\Longrightarrow w \in \hat{\mathfrak{v}}(s) \cdot \hat{\mathfrak{v}}(u) \end{aligned}$$

Case $t = s^{*}$: We have:

$$\begin{aligned} \ell \in \hat{\mathfrak{v}}^{w}(s)^{*} &\Longleftrightarrow \ell \in \hat{\mathfrak{v}}^{w}(s) && \text{(Prop. 9)} \\ &\Longrightarrow w \in \hat{\mathfrak{v}}(s) && \text{(IH)} \\ &\Longrightarrow w \in \hat{\mathfrak{v}}(s)^{*} \end{aligned}$$

□

Thus we have obtained the expected condition for $\mathfrak{v}^{w}$ as follows:

Corollary 11. *Let $\mathfrak{v}$ be a language valuation and w be a word. For all variables x and terms t,*

$$w \in \hat{\mathfrak{v}}(x) \setminus \hat{\mathfrak{v}}(t) \quad \Longrightarrow \quad \ell \in \hat{\mathfrak{v}}^{w}(x) \setminus \hat{\mathfrak{v}}^{w}(t).$$

Proof. For $\ell \in \hat{\mathfrak{v}}^{w}(x)$: By the construction of $\mathfrak{v}^{w}$, $\ell \in \hat{\mathfrak{v}}^{w}(x)$ iff $w \in \hat{\mathfrak{v}}(x)$. For $\ell \notin \hat{\mathfrak{v}}^{w}(t)$: By Lem. 10. □

Theorem 12. *For all variables x and terms t, the following are equivalent:*

1. $\mathsf{LANG} \models x \leq t$*;*
2. $\hat{\mathfrak{v}}(x) \subseteq \hat{\mathfrak{v}}(t)$ *for all language valuations $\mathfrak{v}$ over the set $\{\ell\}$ s.t. $\mathfrak{v}(y) \subseteq \{\mathsf{I}, \ell\}$ for all y;*
3. $\hat{\mathfrak{v}}^{w}(x) \subseteq \hat{\mathfrak{v}}^{w}(t)$ *for all language valuations $\mathfrak{v}$ and words w.*

Proof. 1⇒2, 2⇒3: Trivial, as $\hat{\mathfrak{v}}^{w}(y) \subseteq \{\mathsf{I}, \ell\}$ for all y. 3⇒1: The contraposition is shown by Cor. 11. □

Corollary 13. *The variable inclusion problem (given a variable x and a term t, does $\mathsf{LANG} \models x \leq t$?) is decidable and coNP-complete for KA terms with variable complements.*

Proof. (in coNP): By the condition 2 of Thm. 12, we can give an algorithm as with Cor. 6. (coNP-hard): We give a reduction from the validity problem of propositional formulas in disjunctive normal form, as with Cor. 6. Given a propositional formula φ in disjunctive normal form, let t be the term obtained by the translation in Cor. 6; so, we have that φ is valid iff $\mathsf{LANG} \models \mathsf{I} \leq t$. Then we also have that $\mathsf{LANG} \models \mathsf{I} \leq t$ iff $\mathsf{LANG} \models z \leq z \cdot t$ (where z is a fresh variable); the direction $\Longrightarrow$ is shown by the congruence law, and the direction $\Longleftarrow$ is shown by the substitution law. Therefore, we have that φ is valid iff $\mathsf{LANG} \models z \leq z \cdot t$; thus, the variable inclusion problem is coNP-hard. □

4.1 Generalization from variables to composition-free terms

The proof above applies to not only variables but also terms t having the following property: For all language valuations $\mathfrak{v}$,

$$\text{for all non-empty words } w, \qquad w \in \hat{\mathfrak{v}}(t) \Longrightarrow \ell \in \hat{\mathfrak{v}}^{w}(t). \tag{L_1}$$

(This condition is intended for composition-free terms (Lem. 15). This is generalized to (L_n) in Sect. 5.1.) If t satisfies the condition (L_1), then combining with Lem. 10 (and with Lem. 4 for the empty word I) yields that for all language valuations $\mathfrak{v}$ and words w,

$$w \in \hat{\mathfrak{v}}(t) \setminus \hat{\mathfrak{v}}(s) \quad \Longrightarrow \quad \begin{cases} \ell \in \hat{\mathfrak{v}}^{w}(t) \setminus \hat{\mathfrak{v}}^{w}(s) & (\text{if } w \neq \mathsf{I}) \\ \mathsf{I} \in \hat{\mathfrak{v}}^{w}(t) \setminus \hat{\mathfrak{v}}^{w}(s) & (\text{if } w = \mathsf{I}) \end{cases}.$$

Hence, we have the following:

Theorem 14 (cf. Thm. 12)**.** *For all terms t, s, if t satisfies (L_1), then the following are equivalent:*

1. $\mathsf{LANG} \models t \leq s$*;*
2. $\hat{\mathfrak{v}}(t) \subseteq \hat{\mathfrak{v}}(s)$ *for all language valuations $\mathfrak{v}$ over the set $\{\ell\}$ s.t. $\mathfrak{v}(x) \subseteq \{\mathsf{I}, \ell\}$ for all x;*
3. $\hat{\mathfrak{v}}^{w}(t) \subseteq \hat{\mathfrak{v}}^{w}(s)$ *for all language valuations $\mathfrak{v}$ and words w.*

Proof. As with Thm. 12 (use the above, instead of Cor. 11). □

Thm. 14 can apply to composition-free terms. We say that a term t is *composition-free* if composition $(\cdot)$ nor Kleene-star $(\_^{*})$ does not occur in t.

