A Complete Axiomatisation of a Fragment of Language Algebra

Paul Brunet
University College London, United Kingdom
http://paul.brunet-zamansky.fr
paul@brunet-zamansky.fr

Abstract

We consider algebras of languages over the signature of reversible Kleene lattices, that is the regular operations (empty and unit languages, union, concatenation and Kleene star) together with intersection and mirror image. We provide a complete set of axioms for the equational theory of these algebras. This proof was developed in the proof assistant Coq.

2012 ACM Subject Classification Theory of computation → Algebraic language theory

Keywords and phrases Kleene algebra, language algebra, completeness theorem, axiomatisation

Digital Object Identifier 10.4230/LIPIcs.CSL.2020.11

Supplement Material Coq formalisation: https://github.com/monstrencage/LangAlg

Funding This work was funded by the EPSRC grant IRIS (reference EP/R006865/1).

Acknowledgements I want to thank the anonymous referees who provided valuable comments, and Amina Doumane for her kind assistance.

1 Introduction

We are interested in algebras of languages, equipped with the constants empty language (0), unit language (1, the language containing only the empty word), the binary operations of union (+), intersection (∩), and concatenation (·), and the unary operations of Kleene star ($(-)^\star$) and mirror image ($\overline{(-)}$). It is convenient in this paper to see the Kleene star as a derived operator $e^\star := 1 + e^+$ with the operator e^+ representing the non-zero iteration. We call these algebras *reversible Kleene lattices*. Given a finite set of variables X, and two terms e, f built from variables and the above operations, we say that the equation $e \simeq f$ is *valid* if the corresponding equality holds universally.

In a previous paper [3] we have presented an algorithm to test the validity of such equations, and shown this problem to be ExpSpace-complete. However, we had left open the question of the axiomatisation of these algebras. We address it now, by providing in the current paper a set of axioms from which every valid equation can be derived.

Several fragments of this algebra have been studied:

Kleene algebra (KA): if we restrict ourselves to the operators of regular expressions (0, 1, +, ·, and $(-)^+$), then several axiomatisation have been proposed by Conway[4], before being shown to be complete by Krob [8] and Kozen [6].

Kleene algebra with converse: if we add to KA the mirror operation, then the previous theorem can be extended by switching to a duplicated alphabet, with a letter a' denoting the mirror of the letter a. A small number of identities may be added to KA to get a complete axiomatisation [2].

Identity-free Kleene lattices: this algebra stems from the operators 0, +, ·, ∩ and $(-)^+$. In a recent paper [5] Doumane and Pous provided a complete axiomatisation of this algebra.

28th EACSL Annual Conference on Computer Science Logic (CSL 2020).
Editors: Maribel Fernández and Anca Muscholl; Article No. 11; pp. 11:1–11:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The present work is then an extension of identity-free Kleene lattices, by adding unit and mirror image. We provide in Table 1 a set of axioms which we prove to be complete for the equational theory of language algebra, by reducing to the the completeness theorem of [5]. This proof has been formalised in Coq.

The paper is organised as follows. In Section 2, we introduce some notations and define the various types of expressions used in the paper. We present our axioms and state our main theorem. In Section 3 we deal with a technical lemma having to do with the treatment of the empty word. We proceed in Section 4 to extend the theorem of [5] with the mirror image operator. Section 5 studies in detail terms of the algebra that are below the constant 1, as those play a crucial role in the main proof. We present the proof of our main result in Section 6. We conclude in Section 7 by a discussion on an operator that is missing from our signature, namely constant $\top$ denoting the full language.

On the Coq formalisation

As we have mentioned already, the proofs in this paper have been formalised and checked using the proof assistant Coq. This has several consequences for the present article.

Since Coq offers a very high level of confidence in the proofs it validates, the summary we give here is not meant to convince the reader of our result's validity. Instead we focus on the precise statement of the theorems we proved, and the strategy we employed to establish those. If the reader has doubts as to the validity of some of our claims we refer them to the Coq proof, available on GitHub.

The source of most mistakes when dealing with formal proofs is the correspondence between the statement we want to prove and the one we actually prove. In other words, the main task when assessing the validity of a Coq proof consists in checking that the definitions and assertions in the Coq file match those we have in mind. To that effect, we tried in the present document to remain as close as possible to the Coq script. This might sometimes lead to slightly pedantic definitions, and less than intuitive proofs. We feel however that this is better than the alternative: we use the claims we checked instead of making more intuitive but imprecise arguments.

2 Preliminaries

2.1 Sets, words, and languages

Given a set X, we write $\mathcal{P}(X)$ for the powerset of X and $\mathcal{P}_f(X)$ for the set of finite subsets of X. We will denote the two-elements boolean set as 2. For two sets X, Y, we write $X \times Y$ for their Cartesian product, $X \cup Y$ for their union, and $X \cap Y$ for their intersection. The empty set is denoted by $\emptyset$. We will use the notation $f(A)$ for a set $A \subseteq X$ and a function $f : X \to Y$ to represent the set $\{y \in Y \mid \exists a \in A : f(a) = y\} = \{f(a) \mid a \in A\}$.

Let Σ be an arbitrary alphabet (set), the words over Σ are finite sequences of elements from Σ. The set of all words is written $\Sigma^\star$, and the empty word is written ε. The concatenation of two words u, v is simply denoted by uv. The mirror image of a word u, obtained by reading it backwards, is written $\overline{u}$. For instance $\overline{abc}$ is the word cba.

A language is a set of words, that is an element of $\mathcal{L}\langle\Sigma\rangle := \mathcal{P}(\Sigma^\star)$. We will also use the symbol ε to denote the unit language $\{\varepsilon\}$. The concatenation of two languages L and M, denoted by $L \cdot M$, is obtained by lifting pairwise the concatenation of words: it contains exactly those words that can be obtained as a concatenation uv where $\langle u, v\rangle \in L \times M$. Similarly the mirror image of a language L, denoted by $\overline{L}$, is the set of mirror images of words

Table 1 Axioms of reversible Kleene lattices.

$e+f=f+e$	(1a.1)	$e\cdot(f\cdot g)=(e\cdot f)\cdot g$	(1b.1)
$e+(f+g)=(e+f)+g$	(1a.2)	$e\cdot 0=0=0\cdot e$	(1b.2)
$e+0=e$	(1a.3)	$(e+f)\cdot g=e\cdot g+f\cdot g$	(1b.3)
$e\cap f=f\cap e$	(1a.4)	$e\cdot(f+g)=e\cdot f+e\cdot g$	(1b.4)
$e\cap e=e$	(1a.5)	$e^+=e+e\cdot e^+$	(1b.5)
$e\cap(f\cap g)=(e\cap f)\cap g$	(1a.6)	$e^+=e+e^+\cdot e$	(1b.6)
$(e+f)\cap g=e\cap g+f\cap g$	(1a.7)	$e\cdot f+f=f\Rightarrow e^+\cdot f+f=f$	(1b.7)
$(e\cap f)+e=e$	(1a.8)	$f\cdot e+f=f\Rightarrow f\cdot e^++f=f$	(1b.8)

(a) Distributive lattice. **(b)** Concatenation and iteration.

		$1\cdot e=e=e\cdot 1$	(1d.1)
$\overline{\overline{e}}=e$	(1c.1)	$1\cap(e\cdot f)=1\cap(e\cap f)$	(1d.2)
$\overline{e+f}=\overline{e}+\overline{f}$	(1c.2)	$1\cap\overline{e}=1\cap e$	(1d.3)
$\overline{e\cdot f}=\overline{f}\cdot\overline{e}$	(1c.3)	$(1\cap e)\cdot f=f\cdot(1\cap e)$	(1d.4)
$\overline{e\cap f}=\overline{e}\cap\overline{f}$	(1c.4)	$((1\cap e)\cdot f)\cap g=(1\cap e)\cdot(f\cap g)$	(1d.5)
$\overline{e^+}=\overline{e}^+$	(1c.5)	$(g+(1\cap e)\cdot f)^+=g^++(1\cap e)\cdot(g+f)^+$	(1d.6)

(c) Mirror image. **(d)** Unit.

from L. We write L^n when $L\in\mathcal{L}\langle\Sigma\rangle$ and $n\in\mathbb{N}$ for the iterated concatenation, defined by induction on n by $L^0:=\varepsilon$ and $L^{n+1}:=L\cdot L^n$. The language L^+ is the union of all non-zero iterations of L, i.e. $L^+:=\bigcup_{n>0}L^n$.

