On the Complexity of Reasoning in Kleene Algebra

Dexter Kozen
Department of Computer Science
Upson Hall
Cornell University
Ithaca, NY 14853-7501, USA
kozen@cs.cornell.edu

Abstract

*We study the complexity of reasoning in Kleene algebra and *-continuous Kleene algebra in the presence of extra equational assumptions E; that is, the complexity of deciding the validity of universal Horn formulas $E \to s = t$, where E is a finite set of equations. We obtain various levels of complexity based on the form of the assumptions E. Our main results are: for *-continuous Kleene algebra,*

- *if E contains only commutativity assumptions $pq = qp$, the problem is Π_1^0-complete;*
- *if E contains only monoid equations, the problem is Π_2^0-complete;*
- *for arbitrary equations E, the problem is Π_1^1-complete.*

*The last problem is the universal Horn theory of the *-continuous Kleene algebras. This resolves an open question of Kozen (1994).*

1 Introduction

Kleene algebra (KA) is fundamental and ubiquitous in computer science. Since its invention by Kleene in 1956, it has arisen in various forms in program logic and semantics [17, 28], relational algebra [27, 32], automata theory [24, 25], and the design and analysis of algorithms [1, 15]. Many authors have contributed to the development of Kleene algebra over the years [2, 3, 4, 6, 7, 8, 12, 16, 17, 18, 21, 23, 25, 29, 30, 31]. On the practical side, KA provides a natural and effective tool for equational specification and verification. It has recently been used successfully in numerous applications involving basic safety analysis, low-level program transformations, and concurrency control [9, 10, 22].

1.1 Reasoning with Assumptions

The equational theory of KA alone is *PSPACE*-complete, and this is as efficient as one could expect. However, in practice, one often needs to reason in the presence of *assumptions* of various forms. For example, a *commutativity condition* $pq = qp$ models the fact that the programs p and q can be executed in either order with the same result. Such assumptions are needed to reason about basic program transformations such as constant propagation and moving static computations out of loops. In [22], several useful program transformations are given under commutativity assumptions of the form $pb = bp$, where p is a program and b is a test. This condition models the fact that the execution of the program p does not affect the value of the test b.

Assumptions of the form $pb = bp$ where b is a test do not increase the complexity of KA [11]. Unfortunately, slightly more general commutativity assumptions $pq = qp$, even for p and q atomic, may lead to undecidability. Cohen gave a direct proof of this fact encoding Post's Correspondence Problem (see [22]). This result can also be shown to follow from a 1979 result of Berstel [5] with a little extra work; see §4 below.

These considerations bring up the general theoretical question:

> *How hard is it to reason in Kleene algebra under assumptions of various forms?*

Equivalently and more formally,

> *What is the complexity of deciding the validity of universal Horn formulas $E \to s = t$, where E is a finite set of equations?*

Here "universal" refers to the fact that the atomic symbols of E, s, and t are implicitly universally quantified. This question is quite natural, since the axiomatization of KA is itself a universal Horn axiomatization.

The question becomes particularly interesting in the presence of *-continuity (KA^*). A Kleene algebra is **-continuous* if it satisfies the infinitary condition

$$pq^*r \quad = \quad \sup_{n\geq 0} pq^n r,$$

where the supremum is with respect to the natural order in the Kleene algebra. Not all Kleene algebras are *-continuous, but all known naturally occurring ones are. Moreover, although *-continuity often provides a convenient shortcut in equational proofs, there are no more equations provable with it than without it; that is, the equational theories of KA and KA^* coincide [21].

Because of these considerations, it has become common practice to adopt *-continuity as a matter of course. However, this is not without consequence: although the equational theories of KA and KA^* coincide, their Horn theories do not. Understanding where and how the theories diverge is essential to the understanding of the comparative power and limitations of reasoning in Kleene algebra with and without *-continuity.

1.2 Main Results

In this paper we explore these questions and provide answer to some of them. Our main results are summarized in Table 1. The shaded entries either were previously known or follow easily from known results. The unshaded results are new.

Perhaps the most remarkable of these results is E. This is the general question of the complexity of the universal Horn theory of the *-continuous Kleene algebras. This question was raised by the author in a LICS 1991 paper [19, 21], and has been open since that time. In fact, subject to reinterpretation, the question might even be attributed to Conway (1971) [12, p. 103]. Conway's conjecture is phrased somewhat ambiguously in terms of the universal Horn theory of the regular sets; a literal interpretation is relatively easy to refute [18]. However, it is likely that Conway had the more substantive formulation E in mind.

That the universal Horn theory of KA^* should be so highly complex may be quite surprising in light of the utter simplicity of the axiomatization. We are aware of no other purely equational system with such high complexity. There are a few examples of Π^1_1-completeness results in Propositional Dynamic Logic (PDL), but PDL is a relatively more sophisticated two-sorted system and takes significant advantage of a restricted semantics involving only relational models. Here we make no such restriction: a Kleene algebra or *-continuous Kleene algebra is any algebraic structure satisfying the axioms of §2.1.

1.3 A Universality Property

A cornerstone of our approach is a certain *universality property* relating the class of *-continuous Kleene algebras and a restricted subclass consisting of algebras of the form $\mathsf{REG}\,M$, the regular subsets of an arbitrary monoid M.