Lemma 15. *Every composition-free terms t satisfies the condition (L_1).*

Proof. By easy induction on t. Case $t = x, x^{-}$: By the definition of $\mathfrak{v}^{w}$. Case $t = \mathsf{I}$: By that $w \notin \hat{\mathfrak{v}}(\mathsf{I})$ holds for all non-empty words w. Case $t = \bot$: By $w \notin \hat{\mathfrak{v}}(\bot)$ always. Case $t = s \cup u$: By IH, we have that $w \in \hat{\mathfrak{v}}(s) \cup \hat{\mathfrak{v}}(u) \iff w \in \hat{\mathfrak{v}}(s) \vee w \in \hat{\mathfrak{v}}(u) \Longrightarrow \ell \in \hat{\mathfrak{v}}^{w}(s) \vee \ell \in \hat{\mathfrak{v}}^{w}(u) \iff \ell \in \hat{\mathfrak{v}}^{w}(s) \cup \hat{\mathfrak{v}}^{w}(u)$. □

Corollary 16. *The following problem is coNP-complete for KA terms with variable complements:*

Given a composition-free term t and a term s, does $\mathsf{LANG} \models t \leq s$ hold?

Proof. (coNP-hard): By Cor. 6, as t is possibly I. (in coNP): By Thm. 14 with Lem. 15, we can give an algorithm (from the condition 2 of Thm. 14) as with Cor. 13. □

Remark 17. Lem. 15 fails for non-composition-free terms. For example, when $\mathfrak{v}(x) = \{a\}$, we have

$$aa \in \hat{\mathfrak{v}}(xx) \qquad\qquad \ell \notin \hat{\mathfrak{v}}^{aa}(xx).$$

(Note that $\hat{\mathfrak{v}}^{aa}(xx) = \emptyset$, as $\hat{\mathfrak{v}}^{aa}(x) = \emptyset$ by $\mathfrak{v}(x) = \{a\}$.)

4.2 The universality problem

The *universality problem* is the following problem:

Given a term t, does $\mathsf{LANG} \models \top \leq t$?

As a consequence of Cor. 16, the universality problem is also decidable and coNP-complete.

Corollary 18. *The universality problem is decidable and coNP-complete for KA terms with variable complements.*

Proof. (in coNP): We can apply Cor. 16 because the term $x \cup x^-$ is composition-free and $\mathsf{LANG} \models \top = x \cup x^-$ holds. (coNP-hard): Similar to Cor. 13. Given a propositional formula φ in disjunctive normal form, let t be the term obtained by the translation in Cor. 6; so, we have that φ is valid iff $\mathsf{LANG} \models \mathsf{I} \leq t$. Then we also have that $\mathsf{LANG} \models \mathsf{I} \leq t$ iff $\mathsf{LANG} \models \top \leq \top \cdot t$, which is proved as follows. $\Longrightarrow$: By the congruence law. $\Longleftarrow$: We prove the contraposition. Assume $\mathsf{LANG} \not\models \mathsf{I} \leq t$; then $\mathsf{I} \not\in \hat{\mathfrak{v}}(t)$ for some language valuation $\mathfrak{v}$. Then $\mathsf{I} \not\in \hat{\mathfrak{v}}(\top \cdot t)$ holds; thus, $\mathsf{LANG} \not\models \top \leq \top \cdot t$. Hence, the universality problem is coNP-complete. $\square$

Remark 19. In the standard language equivalence, because $[\mathbf{V}^*] = [\top]$ (and the constant $\top$ is usually not a primitive symbol of regular expressions), the universality problem is always of the form: $[\mathbf{V}^*] = [t]$. However, $\mathsf{LANG} \models \mathbf{V}^* \leq t$ is different from $\mathsf{LANG} \models \top \leq t$, as $\mathsf{LANG} \not\models \mathbf{V}^* = \top$.

Remark 20. Under the standard language equivalence, the universality problem—given a term t, does $[\top] \subseteq [t]$? (i.e., does $[t] = \mathbf{V}^*$?)—is PSPACE-hard [5, 8, 12]. Hence, for KA terms with variable complements, the universality problem w.r.t. languages is strictly easier (cf. Remark 7) than that under the standard language equivalence unless NP = PSPACE.

5 The word inclusion problem

Let $\mathbf{V}' = \{x, x^- \mid x \in \mathbf{V}\}$. The *word inclusion problem* is the following problem:

Given a word w over $\mathbf{V}'$ and a term t, does $\mathsf{LANG} \models w \leq t$?