2.2 Terms: syntax and semantics

Throughout this paper, we will consider expressions over various signatures which we list here. We fix a set of variables X, and let $x,y,\ldots$ range over X.
Expressions: $e,f\in\mathbb{E}_X ::= x \mid 0 \mid 1 \mid e+f \mid e\cdot f \mid e\cap f \mid e^+ \mid \overline{e}$;
One-free expressions: $e,f\in\mathbb{E}'_X ::= x \mid 0 \mid e+f \mid e\cdot f \mid e\cap f \mid e^+ \mid \overline{e}$;
Simple expressions: $e,f\in\mathbb{E}^-_X ::= x \mid 0 \mid e+f \mid e\cdot f \mid e\cap f \mid e^+$;

We will use various sets of axioms, depending on the signature. All of the axioms under consideration are listed in Table 1. We use these axioms to generate equivalence relations over terms. For a type of expressions $\mathbb{T}_X\in\{\mathbb{E}_X,\mathbb{E}'_X,\mathbb{E}^-_X\}$, the *axiomatic equivalence relation*, written $\equiv$ is the smallest congruence on $\mathbb{T}_X$ containing those axioms in Table 1 that only use symbols from the signature of $\mathbb{T}_X$. This means that for $\mathbb{E}^-_X$ we use the axioms from Tables 1a and 1b, for $\mathbb{E}'_X$ we add those from Table 1c and for $\mathbb{E}_X$ we keep all of the axioms of Table 1. We will use the shorthand $e\leqq f$ to mean $e+f\equiv f$. This ensures that $\leqq$ is a partial order with respect to $\equiv$. We list in Table 2 some statements that are provable from the axioms.

▶ Remark 1. Axioms in Tables 1a and 1b are borrowed from [5]. We actually omit two axioms from Pous & Doumane: their axiomatisation include (2a.1) and (2a.3), which happen

Table 2 Some consequences of the axioms.

$e + e \equiv e$	(2a.1)	$e^+ \cdot e^+ \leqq e^+$	(2b.1)
$e \cap 0 \equiv 0$	(2a.2)	$(e^+)^+ \equiv e^+$	(2b.2)
$e \cap (e + f) \equiv e$	(2a.3)	$(1 + e)^+ \equiv 1 + e^+$	(2b.3)

(a) Lattice laws. **(b)** Iteration.

$\overline{0} \equiv 0$	(2c.1)	$e \leqq g \Rightarrow f \leqq g \Rightarrow e + f \leqq g$	(2d.1)
$\overline{1} \equiv 1$	(2c.2)	$g \leqq e \Rightarrow g \leqq f \Rightarrow g \leqq e \cap f$	(2d.2)
$0^+ \equiv 0$	(2c.3)	$e \leqq f \Leftrightarrow e \cap f \equiv e$	(2d.3)
$1^+ \equiv 1$	(2c.4)	$1 \leqq e \cdot f \Leftrightarrow 1 \leqq e \ \wedge\ 1 \leqq f$	(2d.4)

(c) Constants. **(d)** Reasoning rules.

to be derivable from the other identities. The mirror image identities, presented in Table 1c come from [2] (except (1c.4), which is a trivial extension). In Table 1d, we find (1d.1) which is a standard monoid law, as well as (1d.4) and (1d.5) from [1]. Axiom (1d.6) was also present in that paper, although using the Kleene star instead of the non-zero iteration. As far as we know the identities (1d.2) and (1d.3) are new.

Given an expression $e \in \mathbb{T}_X$, a set Σ, and a map $\sigma : X \to \mathcal{L}\langle\Sigma\rangle$, we may interpret e as a language over Σ using the following inductive definition:

$$
\begin{aligned}
&[\![x]\!]_\sigma := \sigma(x) && [\![e+f]\!]_\sigma := [\![e]\!]_\sigma \cup [\![f]\!]_\sigma && [\![e^+]\!]_\sigma := [\![e]\!]_\sigma^+ \\
&[\![0]\!]_\sigma := \emptyset && [\![e \cdot f]\!]_\sigma := [\![e]\!]_\sigma \cdot [\![f]\!]_\sigma && [\![\overline{e}]\!]_\sigma := \overline{[\![e]\!]_\sigma} \\
&[\![1]\!]_\sigma := \varepsilon && [\![e \cap f]\!]_\sigma := [\![e]\!]_\sigma \cap [\![f]\!]_\sigma &&
\end{aligned}
$$

The *semantic equivalence* and *semantic containment* relations on $\mathbb{T}_X$, respectively written $\simeq$ and $\lesssim$, are defined as follows:

$$
\begin{aligned}
e \simeq f &\Leftrightarrow \forall \Sigma, \forall \sigma : X \to \mathcal{L}\langle\Sigma\rangle, [\![e]\!]_\sigma = [\![f]\!]_\sigma . \\
e \lesssim f &\Leftrightarrow \forall \Sigma, \forall \sigma : X \to \mathcal{L}\langle\Sigma\rangle, [\![e]\!]_\sigma \subseteq [\![f]\!]_\sigma .
\end{aligned}
$$

The main result of this paper is a completeness theorem for reversible Kleene lattices:

▶ **Theorem 24** (Main result). $\forall e, f \in \mathbb{E}_X,\ e \equiv f \Leftrightarrow e \simeq f.$

Since all of the axioms in Table 1 are sound for languages, we know that the implication from left to right holds. This paper will thus focus on the converse implication, and will proceed in several steps. Our starting point will be the recently published completeness theorem for identity-free Kleene lattices [5]:

▶ **Theorem 2.** $\forall e, f \in \mathbb{E}_X^-,\ e \equiv f \Leftrightarrow e \simeq f.$

▶ Remark 3. In [5], this theorem is established for interpretations of terms as binary relations instead of languages. However both semantic equivalences coincide for this signature [1].

Table 3 The two definitions of $\sqsubseteq$.

$$\varepsilon \sqsubseteq_1 \varepsilon \qquad \frac{\varepsilon \sqsubseteq_1 v}{\varepsilon \sqsubseteq_1 \bullet v} \qquad \frac{u \sqsubseteq_1 v}{\bullet u \sqsubseteq_1 \bullet v} \qquad \frac{xu \sqsubseteq_1 v \quad x \in \Sigma}{xu \sqsubseteq_1 \bullet v} \qquad \frac{u \sqsubseteq_1 v \quad x \in \Sigma}{xu \sqsubseteq_1 xv}$$

$$u \sqsubseteq_2 u \qquad \varepsilon \sqsubseteq_2 \bullet \qquad \frac{u \sqsubseteq_2 v \quad v \sqsubseteq_2 w}{u \sqsubseteq_2 w} \qquad \frac{u \sqsubseteq_2 v \quad u' \sqsubseteq_2 v'}{uu' \sqsubseteq_2 vv'}$$

3 A remark about the empty word

In several places in the proof, it makes some difference whether or not the empty word belongs to the language of some one-free expression. We show here one way one might manipulate this property, that will be of use later on. The main technical result of this section is the following lemma:

▶ **Proposition 4.** *Given an alphabet Σ, a symbol $\bullet \notin \Sigma$, a map $\sigma : X \to \mathcal{L}\langle\Sigma\rangle$ and a set of variables $\mathcal{X} \subseteq X$, there are maps $\sigma' : X \to \mathcal{L}\langle\Sigma \cup \{\bullet\}\rangle$ and $\phi : (\Sigma \cup \{\bullet\})^\star \to \Sigma^\star$ such that:*

$$\forall a \in \mathcal{X},\ \varepsilon \notin \sigma'(a) \qquad\qquad \forall e \in \mathbb{E}'_X,\ [\![e]\!]_\sigma = \phi\left([\![e]\!]_{\sigma'} \setminus \varepsilon\right).$$

Before we can prove it, we need to introduce a few definitions and intermediate lemmas. Let us fix for the remainder of the section an alphabet Σ, and a new symbol $\bullet \notin \Sigma$. We write $\Sigma' := \Sigma \cup \{\bullet\}$. The monoid homomorphism $\phi : \Sigma'^\star \to \Sigma^\star$ is generated by

$$\phi(\bullet) := \varepsilon \qquad\qquad \forall x \in \Sigma,\ \phi(x) := x.$$

We will need an ordering $\sqsubseteq$ between words over Σ', that corresponds intuitively to "$u \sqsubseteq v$ if u can be obtained by removing some $\bullet$s from v". To define this relation, we provide two deduction systems in Table 3. The definition $\sqsubseteq_1$ can be thought of as being more algorithmic: it is syntax directed (given a pair of words u, v, there is at most one rule with conclusion $u \sqsubseteq_1 v$), and progressing from bottom to top it removes the superfluous $\bullet$s from the right hand side. The other definition is more algebraic. It can be summarised as "the smallest precongruence containing $\varepsilon \sqsubseteq_2 \bullet$". It turns out both definitions are equivalent, and we will simply write $\sqsubseteq$ instead of $\sqsubseteq_i$.

▶ **Lemma 5.** $\sqsubseteq_1 = \sqsubseteq_2$.

Proof. First we prove that $\sqsubseteq_2 \subseteq \sqsubseteq_1$. By proceeding by induction on the derivation $u \sqsubseteq_2 v$, we see that it amounts to showing that $\sqsubseteq_1$ 1) is a preorder (i.e. reflexive and transitive) 2) contains $\varepsilon \sqsubseteq_1 \bullet$ and 3) satisfies the rule $u \sqsubseteq_1 v$ and $u' \sqsubseteq_1 v'$ implies $uu' \sqsubseteq_1 vv'$.

1. Reflexivity and transitivity can be shown by a simple induction on words.
2. By induction on u we can show that $u \sqsubseteq_1 \bullet u$ (which implies $\varepsilon \sqsubseteq_1 \bullet$).
3. Then we may prove:
   - by induction on u that for any v_1, v_2 we have $v_1 \sqsubseteq_1 v_2 \Rightarrow uv_1 \sqsubseteq_1 uv_2$ and
   - by induction on the derivation $u \sqsubseteq_1 v$ that for every w, we get $uw \sqsubseteq_1 vw$.

   These two properties, together with transitivity give us that $\sqsubseteq_1$ is a precongruence.

For the other containment, we show that $u \sqsubseteq_1 v \Rightarrow u \sqsubseteq_2 v$ by a straightforward induction on the derivation $u \sqsubseteq_1 v$. ◀

By induction on the derivation $u \sqsubseteq_2 v$ we may prove the following properties:

$$u \sqsubseteq v \Rightarrow \phi(u) = \phi(v) \tag{3.1}$$

$$u \sqsubseteq v \Rightarrow \overline{u} \sqsubseteq \overline{v} \tag{3.2}$$

By induction on v, and using the definition $\sqsubseteq_1$, we get the following decomposition property:

$$u_1 u_2 \sqsubseteq v \Rightarrow \exists v_1, v_2 : v = v_1 v_2 \wedge u_1 \sqsubseteq v_1 \wedge u_2 \sqsubseteq v_2 \tag{3.3}$$

We make the following observations about words greater than ε:

▶ **Lemma 6.** *For any words $u, v \in \Sigma'$:*
1. *$\varepsilon \sqsubseteq u \Leftrightarrow \phi(u) = \varepsilon$.*
2. *If $\varepsilon \sqsubseteq u, v$ then either $u \sqsubseteq v$ or $v \sqsubseteq u$.*

Proof. 1. By (3.1), we only need to check the right to left implication. We do so by induction on u. $\phi(u) = \varepsilon$ means that u is only composed of $\bullet$s, so there are two case, both being straightforward instances of $\sqsubseteq_1$.
2. This second observation is a consequence of the following statement: if $\varepsilon \sqsubseteq u, v$ then $u \sqsubseteq v$ if and only if the length of u is smaller than the length of v. By a simple induction on $\sqsubseteq_2$ one can show that for any u, v we have the left-to-right implication. For the converse implication we perform the induction on v. ◀

We now arrive at the key property of this ordering:

▶ **Lemma 7.** *Any words u, v such that $\phi(u) = \phi(v)$ have a least upper bound, i.e. a word $u \sqcup v$ such that $u \sqsubseteq u \sqcup v$, $v \sqsubseteq u \sqcup v$ and for any word t such that $u \sqsubseteq t$ and $v \sqsubseteq t$, we have $u \sqcup v \sqsubseteq t$.*

Proof. We pose a word $w = \phi(u) = \phi(v)$, and proceed by induction on w. If $w = \varepsilon$, then by the remark we made earlier u and v are ordered, so the least upper bound is the maximum of the two.

Otherwise, we have $\phi(u) = \phi(v) = aw$. By (yet another) induction, we show that this means we can decompose u and v as follows:

$$u = u_1 a u_2 \qquad v = v_1 a v_2 \qquad \varepsilon \sqsubseteq u_1, v_1 \qquad w = \phi(u_2) = \phi(v_2).$$

So we may use our induction hypothesis to get a least upper bound for u_2 and v_2. Since u_1 and v_1 are both greater than ε, they are ordered. Without loss of generality, let us assume $u_1 \sqsubseteq v_1$. In this case, we claim that $u \sqcup v = v_1 a (u_2 \sqcup v_2)$. It is straightforward to check that $u \sqsubseteq u \sqcup v$ and $v \sqsubseteq u \sqcup v$.

For the remaining property, let t be a word such that $u \sqsubseteq t$ and $v \sqsubseteq t$. We use another decomposition lemma (omitted here), to decompose t as

$$t = t_1 a t_2 \qquad u_1 \sqsubseteq t_1 \qquad u_2 \sqsubseteq t_2 \qquad v_1 \sqsubseteq t_1 \qquad v_2 \sqsubseteq t_2.$$

This allows us to conclude: since $u_2 \sqsubseteq t_2$ and $v_2 \sqsubseteq t_2$, then $u_2 \sqcup v_2 \sqsubseteq t_2$, so:

$$u \sqcup v = v_1 a (u_2 \sqcup v_2) \sqsubseteq t_1 a t_2 = t.$$ ◀

Notice that by (3.1) and Lemma (7) we get that each equivalence class of the relation $\{\langle u, v\rangle \mid \phi(u) = \phi(v)\}$ forms a join-semilattice.

We may now prove Proposition 4:

Proof of Proposition 4. We fix Σ, σ, and $\mathcal{X}$ as in the statement, and define Σ' and $\phi()$ as in the rest of this section. Finally, σ' is defined as $\sigma'(x) := \{u \mid \phi(u) \in \sigma(x) \wedge (x \in \mathcal{X} \Rightarrow u \neq \varepsilon)\}$.

It is straightforward to check that $\phi(\sigma'(x)) = \sigma(x)$ for any variable x. Therefore we only need to check that this property is preserved by the operators of one-free expressions. For any languages L, M, the following distributivity laws hold:

$$\phi\left(\overline{L}\right) = \overline{\phi(L)} \qquad \phi(L \cdot M) = \phi(L) \cdot \phi(M)$$
$$\phi\left(L^+\right) = \phi(L)^+ \qquad \phi(L \cup M) = \phi(L) \cup \phi(M)$$

However, it is not the case in general that $\phi(L \cap M) = \phi(L) \cap \phi(M)$. To make the induction go through, we will need to show that this identity holds for all the languages generated from the languages $\sigma'(x)$ by the operations $0, \cdot, +, \cap, (-)^+, \overline{(-)}$. This is achieved by identifying some sufficient condition for $\phi(L \cap M) = \phi(L) \cap \phi(M)$, and showing that this condition is satisfied by every language of the shape $[\![e]\!]_{\sigma'}$.