As Remark 17 shows, we cannot apply the method given in Sect. 4 straightforwardly. Nevertheless, we can solve this problem by generalizing the language valuation of Def. 8, as follows. The valuations in Defs. 8, 21 are given by the first author.

Definition 21 (*words-to-letters valuations*)**.** For a language valuation $\mathfrak{v}$ over a set X and words $w_0, \ldots, w_{n-1}$ over X (where $n \geq 0$), the language valuation $\mathfrak{v}^{\langle w_0, \ldots, w_{n-1} \rangle}$ over the set $\{\ell_0, \ldots, \ell_{n-1}\}$ (where $\ell_0, \ldots, \ell_{n-1}$ are pairwise distinct letters) is defined as follows:

$$\mathfrak{v}^{\langle w_0, \ldots, w_{n-1} \rangle}(x) \quad \triangleq \quad \{\ell_i \ldots \ell_{j-1} \mid 0 \leq i \leq j \leq n \wedge w_i \ldots w_{j-1} \in \mathfrak{v}(x)\}.$$

(Note that the language valuation $\mathfrak{v}^w$ (Def. 8) is the case $n = 1$ of Def. 21 and the language valuation $\mathfrak{v}^{\langle\rangle}$ in the proof of Thm. 5 is the case $n = 0$ of Def. 21.)

By using words-to-letters valuations, we can naturally strengthen the results in Sect. 4 from variables to words. We prepare the following fact:

Proposition 22 (cf. Prop. 9)**.** *For all languages L, K and words w,*

$$\begin{aligned} w \in L \cup K &\iff w \in L \vee w \in K \\ w \in L \cdot K &\iff \exists v, v' \text{ s.t. } w = vv',\ v \in L \wedge v' \in K \\ w \in L^* &\iff \exists n \in \mathbb{N}, \exists v_0, \ldots, v_{n-1} \text{ s.t. } w = v_0 \ldots v_{n-1}, \forall i < n,\ v_i \in L. \end{aligned}$$

Proof. By definition. $\square$

Lemma 23 (cf. Lem. 10)**.** *Let $\mathfrak{v}$ be a language valuation and $w_0, \dots, w_{n-1}$ be words (where $n \geq 0$). For all terms t and $0 \leq i \leq j \leq n$, we have:*

$$\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(t) \quad \Longrightarrow \quad w_i \dots w_{j-1} \in \hat{\mathfrak{v}}(t).$$

Proof. By induction on t.

Case $t = x, x^-$: By the construction of $\mathfrak{v}^{\langle w_0,\dots,w_{n-1}\rangle}$, $\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(x)$ iff $w_i \dots w_{j-1} \in \hat{\mathfrak{v}}(x)$. (Hence, we also have $\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(x^-)$ iff $w_i \dots w_{j-1} \in \hat{\mathfrak{v}}(x^-)$.)

Case $t = \bot$, Case $t = \mathsf{I}$ where $i < j$: By $\ell_i \dots \ell_{j-1} \notin \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(t)$.

Case $t = \mathsf{I}$ where $i = j$: By $\mathsf{I} \in \hat{\mathfrak{v}}(\mathsf{I})$.

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

$$\begin{aligned}
\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(s \cup u) &\Longleftrightarrow \ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(s) \vee \ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(u) && \text{(Prop. 22)}\\
&\Longrightarrow w_i \dots w_{j-1} \in \hat{\mathfrak{v}}(s) \vee w_i \dots w_{j-1} \in \hat{\mathfrak{v}}(u) && \text{(IH)}\\
&\Longrightarrow w_i \dots w_{j-1} \in \hat{\mathfrak{v}}(s \cup u).
\end{aligned}$$

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

$$\begin{aligned}
\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(s \cdot u) &\Longleftrightarrow \bigvee_{i \leq k \leq j} (\ell_i \dots \ell_{k-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(s) \wedge \ell_k \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(u))\\
& && \text{(Prop. 22)}\\
&\Longrightarrow \bigvee_{i \leq k \leq j} (w_i \dots w_{k-1} \in \hat{\mathfrak{v}}(s) \wedge w_k \dots w_{j-1} \in \hat{\mathfrak{v}}(u)) && \text{(IH)}\\
&\Longrightarrow w_i \dots w_{j-1} \in \hat{\mathfrak{v}}(s \cdot u).
\end{aligned}$$

Case $t = s^*$: We have

$$\begin{aligned}
\ell_i \dots \ell_{j-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(s^*) &\Longleftrightarrow \exists m \in \mathbb{N}, \bigvee_{i = k_0 \leq k_1 \leq \dots \leq k_m = j} \bigwedge_{l=1}^{m} (\ell_{k_{l-1}} \dots \ell_{k_l - 1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(s)) && \text{(Prop. 22)}\\
&\Longrightarrow \exists m \in \mathbb{N}, \bigvee_{i = k_0 \leq k_1 \leq \dots \leq k_m = j} \bigwedge_{l=1}^{m} (w_{k_{l-1}} \dots w_{k_l - 1} \in \hat{\mathfrak{v}}(s)) && \text{(IH)}\\
&\Longrightarrow w_i \dots w_{j-1} \in \hat{\mathfrak{v}}(s^*).
\end{aligned}$$