A good choice for such a condition is the property "being upwards-closed with respect to $\sqsubseteq$", i.e. languages L such that whenever $u \in L$ and $u \sqsubseteq v$, then $v \in L$. Clearly $\sigma'(x)$ is closed for any variable x. Since the property "being closed" is preserved by each operation in the signature of $\mathbb{E}'_X$, we deduce that for any expression $e \in \mathbb{E}'_X$ the language $[\![e]\!]_{\sigma'}$ is closed.

Thankfully, for closed languages the missing identity $\phi(L \cap M) = \phi(L) \cap \phi(M)$ holds, thanks to Lemma 7. Thus we may conclude by induction on the expressions that $[\![e]\!]_\sigma = \phi([\![e]\!]_{\sigma'})$. For the last step, notice that $\varepsilon \sqsubseteq \bullet$ and $\phi(\varepsilon) = \phi(\bullet)$. Since $[\![e]\!]_{\sigma'}$ is closed, if $\varepsilon \in [\![e]\!]_{\sigma'}$ then $\bullet \in [\![e]\!]_{\sigma'}$, thus $\phi([\![e]\!]_{\sigma'} \setminus \varepsilon) = \phi([\![e]\!]_{\sigma'}) = [\![e]\!]_\sigma$. ◀

By setting the set $\mathcal{X}$ in the previous proposition to the full set X, we get the straightforward corollary, which will prove useful in the next section.

▶ **Corollary 8.** *Let e be a one-free expression, then for any expression $f \in \mathbb{E}_X$ we have*

$$e \lesssim f \Leftrightarrow \forall \Sigma, \forall \sigma : X \to \mathcal{L}\langle\Sigma\rangle, \varepsilon \notin \bigcup_{x \in X} \sigma(x) \Rightarrow [\![e]\!]_\sigma \subseteq [\![f]\!]_\sigma .$$

4 Mirror image

In this section, we show a completeness theorem for one-free expressions. In order to get this result we will use translations between $\mathbb{E}'_X$ and $\mathbb{E}^-_{X \times 2}$. An expression $e \in \mathbb{E}'_X$ is *clean*, written $e \in \mathbb{C}_X$, if the mirror operator is only applied to variables. First, notice that we may restrict ourselves to clean expressions thanks to the following inductive function:

$$\begin{aligned}
&\Upsilon : \mathbb{E}'_X \times 2 \to \mathbb{E}'_X \\
&\langle 0, b\rangle \mapsto 0 && \langle e^+, b\rangle \mapsto \Upsilon\langle e, b\rangle^+ \\
&\langle x, \top\rangle \mapsto x && \langle \overline{e}, \top\rangle \mapsto \Upsilon\langle e, \bot\rangle \\
&\langle x, \bot\rangle \mapsto \overline{x} && \langle \overline{e}, \bot\rangle \mapsto \Upsilon\langle e, \top\rangle \\
&\langle e + f, b\rangle \mapsto \Upsilon\langle e, b\rangle + \Upsilon\langle f, b\rangle && \langle e \cdot f, \top\rangle \mapsto \Upsilon\langle e, \top\rangle \cdot \Upsilon\langle f, \top\rangle \\
&\langle e \cap f, b\rangle \mapsto \Upsilon\langle e, b\rangle \cap \Upsilon\langle f, b\rangle && \langle e \cdot f, \bot\rangle \mapsto \Upsilon\langle f, \bot\rangle \cdot \Upsilon\langle e, \bot\rangle .
\end{aligned}$$

We can show by induction on terms the following properties of Υ:

$$\forall \langle e, b\rangle \in \mathbb{E}'_X \times 2, \Upsilon\langle e, b\rangle \in \mathbb{C}_X. \tag{4.1}$$
$$\forall e \in \mathbb{E}'_X, \Upsilon\langle e, \top\rangle \equiv e \text{ and } \Upsilon\langle e, \bot\rangle \equiv \overline{e}. \tag{4.2}$$

We now define translations between clean expressions and simple expressions:

- $\uparrow(-) : \mathbb{C}_X \to \mathbb{E}^-_{X\times 2}$ replaces mirrored variables $\overline{x}$ with $\langle x, \bot\rangle$ and variables x with $\langle x, \top\rangle$;
- $\downarrow(-) : \mathbb{E}^-_{X\times 2} \to \mathbb{C}_X$ replaces $\langle x, \top\rangle$ with x and $\langle x, \bot\rangle$ with $\overline{x}$.

We can easily show by induction the following properties:

$$\forall e \in \mathbb{C}_X, \downarrow\uparrow e = e. \tag{4.3}$$
$$\forall e, f \in \mathbb{E}^-_{X\times 2}, e \equiv f \Rightarrow \downarrow e \equiv \downarrow f. \tag{4.4}$$

The last step to obtain the completeness theorem for $\mathbb{E}'_X$ is the following claim:

▷ Claim 9. $\forall e, f \in \mathbb{C}_X,\ e \simeq f \Rightarrow \uparrow e \simeq \uparrow f$.

▶ **Lemma 10.** *If Claim 9 holds, then* $\forall e, f \in \mathbb{E}'_X,\ e \equiv f \Leftrightarrow e \simeq f$.

Proof. By soundness, we know that $e \equiv f \Rightarrow e \simeq f$. For the converse implication:

$$\begin{aligned}
e \simeq f &\Rightarrow \Upsilon\langle e, \top\rangle \simeq \Upsilon\langle f, \top\rangle && \text{By soundness and Equation (4.2).}\\
&\Rightarrow \uparrow\Upsilon\langle e, \top\rangle \simeq \uparrow\Upsilon\langle f, \top\rangle && \text{By Claim 9.}\\
&\Rightarrow \uparrow\Upsilon\langle e, \top\rangle \equiv \uparrow\Upsilon\langle f, \top\rangle && \text{By Theorem 2.}\\
&\Rightarrow \downarrow\uparrow\Upsilon\langle e, \top\rangle \equiv \downarrow\uparrow\Upsilon\langle f, \top\rangle && \text{By Equation (4.4).}\\
&\Rightarrow \Upsilon\langle e, \top\rangle \equiv \Upsilon\langle f, \top\rangle && \text{By Equation (4.3).}\\
&\Rightarrow e \equiv f && \text{By Equation (4.2).}
\end{aligned}$$

◀

Hence, we only need to show Claim 9 to conclude. To that end, we show that for any clean expression e, any interpretation of $\uparrow e$ can be obtained by applying some transformation to some interpretation of e. Thanks to Corollary 8, we may restrict our attention to interpretation that avoid the empty word. This seemingly mundane restriction turns out to be of significant importance: if the empty word is allowed, the proof of Lemma 11 becomes much more involved. More precisely, we prove the following lemma:

▶ **Lemma 11.** *Let Σ be some set and $\sigma : X \times 2 \to \mathcal{L}\langle\Sigma\rangle$ some interpretation such that $\forall x, \varepsilon \notin \sigma(x)$. There exists an alphabet Σ', an interpretation $\sigma' : X \to \mathcal{L}\langle\Sigma'\rangle$ and a function $\psi : \mathcal{L}\langle\Sigma'\rangle \to \mathcal{L}\langle\Sigma\rangle$ such that: $\forall e \in \mathbb{C}_X,\ [\![\uparrow e]\!]_\sigma = \psi([\![e]\!]_{\sigma'})$.*

Proof. We fix Σ and $\sigma : X \times 2 \to \mathcal{L}\langle\Sigma\rangle$ as in the statement. Like in the proof of Proposition 4, we set $\Sigma' = \Sigma \cup \{\bullet\}$, with $\bullet$ a fresh letter, and write $\phi(u)$ for the word obtained from $u \in \Sigma'^\star$ by erasing every occurrence of $\bullet$. Additionally we define the function $\eta : \Sigma^\star \to \Sigma'^\star$ as follows:

$$\eta(\varepsilon) := \varepsilon \qquad\qquad \eta(a\,u) := \bullet\, a\, \eta(u) \quad (\langle a, u\rangle \in \Sigma \times \Sigma^\star).$$

Clearly, $\phi(\eta(u)) = u$ and $\eta(u\,v) = \eta(u)\,\eta(v)$. We may now define σ' and ψ:

$$\sigma'(x) := \{\eta(u) \mid u \in \sigma\langle x, \top\rangle\} \cup \left\{\overline{\eta(u)} \;\middle|\; u \in \sigma\langle x, \bot\rangle\right\} \qquad \psi(L) := \{u \mid \eta(u) \in L\}.$$

This is where the restriction $\varepsilon \notin \sigma(x)$ comes in. Indeed a word w cannot be written both as $w = \eta(u_1)$ and as $w = \overline{\eta(u_2)}$ unless $w = u_1 = u_2 = \varepsilon$. Since σ does not contain the empty word, we may show that $\psi(\sigma'(x)) = \sigma\langle x, \top\rangle$ and $\psi\left(\overline{\sigma'(x)}\right) = \sigma\langle x, \bot\rangle$.

ψ distributes over the union and intersection operators. However, it does not hold in general that $\psi(L \cdot M) = \psi(L) \cdot \psi(M)$. Like in the proof of Proposition 4 we will therefore identify a predicate on languages that is sufficient for this identity to hold, is satisfied by

$\sigma'(x)$, and is stable by $\cdot, \cap, +, (-)^+, \overline{(-)}$. In this case we find that an adequate candidate is "L contains only valid words", where the set $\mathbb{V}$ of valid words is defined as follows:

$$\frac{u \in \Sigma^+}{\eta(u) \in \mathbb{V}} \qquad \frac{u \in \mathbb{V}}{\overline{u} \in \mathbb{V}} \qquad \frac{u \in \mathbb{V} \quad v \in \mathbb{V}}{u\,v \in \mathbb{V}}$$

Alternatively, the elements of $\mathbb{V}$ are words over Σ' that can be written as a product $\alpha_1 \dots \alpha_n$ with $1 \leqslant n$ and each $\alpha_i \in (\Sigma \cdot \bullet) \cup (\bullet \cdot \Sigma)$. One may see from the definitions that $\sigma'(x) \subseteq \mathbb{V}$. $\mathbb{V}$ can also be seen to be trivially closed by concatenation and mirror image. Since the remaining operators are either idempotent (union and intersection) or derived (iteration), we get that $[\![e]\!]_{\sigma'} \subseteq \mathbb{V}$. This enables us to conclude thanks to the following property:

$$\forall u_1, u_2 \in \mathbb{V},\ \eta(u) = u_1\,u_2 \Rightarrow \exists v_1, v_2 : u_1 = \eta(v_1) \wedge u_2 = \eta(v_2) \wedge u = v_1\,v_2. \tag{4.5}$$

This property enables us to show that $\psi(L \cdot M) = \psi(L) \cdot \psi(M)$ and $\psi(L^+) = \psi(L)^+$, for languages of valid words L, M. Hence we obtain by induction on expressions that for any term $e \in \mathbb{C}_X$, it holds that $[\![\uparrow e]\!]_\sigma = [\![e]\!]_{\sigma'}$. ◀

▶ **Theorem 12.** $\forall e, f \in \mathbb{E}'_X,\ e \equiv f \Leftrightarrow e \simeq f$.

Proof. Thanks to Lemma 10, we only need to check Claim 9. Let e, f be two clean expressions such that $e \simeq f$, we want to prove $\uparrow e \simeq \uparrow f$. According to Corollary 8, we need to compare $[\![\uparrow e]\!]_\sigma$ and $[\![\uparrow f]\!]_\sigma$ for some $\sigma : X \times 2 \to \mathcal{L}\langle\Sigma\rangle$ such that $\varepsilon \notin \bigcup_{x \in X \times 2} \sigma(x)$. By Lemma 11, we may express these languages as respectively $\psi([\![e]\!]_{\sigma'})$ and $\psi([\![f]\!]_{\sigma'})$. Since $e \simeq f$, we get that $[\![e]\!]_{\sigma'} = [\![f]\!]_{\sigma'}$, thus proving the desired identity and concluding the proof. ◀

5 Interlude: tests

Before we start with the main proof, we define *tests* and establish a few result about them. Given a list of variables $u \in X^\star$, we define the term θ_u by induction on u as $\theta_\varepsilon := 1$ and $\theta_{a\,u} := a \cap \theta_u$. Thanks to the following remark, we will hereafter consider θ_A for $A \in \mathcal{P}_f(X)$:

▶ Remark 13. Let u, v be two lists of variables containing the same letters (meaning a variable appears in u if and only if it appears in v). Then $\theta_u \equiv \theta_v$.

The following property explains our choice of terminology: the function $\lambda\sigma. [\![\theta_A]\!]_\sigma$ can be seen as a boolean predicate testing whether the empty word is in each of the $\sigma(a)$ for $a \in A$.

▶ **Lemma 14.** *Let Σ be some alphabet and $\sigma : X \to \mathcal{L}\langle\Sigma\rangle$. Then either $\forall a \in A,\ \varepsilon \in \sigma(a)$, in which case $[\![\theta_A]\!]_\sigma = \varepsilon$, or $[\![\theta_A]\!]_\sigma = \emptyset$.*

Tests satisfy the following universal identities, with $A, B \in \mathcal{P}_f(X)$ and $e, f \in \mathbb{E}_X$:

$$\theta_A \leqq 1 \tag{5.1}$$
$$\theta_A \cap \theta_B \equiv \theta_A \cdot \theta_B \equiv \theta_{A \cup B} \tag{5.2}$$
$$\theta_A \equiv \theta_A \cdot \theta_A \tag{5.3}$$
$$a \in A \Rightarrow \theta_A \leqq a \tag{5.4}$$
$$\theta_A \cdot e \equiv e \cdot \theta_A \tag{5.5}$$
$$(\theta_A \cdot e) \cap (\theta_B \cdot f) \equiv \theta_{A \cup B} \cdot (e \cap f) \tag{5.6}$$
$$\theta_A^+ \equiv \overline{\theta_A} \equiv \theta_A. \tag{5.7}$$

We now want to compare tests with other tests or with expressions. Let us define the following interpretation for any finite set $A \in \mathcal{P}_f(X)$.

$$\begin{aligned} \sigma_A : \quad X &\to \mathcal{L}\langle\emptyset\rangle \\ x &\mapsto \begin{cases} \varepsilon & \text{if } x \in A \\ \emptyset & \text{otherwise.} \end{cases} \end{aligned}$$

Note that the alphabet here does not matter, since we only want the unit language and the empty language. This interpretation enables us to establish the following lemma:

▶ **Lemma 15.** *For any $A, B \in \mathcal{P}_f(X)$, the following are equivalent:*

(i) $\varepsilon \in [\![\theta_B]\!]_{\sigma_A}$ **(ii)** $B \subseteq A$ **(iii)** $\theta_A \leqq \theta_B$ **(iv)** $\theta_A \lesssim \theta_B$.