$\square$

Corollary 24 (cf. Cor. 11)**.** *Let $\mathfrak{v}$ be a language valuation, w be a word, $w_0, \dots, w_{n-1}$ be words s.t. $w = w_0 \dots w_{n-1}$. For all words v over $\mathbf{V}'$ of length n and all terms t,*

$$w \in \hat{\mathfrak{v}}(v) \setminus \hat{\mathfrak{v}}(t) \quad \Longrightarrow \quad \ell_0 \dots \ell_{n-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(v) \setminus \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(t).$$

Proof. For $\ell_0 \dots \ell_{n-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(v)$: Let $v = x_0 \dots x_{n-1}$. For each $i < n$, by the construction of $\mathfrak{v}^{\langle w_0,\dots,w_{n-1}\rangle}$, $\ell_i \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(x_i)$ iff $w_i \in \hat{\mathfrak{v}}(x_i)$. Thus, we have that $\ell_0 \dots \ell_{n-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(v)$. For $\ell_0 \dots \ell_{n-1} \notin \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(t)$: By Lem. 23. $\square$

Theorem 25 (cf. Thm. 12)**.** *For all words v over $\mathbf{V}'$ of length n and all terms t, the following are equivalent:*

1. $\mathsf{LANG} \models v \leq t$;
2. $\hat{\mathfrak{v}}(v) \subseteq \hat{\mathfrak{v}}(t)$ for all language valuations $\mathfrak{v}$ s.t. $\mathfrak{v}(x) \subseteq \{\ell_i \dots \ell_j \mid 0 \leq i \leq j \leq n\}$ for all x;
3. $\hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(v) \subseteq \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(t)$ for all language valuations $\mathfrak{v}$ and words $w_0,\dots,w_{n-1}$.

Proof. 1⇒2, 2⇒3: Trivial. 3⇒1: The contraposition is shown by Cor. 24. □

Corollary 26 (cf. Cor. 13)**.** *The word inclusion problem (given a word w and a term t, does $\mathsf{LANG} \models w \leq t$?) is decidable and coNP-complete for KA terms with variable complements.*

Proof. (coNP-hard): By Cor. 6, as w is possibly I. (in coNP): By the condition 2 of Thm. 25, we can give an algorithm as with Cor. 13. □

5.1 Generalization from words to star-free terms

We can apply Thm. 25 to not only words over $\mathbf{V}'$ but also terms t having the following property:

For all language valuations $\mathfrak{v}$ and non-empty words w, for some $w_0,\dots,w_{n-1}$ s.t. $w = w_0 \dots w_{n-1}$,

$$w \in \hat{\mathfrak{v}}(t) \Longrightarrow \ell_0 \dots \ell_{n-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(t). \tag{L_n}$$

If t satisfies the condition (L_n), then combining with Lem. 23 (and with Lem. 4 for the empty word I) yields that for all language valuations $\mathfrak{v}$ and words w, for some words $w_0,\dots,w_{n-1}$ s.t. $w = w_0 \dots w_{n-1}$, we have:

$$w \in \hat{\mathfrak{v}}(t) \setminus \hat{\mathfrak{v}}(s) \quad \Longrightarrow \quad \begin{cases} \ell_0 \dots \ell_{n-1} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(t) \setminus \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(s) & (\text{if } w \neq \mathsf{I}) \\ \mathsf{I} \in \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(t) \setminus \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(s) & (\text{if } w = \mathsf{I}) \end{cases}.$$

Hence, we have the following:

Theorem 27 (cf. Thm. 14)**.** *For all terms t,s, if t satisfies (L_n), the following are equivalent:*

1. $\mathsf{LANG} \models t \leq s$;
2. $\hat{\mathfrak{v}}(t) \subseteq \hat{\mathfrak{v}}(s)$ for all language valuations $\mathfrak{v}$ over the set $\{\ell_0,\dots,\ell_{n-1}\}$ s.t. $\mathfrak{v}(x) \subseteq \{\ell_i \dots \ell_j \mid 0 \leq i \leq j \leq n\}$ for all x;
3. $\hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(t) \subseteq \hat{\mathfrak{v}}^{\langle w_0,\dots,w_{n-1}\rangle}(s)$ for all language valuations $\mathfrak{v}$ and words $w_0,\dots,w_{n-1}$.

Proof. As with Thm. 25 (use the above, instead of Cor. 24). □

By using Thm. 27, we can generalize Cor. 26 from words to star-free terms. We say that a term t is *star-free* if the Kleene-star ($\_^*$) does not occur in t.

Lemma 28 (cf. Lem. 15)**.** *Every star-free term t satisfies (L_n) for some n.*

Proof. Because the set $[t]_{\mathbf{V}'}$ is finite as t is star-free, let n be the maximal length among words in $[t]_{\mathbf{V}'}$. Let $\mathfrak{v}$ be a language valuation and let w be a non-empty word such that $w \in \hat{\mathfrak{v}}(t)$. Since $\hat{\mathfrak{v}}(t) = \hat{\mathfrak{v}}([t]_{\mathbf{V}'})$ (Lem. 2), there is a word $v \in [t]_{\mathbf{V}'}$ such that $w \in \hat{\mathfrak{v}}(v)$. Let $v = x_0 \dots x_{m-1}$ (note that $m \geq 1$, as w is non-empty and $w \in \hat{\mathfrak{v}}(v)$). Since $w \in \hat{\mathfrak{v}}(x_0 \dots x_{m-1})$, there are $w_0,\dots,w_{m-1}$ of $w = w_0 \dots w_{m-1}$ such that $w_i \in \hat{\mathfrak{v}}(x_i)$ for every i. Let $\mathfrak{v}' \triangleq \mathfrak{v}^{\langle w_0,\dots,w_{m-1},\mathsf{I},\dots,\mathsf{I}\rangle}$, where the length of the sequence is n. Then, we have $\ell_i \in \hat{\mathfrak{v}}'(x_i)$ for every $0 \leq i \leq m-2$ and $\ell_{m-1}\ell_m \dots \ell_{n-1} \in \hat{\mathfrak{v}}'(x_{m-1})$; thus, $\ell_0 \dots \ell_{n-1} \in \hat{\mathfrak{v}}'(x_0 \dots x_{m-1}) = \hat{\mathfrak{v}}'(v) \subseteq \hat{\mathfrak{v}}'(t)$. Hence, this completes the proof. □

Corollary 29. *The following problem is coNP-complete for KA terms with variable complements:*

Given a star-free term t and a term s, does $\mathsf{LANG} \models t \leq s$*?*

Proof. (coNP-hard): By Cor. 6, as t is possibly I. (in coNP): By Lem. 28, we can give an algorithm as with Cor. 26. □

5.2 words-to-letters valuation property

Finally, we show the following property; thus, we have that words-to-letters valuations are sufficient for the equational theory of (full) KA terms with variable complements.

Corollary 30 (words-to-letters valuation property)**.** *For all terms t, s, the following are equivalent:*

1. $\mathsf{LANG} \not\models t \leq s$;
2. *there is a words-to-letters valuation $\mathfrak{v}$ such that $\hat{\mathfrak{v}}(t) \not\subseteq \hat{\mathfrak{v}}(s)$.*

Proof. 2⇒1: Trivial. 1⇒2: Since $\mathsf{LANG} \not\models t \leq s$, there is a language valuation $\mathfrak{v}$ such that $\hat{\mathfrak{v}}(t) \not\subseteq \hat{\mathfrak{v}}(s)$. Since $\hat{\mathfrak{v}}(t) = \hat{\mathfrak{v}}([t]_{\mathbf{V}'})$ (Lem. 2), there is a word $v \in [t]_{\mathbf{V}'}$ such that $\hat{\mathfrak{v}}(v) \not\subseteq \hat{\mathfrak{v}}(s)$ (i.e., $\mathsf{LANG} \not\models v \leq s$). Let n be the length of v. By Thm. 25, there are a words-to-letters valuation $\mathfrak{v}$ and words $w_0, \ldots, w_{n-1}$ such that $\hat{\mathfrak{v}}^{\langle w_0,\ldots,w_{n-1}\rangle}(v) \not\subseteq \hat{\mathfrak{v}}^{\langle w_0,\ldots,w_{n-1}\rangle}(s)$. Thus $\hat{\mathfrak{v}}^{\langle w_0,\ldots,w_{n-1}\rangle}(t) \not\subseteq \hat{\mathfrak{v}}^{\langle w_0,\ldots,w_{n-1}\rangle}(s)$, as $v \in [t]_{\mathbf{V}'}$ (Lem. 2). Hence this completes the proof. □

6 On the equational theory of words with variable complements

We prove that the equational theory of words over $\mathbf{V}'$ coincides with the word equivalence (Thm. 32). We give language valuations for separating two distinct words based on words-to-letters valuations.

Lemma 31. *Let $w = x_0 \ldots x_{n-1}$ and $v = y_0 \ldots y_{m-1}$ be words over $\mathbf{V}'$, where $n \leq m$. Let $\mathfrak{v}$ be a language valuation over $\{\ell_0, \ldots, \ell_{n-1}\}$ such that*

- *for all $i < n$, $\ell_i \in \hat{\mathfrak{v}}(x_i)$;*
- *for all $i < m$ and $i \leq j \leq n$, $\ell_i \ldots \ell_{j-1} \in \hat{\mathfrak{v}}(y_i)$ iff $(y_i = x_i \land j = i+1)$.*

If $w \neq v$, then $\ell_0 \ldots \ell_{n-1} \in \hat{\mathfrak{v}}(w) \setminus \hat{\mathfrak{v}}(v)$. Such a language valuation $\mathfrak{v}$ always exists.