Proof. Assume (i) holds, i.e. $\varepsilon \in [\![\theta_B]\!]_{\sigma_A}$. By Lemma 14 this means that for every $a \in B$ we have $\varepsilon \in \sigma_A(a)$ which by definition of σ_A ensures that $a \in A$. Thus we have shown that (ii) holds. We show that (ii) implies (iii) by induction on the size of B:

- if $B = \emptyset$, by Equation (5.1) $\theta_A \leqq 1 = \theta_\emptyset$.
- if $B = \{a\} \cup B'$ with $a \notin B'$, since $B \subseteq A$ we have $a \in A$ and $B' \subseteq A$. By induction hypothesis we know that $\theta_A \leqq \theta'_B$. By Remark 13 we get that $\theta_A \equiv a \cap \theta_A$. Hence we get:

$$\theta_A \equiv a \cap \theta_A \leqq a \cap \theta'_B = \theta_B.$$

Thanks to soundness we have that (iii) implies (iv). For the last implication, notice that by construction of σ_A we have $\varepsilon \in [\![\theta_A]\!]_{\sigma_A}$. Therefore if $\theta_A \lesssim \theta_B$ then we can conclude that $\varepsilon \in [\![\theta_A]\!]_{\sigma_A} \subseteq [\![\theta_B]\!]_{\sigma_A}$. ◀

We now define a function $\mathbb{I} : \mathbb{E}_X \to \mathcal{P}_f(\mathcal{P}_f(X))$, whose purpose is to represent as a sum of tests the intersection of an arbitrary expression with 1:

$$\mathbb{I}(0) := \emptyset \qquad \mathbb{I}(1) := \{\emptyset\} \qquad \mathbb{I}(x) := \{\{x\}\} \qquad \mathbb{I}(e+f) := \mathbb{I}(e) \cup \mathbb{I}(f)$$

$$\mathbb{I}(e \cdot f) = \mathbb{I}(e \cap f) := \{A \cup B \mid \langle A, B\rangle \in \mathbb{I}(e) \times \mathbb{I}(f)\} \qquad \mathbb{I}(e^+) = \mathbb{I}(\overline{e}) := \mathbb{I}(e).$$

▶ **Lemma 16.** $\forall e \in \mathbb{E}_X,\ 1 \cap e \equiv \sum_{C \in \mathbb{I}(e)} \theta_C$.

▶ **Corollary 17.** $\forall e \in \mathbb{E}_X, \forall A \in \mathcal{P}_f(X),\ \theta_A \leqq e \Leftrightarrow \theta_A \lesssim e$.

Proof. We only need to show the implication from right to left. Assume $\theta_A \lesssim e$. This implies $1 \cap \theta_A \lesssim 1 \cap e$, and since $\theta_A \leqq 1$ we know that $1 \cap \theta_A \equiv \theta_A$ which by soundness implies $\theta_A \simeq 1 \cap \theta_A$. Combining this with Lemma 16, we get that $\theta_A \simeq 1 \cap \theta_A \lesssim 1 \cap e \simeq \sum_{C \in \mathbb{I}(e)} \theta_C$. By Lemma 15, we know that $\varepsilon \in [\![\theta_A]\!]_{\sigma_A}$, which means that $\varepsilon \in \left[\!\!\left[\sum_{C \in \mathbb{I}(e)} \theta_C\right]\!\!\right]_{\sigma_A} = \bigcup_{C \in \mathbb{I}(e)} [\![\theta_C]\!]_{\sigma_A}$. Therefore there must be some $B \in \mathbb{I}(e)$ such that $\varepsilon \in [\![\theta_B]\!]_{\sigma_A}$ which by Lemma 15 tells us that $\theta_A \leqq \theta_B$. We may now conclude:

$$\theta_A \leqq \theta_B \leqq \sum_{C \in \mathbb{I}(e)} \theta_C \equiv 1 \cap e \leqq e.$$ ◀

▶ Remark 18. The word "test" is reminiscent of Kleene algebra with tests (KAT)[7]. Indeed according to Equation (5.1) our tests are sub-units, like in KAT. However unlike in KAT, there are non-test terms t such that $t \leqq 1$. In general such terms are sums of tests, as can be inferred from Lemma 16 (because for every sub-unit $e \leqq 1$, we have $e \equiv 1 \cap e \equiv \sum_{C \in \mathbb{I}(e)} \theta_C$).

6 Completeness of reversible Kleene lattices

To tackle this completeness proof, we will proceed in three steps. Since we already proved soundness, and since an equality can be equivalently expressed as a pair of containments, we start from the following statement:

$$\forall e, f \in \mathbb{E}_X,\ e \lesssim f \Rightarrow e \leqq f.$$

First, we will show that any expression in $\mathbb{E}_X$ can be equivalently written as a sum of terms that are either tests or products $\theta_A \cdot e$ of a test and a one-free expression. The case of tests having been dispatched already (Corollary 17), this reduces the problem to:

$$\forall e \in \mathbb{E}'_X, \forall A \in \mathcal{P}_f(X), \forall f \in \mathbb{E}_X,\ \theta_A \cdot e \lesssim f \Rightarrow \theta_A \cdot e \leqq f.$$

Second, we will show that for any pair $\langle A, f\rangle \in \mathcal{P}_f(X) \times \mathbb{E}_X$, there exists an expression $\langle f\rangle_A \in \mathbb{E}_X$ such that $\theta_A \cdot \langle f\rangle_A \leqq f$ and whenever $\theta_A \cdot e \lesssim f$ we have $e \lesssim \langle f\rangle_A$. This further reduces the problem into:

$$\forall e \in \mathbb{E}'_X, \forall f \in \mathbb{E}_X,\ e \lesssim f \Rightarrow e \leqq f.$$

For the third and last step, we show that for any expression $f \in \mathbb{E}_X$, there is an expression $[f] \in \mathbb{E}'_X$ such that $[f] \leqq f$ and whenever $e \lesssim f$ for $e \in \mathbb{E}'_X$ we have $e \lesssim [f]$. This is enough to conclude thanks to Theorem 12.

In the next three subsections, we introduce constructions and prove lemmas necessary for each step. Then, in Section 6.4 we put them all together to show the main result.

6.1 First step: normal forms

A normal form is either an expression of the shape θ_A or of the shape $\theta_A \cdot e$ with $e \in \mathbb{E}'_X$. We denote by $\mathbb{NF}$ the set of normal forms. The main result of this section is the following:

▶ **Lemma 19.** *For any $e \in \mathbb{E}_X$ there exists a finite set $\mathcal{N}(e) \subseteq \mathbb{NF}$ such that $e \equiv \sum_{\eta \in \mathcal{N}(e)} \eta$.*

Proof. We show by induction on e how to build $\mathcal{N}(e)$. The correctness of the construction is fairly straightforward, and is left as an exercise : we will only state the relevant proof obligations when appropriate.

For constants, variables, and unions, the choice is rather obvious:

$$\mathcal{N}(0) := \emptyset \qquad \mathcal{N}(1) := \{\theta_\emptyset\} \qquad \mathcal{N}(x) := \{\theta_\emptyset \cdot x\} \qquad \mathcal{N}(e+f) := \mathcal{N}(e) \cup \mathcal{N}(f).$$

The case of mirror image is also rather straightforward:

$$\mathcal{N}(\overline{e}) := \{\theta_A \mid \theta_A \in \mathcal{N}(e)\} \cup \left\{\theta_A \cdot \overline{e'} \;\middle|\; \theta_A \cdot e' \in \mathcal{N}(e)\right\}.$$

For concatenations, we define the product $\eta \odot \gamma$ of two normal forms $\eta, \gamma \in \mathbb{NF}$ as:

$$\theta_A \odot \theta_B := \theta_{A\cup B} \qquad \theta_A \odot \theta_B \cdot e := \theta_A \cdot e \odot \theta_B := \theta_{A\cup B} \cdot e \qquad \theta_A \cdot e \odot \theta_B \cdot f := \theta_{A\cup B} \cdot (e \cdot f).$$

We then define $\mathcal{N}(e \cdot f) := \{\eta \odot \gamma \mid \langle \eta, \gamma\rangle \in \mathcal{N}(e) \times \mathcal{N}(f)\}$. For correctness of the construction, we would have to prove that $\forall \eta, \gamma \in \mathbb{NF},\ \eta \cdot \gamma \equiv \eta \odot \gamma$.