Proof. Since $\ell_i \in \mathfrak{v}(x_i)$ for all i, we have $\ell_0 \ldots \ell_{n-1} \in \hat{\mathfrak{v}}(w)$. Assume that $\ell_0 \ldots \ell_{n-1} \in \hat{\mathfrak{v}}(y_0 \ldots y_{m-1})$. For $i = 0$, by the condition of $\hat{\mathfrak{v}}(y_i)$ (i.e., $\ell_i \ldots \ell_{j-1} \notin \hat{\mathfrak{v}}(y_i)$ unless $j = i+1$), we should have $\ell_i \in \hat{\mathfrak{v}}(y_i)$, $\ell_{i+1} \ldots \ell_{n-1} \in \hat{\mathfrak{v}}(y_{i+1} \ldots y_{m-1})$, and $y_i = x_i$. By using the same argument iteratively, the condition above should hold for all $i < n$; thus, we have $\mathsf{I} \in \hat{\mathfrak{v}}(y_n \ldots y_{m-1})$ and $y_0 \ldots y_{n-1} = x_0 \ldots x_{m-1}$. Since $\mathsf{I} \notin \hat{\mathfrak{v}}(y_n)$, we have $y_n \ldots y_{m-1} = \mathsf{I}$; thus, $m = n$. However this yields $w = x_0 \ldots x_{n-1} = y_0 \ldots y_{m-1} = v$, which contradicts the assumption. Hence $\ell_0 \ldots \ell_{n-1} \in \hat{\mathfrak{v}}(w) \setminus \hat{\mathfrak{v}}(v)$. Additionally, such a language valuation $\mathfrak{v}$ always exists as follows. If some conditions conflict, then the first condition ($\ell_i \in \hat{\mathfrak{v}}(x_i)$) and the second condition when $j = i+1$ ($\ell_i \in \hat{\mathfrak{v}}(y_i)$) are for some i. If $y_i = x_i$, then $\ell_i \in \hat{\mathfrak{v}}(y_i) = \hat{\mathfrak{v}}(x_i)$, so it does not conflict to the condition $\ell_i \in \hat{\mathfrak{v}}(x_i)$; If $y_i = x_i^-$ (or $y_i^- = x_i$), then $\ell_i \notin \hat{\mathfrak{v}}(y_i)$, so it does not conflict to the condition $\ell_i \in \hat{\mathfrak{v}}(x_i)$. Otherwise, they are not conflicted, as the variables occurring in x_i and y_i are different. Thus, in either case, conditions are not conflicted. Hence, this completes the proof. □

Theorem 32 (Completeness for words with variable complements)**.** *For all words w, v over $\mathbf{V}'$,*

$$\mathsf{LANG} \models w = v \quad \Longleftrightarrow \quad w = v.$$

Proof. $\Longleftarrow$: Clear. $\Longrightarrow$: The contraposition is shown by Lem. 31. □

Remark 33. Since $[w]_{\mathbf{V}'} = \{w\}$, Thm. 32 also shows that: for all words w, v over $\mathbf{V}'$,

$$[w]_{\mathbf{V}'} = [v]_{\mathbf{V}'} \quad \Longleftrightarrow \quad \mathsf{LANG} \models w = v.$$

However, for general terms, the direction $\Longleftarrow$ fails: For example, when $x \neq y$,

$$\mathsf{LANG} \models x \cup x^- = y \cup y^- \qquad\qquad [x \cup x^-]_{\mathbf{V}'} \neq [y \cup y^-]_{\mathbf{V}'}.$$

(The direction $\Longrightarrow$ always holds by Lem. 2.) Thus, we need more axioms to characterize the equational theory.

Remark 34. As (‡) and Remark 1, for all words w, v over $\mathbf{V}'$,

$$\mathsf{LANG} \models w = v \quad \Longrightarrow \quad [w] = [v].$$

However, the converse direction fails even for words (e.g., $w = x^-$ and $v = x^- x^-$).

6.1 Separating one-variable words with small number of letters

We write LANG_n for the class of language models over a set of cardinality at most n. We write

$$\mathsf{LANG}_n \models t = s \quad \stackrel{\Delta}{\Longleftrightarrow} \quad \hat{\mathfrak{v}}(t) = \hat{\mathfrak{v}}(s) \text{ holds for all (language) valuations } \mathfrak{v} \text{ on } S\text{-algebras in } \mathsf{LANG}_n.$$

Notice that words-to-letters valuations need an unbounded number of letters; so the proof of Thm. 32 cannot directly apply to the class LANG_n. Nevertheless, for *one-variable* words (i.e., words over the set $\{z, z^-\}$ where z is a variable), we can show completeness theorems (cf. Thm. 32) of the equational theory over LANG_n, as Thms. 35, 36. The valuation in the proof of Thm. 36 is given by the second author.