For intersections, we define $\otimes : \mathbb{NF} \times \mathbb{NF} \to \mathcal{P}_f(\mathbb{NF})$:

$$\theta_A \otimes \theta_B := \{\theta_{A\cup B}\} \qquad \theta_A \otimes \theta_B \cdot e := \theta_A \cdot e \otimes \theta_B := \{\theta_{A\cup B\cup C} \mid C \in \mathbb{I}(e)\}$$
$$\theta_A \cdot e \otimes \theta_B \cdot f := \theta_{A\cup B} \cdot (e \cap f).$$

We then define $\mathcal{N}(e \cap f) := \bigcup_{\langle \eta,\gamma\rangle \in \mathcal{N}(e)\times\mathcal{N}(f)} \eta \otimes \gamma$.

Finally, for iterations we use the following definition:

$$\mathcal{N}(e^+) := \{\theta_A \mid \theta_A \in \mathcal{N}(e)\} \cup \left\{ \theta_{\bigcup_i A_i} \cdot \left(\sum_i e_i\right)^+ \;\middle|\; \{\theta_{A_i} \cdot e_i \mid i \leqslant n\} \subseteq \mathcal{N}(e) \right\}. \qquad ◀$$

▶ Remark. In [1], a similar lemma was proved (Lemma 3.4). However, the proof in that paper is slightly wrong, as it fails to consider the cases $\theta_A \cap \theta_B$ (easy) and $\theta_A \cap \theta_B \cdot e$ (more involved).

6.2 Second step: removing tests on the left

Here we want to transform an inequation $\theta_A \cdot e \lesssim f$, into one one the shape $e \lesssim \langle f\rangle_A$, while maintaining that $\theta_A \cdot \langle f\rangle_A \leqq f$. The construction of $\langle f\rangle_A$ is fairly straightforward, the intuition being that θ_A forces us to only consider interpretations such that $a \in A \Rightarrow \varepsilon \in [\![a]\!]_\sigma$. Therefore, for any $a \in A$ we replace in f every occurrence of a with $1 + a$.

▶ **Lemma 20.** $\theta_A \cdot \langle f\rangle_A \leqq f \leqq \langle f\rangle_A$.

Proof. Since $a \leqq 1 + a$, we can show by induction that $f \leqq \langle f\rangle_A$. Also, if $a \in A$:

$$\begin{aligned} \theta_A \cdot (1+a) &\equiv \theta_A + \theta_A \cdot a && \text{By (1d.1) and (1b.4)}\\ &\equiv \theta_A \cdot \theta_A + \theta_A \cdot a && \text{By (5.3)}\\ &\equiv \theta_A \cdot (\theta_A + a) && \text{By (1b.4)}\\ &\equiv \theta_A \cdot a. && \text{By (5.4)} \end{aligned}$$

This proves for the case of variables that $\theta_A \cdot \langle f\rangle_A \leqq f$, and can be generalised to arbitrary expressions by a simple induction. ◀

For the other property, we rely on the following lemma:

▶ **Lemma 21.** *Let Ξ be some alphabet, and $\sigma : X \to \mathcal{L}\langle\Xi\rangle$ be an interpretation such that $\forall x \in X,\ \varepsilon \notin \sigma(x)$. Then $[\![\langle f\rangle_A]\!]_\sigma = [\![\langle f\rangle_A]\!]_\tau$, where* $\tau : \begin{array}{rcl} X & \to & \mathcal{L}\langle\Xi\rangle \\ x & \mapsto & \sigma(x) \cup \{\varepsilon \mid x \in A\}. \end{array}$

Proof. The result follows from a straightforward induction, the only interesting case being that of variables $x \in A$. This case is a simple consequence of our definitions:

$$[\![1+a]\!]_\tau = \varepsilon \cup \tau(a) = \varepsilon \cup \sigma(a) \cup \varepsilon = \varepsilon \cup \sigma(a) = [\![1+a]\!]_\sigma . \qquad ◀$$

▶ **Corollary 22.** *Let $\langle A, e\rangle \in \mathcal{P}_f(X) \times \mathbb{E}'_X$ such that $\theta_A \cdot e \lesssim f$, then $e \lesssim \langle f\rangle_A$.*

Proof. Since by Lemma 20 we have $f \leqq \langle f\rangle_A$ by soundness and transitivity of $\lesssim$ we have $\theta_A \cdot e \lesssim \langle f\rangle_A$. We want to show that $e \lesssim \langle f\rangle_A$, so by Corollary 8 we only need to check that for any interpretation $\sigma : X \to \mathcal{L}\langle\Sigma\rangle$ such that $\varepsilon \notin \bigcup_{x\in X}\sigma(x)$ we have $[\![e]\!]_\sigma \subseteq [\![\langle f\rangle_A]\!]_\sigma$. If we take τ like in Lemma 21, we get that 1) since for every variable $\sigma(x) \subseteq \tau(x)$, $[\![e]\!]_\sigma \subseteq [\![e]\!]_\tau$ and 2) since for every $a \in A$ we have $\varepsilon \in \tau(a)$, we get $[\![\theta_A]\!]_\tau = \varepsilon$. Together these tell us that $[\![e]\!]_\sigma \subseteq [\![e]\!]_\tau = \varepsilon \cdot [\![e]\!]_\tau = [\![\theta_A \cdot e]\!]_\tau$. Since $\theta_A \cdot e \lesssim \langle f\rangle_A$ we know that $[\![\theta_A \cdot e]\!]_\tau \subseteq [\![\langle f\rangle_A]\!]_\tau$, and by Lemma 21 we know $[\![\langle f\rangle_A]\!]_\sigma = [\![\langle f\rangle_A]\!]_\tau$. We may therefore conclude that $[\![e]\!]_\sigma \subseteq [\![\theta_A \cdot e]\!]_\tau \subseteq [\![\langle f\rangle_A]\!]_\tau = [\![\langle f\rangle_A]\!]_\sigma$. ◀

6.3 Third step: removing tests on the right

This last step relies on Proposition 4 and Lemma 19.

▶ **Lemma 23.** *For any expression $f \in \mathbb{E}_X$, there exists a one-free expression $[f] \in \mathbb{E}'_X$ such that $[f] \leqq f$ and for any one-free expression $e \in \mathbb{E}'_X$ such that $e \lesssim f$ we have $e \lesssim [f]$. In other words, $[f]$ is the maximum of the set $\{e \in \mathbb{E}'_X \mid e \leqq f\}$.*

Proof. We define $[f] := \sum_{\theta_\emptyset \cdot f' \in \mathcal{N}(f)} f'$. We can easily check that $[f] \leqq f$:

$$[f] \equiv 1 \cdot [f] = \theta_\emptyset \cdot \sum_{\theta_\emptyset \cdot f' \in \mathcal{N}(f)} f' \equiv \sum_{\theta_\emptyset \cdot f' \in \mathcal{N}(f)} \theta_\emptyset \cdot f' \leqq \sum_{\eta \in \mathcal{N}(f)} \eta \equiv f.$$