For a word $w = x_0 \dots x_{n-1} \in \{z, z^-\}^*$ and $x \in \{z, z^-\}$, we write $\|w\|_x$ for the number $\#(\{0 \leq i < n \mid x_i = x\})$. For a letter a and $n \in \mathbb{N}$, we write a^n for the word $a \dots a$ of length n.

Theorem 35. *For all words $w, v \in \{z, z^-\}^*$, we have:*

$$\mathsf{LANG}_1 \models w = v \quad \Longleftrightarrow \quad \|w\|_z = \|v\|_z \wedge \|w\|_{z^-} = \|v\|_{z^-}.$$

Proof. $\Longleftarrow$: By the following commutative law: for all language valuations $\mathfrak{v}$ over a set of cardinality at most 1, $\hat{\mathfrak{v}}(zz^-) = \hat{\mathfrak{v}}(z^- z)$. $\Longrightarrow$: If $\|w\|_z < \|v\|_z$, then let $\mathfrak{v}$ be the language valuation defined by $\mathfrak{v}(z) = \{a\}$. Then $a^{\|w\|_z} \in \hat{\mathfrak{v}}(w) \setminus \hat{\mathfrak{v}}(v)$; thus $\mathsf{LANG}_1 \not\models w = v$. If $\|w\|_{z^-} < \|v\|_{z^-}$, then let $\mathfrak{v}$ be the language valuation defined by $\mathfrak{v}(z) = \{a\}^* \setminus \{a\}$. Then $a^{\|w\|_{z^-}} \in \hat{\mathfrak{v}}(w) \setminus \hat{\mathfrak{v}}(v)$. If $\|w\|_z > \|v\|_z$ (resp. $\|w\|_{z^-} > \|v\|_{z^-}$), then similarly to the cases above. □

Theorem 36. *For all words $w, v \in \{z, z^-\}^*$, we have:*

$$\mathsf{LANG}_2 \models w = v \quad \Longleftrightarrow \quad \mathsf{LANG} \models w = v \quad \Longleftrightarrow \quad w = v.$$

Proof. The two $\Longleftarrow$ are clear by definition. We prove $w \neq v \Longrightarrow \mathsf{LANG}_2 \not\models w = v$. By Thm. 35, it suffices to show the case when $\|w\|_z = \|v\|_z$ and $\|w\|_{z^-} = \|v\|_{z^-}$. Let $n \triangleq \|w\|_{z^-} = \|v\|_{z^-}$ and let w, v be as follows:

$$w = z^{c_0} z^- z^{c_1} \dots z^- z^{c_n}$$
$$v = z^{d_0} z^- z^{d_1} \dots z^- z^{d_n}.$$

Since $w \neq v$, there is $i \leq n$ such that $c_j = d_j$ for all $j < i$ and $c_i \neq d_i$. Without loss of generality, we can assume $c_i < d_i$. Now, we consider the following language valuation $\mathfrak{v}$ over $A \triangleq \{a,b\}$:

$$\mathfrak{v}(z) \triangleq [(aA^*) \cap (A^*a)] = \{c_0 \dots c_{n-1} \in \{a,b\}^* \mid n \geq 1, c_0 = a, c_{n-1} = a\}.$$

Then $a^{(\sum_{j=0}^{i} c_j)} b a^{(\sum_{j=i+1}^{n} c_j)} \in \hat{\mathfrak{v}}(w)$, as $a \in \hat{\mathfrak{v}}(z)$ and $\mathsf{I}, b \in \hat{\mathfrak{v}}(z^-)$. Assume, towards contradiction, that $a^{(\sum_{j=0}^{i} c_j)} b a^{(\sum_{j=i+1}^{n} c_j)} \in \hat{\mathfrak{v}}(v)$. Each z occurring in v should map to a, as $(\sum_{j=0}^{n} c_j) = (\sum_{j=0}^{n} d_j)$ and every word in $\mathfrak{v}(z)$ except for a has at least two occurrences of a. The $(\sum_{j=0}^{i} c_j)$-th occurrence and $((\sum_{j=0}^{i} c_j) + 1)$-th occurrence of z^- are adjacent (since $(\sum_{j=0}^{i} c_j) < (\sum_{j=0}^{i} d_j)$). Combining them yields $b \in \hat{\mathfrak{v}}(\mathsf{I})$, thus reaching a contradiction. Hence, $a^{(\sum_{j=0}^{i} c_j)} b a^{(\sum_{j=i+1}^{n} c_j)} \in \hat{\mathfrak{v}}(w) \setminus \hat{\mathfrak{v}}(v)$. This completes the proof. □

The proof of Thms. 35, 36 only applies to one-variable words. We leave open Thms. 35, 36 for many variables words (cf. Thm. 32).

7 Conclusion and future work

We have introduced words-to-letters valuations. By using them, we have shown the decidability and complexity of the identity/variable/word inclusion problems (Cors. 6, 13, 26) and the universality problem (Cor. 18) of the equational theory of KA terms with variable complements w.r.t. languages; in particular, the inequational theory $t \leq s$ is coNP-complete when t does not contain Kleene-star (Cor. 29). Additionally, we have proved a completeness theorem for words with variable complements w.r.t. languages (Thm. 32); moreover, for one-variable words, the equational theory over LANG coincides with that over LANG_2 (Thm. 36).