For the other property, we rely on Proposition 4. Assume $e \lesssim f$, we want to show that $e \lesssim [f]$. By Corollary 8, it is enough to check that $[\![e]\!]_\sigma \subseteq [\![[f]]\!]_\sigma$ for interpretations σ such that $\forall x \in X, \varepsilon \notin \sigma(x)$. Let σ be such an interpretation, and u some word such that $u \in [\![e]\!]_\sigma$. Notice that the condition on σ ensures that $\forall x \in X, 1 \cap \sigma(x) = \emptyset$, hence $[\![\theta_A]\!]_\sigma \neq \emptyset$ implies that $A = \emptyset$ by Lemma 14. Also, because $\sigma(x)$ never contains the empty word and e does not feature the constant 1, u must be different from ε. Since $e \lesssim f$, we already know that $u \in [\![f]\!]_\sigma$. By Lemma 19 and soundness, we know that there is a normal form $\eta \in \mathcal{N}(f)$ such that $u \in [\![\eta]\!]_\sigma$. Since $u \neq \varepsilon$, η cannot be a test: that would imply by (5.1) that $\eta \leqq 1$, hence $[\![\eta]\!]_\sigma \subseteq [\![1]\!]_\sigma = \varepsilon$. Therefore we know that there is a term $\theta_A \cdot f' \in \mathcal{N}(f)$ such that $u \in [\![\theta_A \cdot f']\!]_\sigma$. This means that $u \in [\![f']\!]_\sigma$ and $\varepsilon \in [\![\theta_A]\!]_\sigma$. As we have noticed before, this means that $A = \emptyset$. Thus we get $u \in [\![f']\!]_\sigma$ and $\theta_\emptyset \cdot f' \in \mathcal{N}(f)$, which ensures that $u \in [\![[f]]\!]_\sigma$. ◀

6.4 Main theorem

We may now prove the main result of this paper:

▶ **Theorem 24** (Main result). *$\forall e, f \in \mathbb{E}_X,\ e \equiv f \Leftrightarrow e \simeq f$.*

Proof. Since $e \equiv f \Leftrightarrow e \leqq f \wedge f \leqq e$ and $e \simeq f \Leftrightarrow e \lesssim f \wedge f \lesssim e$, we focus instead on proving that $e \leqq f \Leftrightarrow e \lesssim f$. By soundness we know that $e \leqq f \Rightarrow e \lesssim f$, so we only need to show the converse implication.

Let $e, f \in \mathbb{E}_X$ such that $e \lesssim f$. By Lemma 19 we can show that $e \equiv \sum_{\eta \in \mathcal{N}(e)} \eta$. Let $\eta \in \mathcal{N}(e)$. Thanks to the properties of $\lesssim$ we have that $\eta \lesssim f$. There are two cases for η:

- either $\eta = \theta_A$ for some $A \in \mathcal{P}_f(X)$, in which case we have $\eta \leqq f$ by Corollary 17;
- or $\eta = \theta_A \cdot e'$ with $A \in \mathcal{P}_f(X)$ and $e' \in \mathbb{E}'_X$. In that case, by Corollary 22 we have $e' \lesssim \langle f \rangle_A$, and by Lemma 23 we get $e' \lesssim [\langle f \rangle_A]$. Since both e' and $[\langle f \rangle_A]$ are one-free, we may apply Theorem 12 to get a proof that $e' \leqq [\langle f \rangle_A]$. Therefore

$$\begin{aligned} \eta = \theta_A \cdot e' \leqq \theta_A \cdot [\langle f \rangle_A] &\leqq \theta_A \cdot \langle f \rangle_A && \text{By Lemma 23.} \\ &\leqq f && \text{By Lemma 20.} \end{aligned}$$

In both cases we have established that $\eta \leqq f$, so by monotonicity we show that

$$e \equiv \sum_{\eta \in \mathcal{N}(e)} \eta \leqq \sum_{\eta \in \mathcal{N}(e)} f \leqq f.$$ ◀

7 The "top" problem

In reversible Kleene lattices, union and intersection form a distributive lattice, and 0 acts as both the unit of union and the annihilator of intersection. All that is missing to get a bounded distributive lattice is the unit of intersection and annihilator of union, namely the constant $\top$, to be interpreted as the full language. However, this turns out to be more complicated than one might think.

The first idea that comes to mind is to add the sole axiom $\top + e = \top$. This axiom just says that for any expression $e \leqq \top$, and is enough to show that $e \cap \top \equiv \top \cap e \equiv e$. It is obviously sound, so we get soundness of the resulting axiomatic equivalence. This axiomatic equivalence can be reduced without too much difficulty to that of reversible Kleene lattices, thanks to the following remark:

▶ Remark 25. If we write $\mathbb{E}_X^\top$ for expressions with $\top$, let $\phi : \mathbb{E}_X^\top \to \mathbb{E}_{X+1}$ be the function that replaces every occurrence of $\top$ with $\left(\sum_{a \in X+1}(a + \overline{a})\right)^\star$. Then the following identity holds: $\forall e, f \in \mathbb{E}_X^\top,\ e \equiv f \Leftrightarrow \phi(e) \equiv \phi(f)$.

This same construction, when applied to expressions without intersections, yields a completeness proof. In the presence of intersection however it is not complete. We illustrate this with two examples.

▶ **Example 26** (Levi's lemma)**.** Levi's lemma for strings [9] states that whenever we have two factorisations of the same word, i.e. $u_1\, u_2 = v_1\, v_2$, then either $\exists w,\ u_1 = v_1\, w \wedge v_2 = w\, u_2$ or $\exists w,\ v_1 = u_1\, w \wedge u_2 = w\, v_2$. If we now move from words to languages, it means that every word that can be obtained simultaneously as $L_1 \cdot L_2$ and $M_1 \cdot M_2$ also belongs to either $L_1 \cdot \top \cdot M_2$ or $M_1 \cdot \top \cdot L_2$. In other words, the following inequation holds:

$$(e_1 \cdot e_2) \cap (f_1 \cdot f_2) \lesssim (e_1 \cdot \top \cdot f_2) + (f_1 \cdot \top \cdot e_2)\,.$$

However this equation is not derivable. This law also contrasts with the properties we can observe in every fragment of this algebra that we have studied: in every case, if a term without $\star$ or $+$ is smaller than a term $e + f$, then it must be smaller than either e or f. One can plainly see that it is not the case here.

▶ **Example 27** (Factorisation)**.** Another troubling example is the following:

$$(a \cdot b) \cap (a \cdot c) \lesssim a \cdot ((\top \cdot b) \cap (\top \cdot c))\,.$$

As before, this inequation is valid, but it is not derivable, and it does not involve unions. This suggests that the (in-)equational theory of languages with just the signature $\langle \cdot, \cap, \top \rangle$ is already non-trivial. We believe that the key to adding $\top$ to Kleene lattices lies with a better understanding of the theory of this smaller signature.

References

1 Hajnal Andréka, Szabolcs Mikulás, and István Németi. The Equational Theory of Kleene Lattices. *Theor. Comput. Sci.*, 412(52):7099–7108, 2011. doi:10.1016/j.tcs.2011.09.024.

2 S. L. Bloom, Z. Ésik, and Gh. Stefanescu. Notes on Equational Theories of Relations. *algebra universalis*, 33(1):98–126, March 1995. doi:10.1007/BF01190768.

3 Paul Brunet. Reversible Kleene Lattices. In Kim G. Larsen, Hans L. Bodlaender, and Jean-Francois Raskin, editors, *42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017)*, volume 83 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 66:1–66:14, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.MFCS.2017.66.

4 John H. Conway. *Regular algebra and finite machines*. Chapman and Hall Mathematics Series, 2012.

5 Amina Doumane and Damien Pous. Completeness for Identity-Free Kleene Lattices. In Sven Schewe and Lijun Zhang, editors, *29th International Conference on Concurrency Theory (CONCUR 2018)*, volume 118 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 18:1–18:17, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.CONCUR.2018.18.

6 D. Kozen. A Completeness Theorem for Kleene Algebras and the Algebra of Regular Events. *Information and Computation*, 110(2):366–390, May 1994. doi:10.1006/inco.1994.1037.

7 Dexter Kozen. Kleene Algebra with Tests. *ACM Trans. Program. Lang. Syst.*, 19(3):427–443, 1997. doi:10.1145/256167.256195.

8 Daniel Krob. Complete Systems of B-Rational Identities. *Theoretical Computer Science*, 89(2):207–343, October 1991. doi:10.1016/0304-3975(91)90395-I.

9 Frank W. Levi. On semigroups. *Bull. Calcutta Math. Soc*, 36(141-146):82, 1944.