A natural interest is to extend our decidability results, e.g., for full KA terms with variable complements. As Cor. 30 shows, even for full terms, words-to-letters valuations are sufficient valuations in investigating the equational theory. The first author conjectures that the equational theory of KA terms with variable complements is decidable, possibly by combining the technique like *saturable paths* [9] (which were introduced for the equational theory w.r.t. binary relations). Additionally, we leave open the (finite) axiomatizability of the equational theory (including that over sets of bounded cardinality; cf. Sect. 6.1).

Acknowledgement

We would like to thank the anonymous reviewers for their useful comments. This work was supported by JSPS KAKENHI Grant Number JP21K13828 and JST ACT-X Grant Number JPMJAX210B, Japan.

References

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

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

[3] John H. Conway (1971): *Regular Algebra and Finite Machines*. Chapman and Hall.

[4] Stephen A. Cook (1971): *The complexity of theorem-proving procedures*. In: STOC, ACM, p. 151–158, doi:10.1145/800157.805047.

[5] Harry B. Hunt III, Daniel J. Rosenkrantz & Thomas G. Szymanski (1976): *On the equivalence, containment, and covering problems for the regular and context-free languages*. Journal of Computer and System Sciences 12(2), pp. 222–268, doi:10.1016/S0022-0000(76)80038-4.

[6] S. C. Kleene (1951): *Representation of Events in Nerve Nets and Finite Automata*. Technical Report, RAND Corporation. Available at `https://www.rand.org/pubs/research_memoranda/RM704.html`.

[7] Dexter Kozen & Frederick Smith (1996): *Kleene algebra with tests: Completeness and decidability*. In: *CSL*, *LNCS* 1258, Springer, pp. 244–259, doi:10.1007/3-540-63172-0_43.

[8] A. R. Meyer & L. J. Stockmeyer (1972): *The equivalence problem for regular expressions with squaring requires exponential space*. In: *SWAT*, IEEE, pp. 125–129, doi:10.1109/SWAT.1972.29.

[9] Yoshiki Nakamura (2023): *Existential Calculi of Relations with Transitive Closure: Complexity and Edge Saturations*. In: *LICS*, IEEE, pp. 1–13, doi:10.1109/LICS56636.2023.10175811.

[10] Kan Ching Ng (1984): *Relation algebras with transitive closure*. Ph.D. thesis, University of California.

[11] Damien Pous & Jana Wagemaker (2022): *Completeness Theorems for Kleene Algebra with Top*. In: *CONCUR*, 243, Schloss Dagstuhl, pp. 26:1–26:18, doi:10.4230/LIPICS.CONCUR.2022.26.

[12] Larry J. Stockmeyer & Albert R. Meyer (1973): *Word problems requiring exponential time (Preliminary Report)*. In: *STOC*, ACM, pp. 1–9, doi:10.1145/800125.804029.

[13] Alfred Tarski (1941): *On the Calculus of Relations*. The Journal of Symbolic Logic 6(3), pp. 73–89, doi:10.2307/2268577.

[14] Ken Thompson (1968): *Programming Techniques: Regular expression search algorithm*. Communications of the ACM 11(6), pp. 419–422, doi:10.1145/363347.363387.

A A direct proof of the coincidence between the equational theory w.r.t. languages and the language equivalence for KA terms

(In this section, we use the notations of Sect. 2.)

We say that a term t is a *KA term* if the complement $(\_^-)$ does not occur in t. Recall language valuations for languages in Sect. 2.3.

Lemma 37 (cf. Lem. 2)**.** *Let $\mathfrak{v}$ be a language valuation. For all KA terms t, we have*

$$\hat{\mathfrak{v}}(t) \quad = \quad \hat{\mathfrak{v}}([t]).$$

Proof. By Lem. 2, as $[t] = [t]_{\mathbf{V}'}$ (since KA terms do not contain the complement). □

Theorem 38. *For all KA terms t, s,*

$$\mathsf{LANG} \models t = s \quad \Longleftrightarrow \quad [t] = [s].$$

Proof. We have

$$\begin{aligned}
\mathsf{LANG} \models t = s \quad &\Longrightarrow \quad [t] = [s] && ([_] \text{ is an instance of language valuations})\\
&\Longrightarrow \quad \text{for all language valuations } \mathfrak{v}, \hat{\mathfrak{v}}([t]) = \hat{\mathfrak{v}}([s])\\
&\Longleftrightarrow \quad \text{for all language valuations } \mathfrak{v}, \hat{\mathfrak{v}}(t) = \hat{\mathfrak{v}}(s) && (\text{Lem. 37})\\
&\Longleftrightarrow \quad \mathsf{LANG} \models t = s. && (\text{By definition})
\end{aligned}$$

□