Formally, the universality property says that any monoid homomorphism $h : M \to K$ from a monoid M to the multiplicative monoid of a *-continuous Kleene algebra K extends uniquely to a Kleene algebra homomorphism $\widehat{h} : \mathsf{REG}\, M \to K$. In category-theoretic terms, the map $M \mapsto \mathsf{REG}\, M$ constitutes a left adjoint to the forgetful functor taking a *-continuous Kleene algebra to its multiplicative monoid.

In practice, this property will allow us to restrict our attention to algebras of the form $\mathsf{REG}\,\Sigma^*/E$ when dealing with universal Horn formulas $E \to s = t$, where E consists of monoid equations.

This relationship exists only because of the fact that in *-continuous Kleene algebras, suprema of definable sets exist [20, Lemma 7.1, p. 35]. It is not valid for Kleene algebras in general.

We develop this connection in more detail in §2.3.

1.4 Other Results

The results D and H apply to Kleene algebras with tests and were proved in [11]. The decision problems in the column labeled KA are all r.e. because of the finitary axiomatization of KA given in §2.1. The r.e.-hardness of A and B follows from the fact that these problems encode the word problem for finitely presented monoids, shown r.e.-hard independently by Post and Markov in 1947 (see [13, Theorem 4.3, p. 98]). The *EXPSPACE*-hardness of C follows from the *EXPSPACE*-hardness of the word problem for commutative monoids [26]. It is not known whether C is decidable.

2 Preliminary Definitions

2.1 Kleene Algebra

A *Kleene algebra* is a structure $(K,\ +,\ \cdot,\ ^*,\ 0,\ 1)$ satisfying the following equations and equational implications:

$$\begin{aligned} p + (q + r) &= (p + q) + r \\ p + q &= q + p \\ p + 0 = p + p &= p \\ p(qr) &= (pq)r \end{aligned}$$

Form of assumptions	KA	KA*
unrestricted	Ⓐ Σ^0_1-complete	Ⓔ Π^1_1-complete
monoid equations	Ⓑ Σ^0_1-complete	Ⓕ Π^0_2-complete
$pq = qp$	Ⓒ *EXPSPACE*-hard	Ⓖ Π^0_1-complete
$pb = bp$	Ⓓ *PSPACE*-complete	Ⓗ *PSPACE*-complete

Table 1. Main Results

$$\begin{array}{rcl} 1p = p1 & = & p \\ 0p = p0 & = & 0 \\ p(q+r) & = & pq + pr \\ (p+q)r & = & pr + qr \\ 1 + pp^* & \leq & p^* \\ 1 + p^*p & \leq & p^* \\ px \leq x & \to & p^*x \leq x \\ xp \leq x & \to & xp^* \leq x \end{array}$$

where $\leq$ refers to the natural partial order:

$$p \leq q \quad \stackrel{\text{def}}{\Longleftrightarrow} \quad p + q = q.$$

We abbreviate $p \cdot q$ as pq and avoid parentheses by assigning the precedence $^* > \cdot > +$ to the operators.

A Kleene algebra is *-*continuous* if

$$pq^*r \quad = \quad \sup_{n \geq 0} pq^n r \qquad (1)$$

where $q^0 = 1$, $q^{n+1} = qq^n$, and the supremum is with respect to the natural order $\leq$. The *-continuity condition (1) can be regarded as the conjunction of infinitely many axioms $pq^nr \leq pq^*r$, $n \geq 0$, and the infinitary Horn formula

$$\bigwedge_{n \geq 0} (pq^n r \leq y) \quad \to \quad pq^*r \leq y. \qquad (2)$$

The category of Kleene algebras and Kleene algebra homomorphisms is denoted KA. The full subcategory of *-continuous Kleene algebras is denoted KA*.

A *term* is just a regular expression over some finite alphabet Σ. Terms are denoted $s, t, u, \ldots$. An *interpretation* over a Kleene algebra K is a map $I : \Sigma \to K$. Every interpretation I extends uniquely to a homomorphism $I : \{\text{terms over } \Sigma\} \to K$. An equation $s = t$ is *true* under interpretation I if $I(s) = I(t)$. More generally, if E is a set of equations, the Horn formula $E \to s = t$ is *true* under I if either $I(s) = I(t)$ or some equation of E is not true under I. We write

$$K, I \quad \vDash \quad \varphi$$

if φ is true in K under I. We write

$$\mathsf{KA} \quad \vDash \quad \varphi$$

if φ is true in all Kleene algebras under all interpretations. The *equational theory of Kleene algebras*, denoted $\mathcal{E}\,$KA, is the set of equations true in all Kleene algebras under all interpretations. The *universal Horn theory of Kleene algebras*, denoted $\mathcal{H}\,$KA, is the set of all finite equational implications $E \to s = t$ true in all Kleene algebras under all interpretations.

Similar definitions hold for the *-continuous Kleene algebras, using KA* in place of KA.

2.2 Regular Sets over a Monoid

Let M be a monoid with identity 1_M. The powerset 2^M forms a natural *-continuous Kleene algebra under the operations

$$\begin{array}{rcl} A + B & = & A \cup B \\ AB & = & \{xy \mid x \in A,\ y \in B\} \\ A^* & = & \bigcup_{n \geq 0} A^n \\ 0 & = & \varnothing \\ 1 & = & \{1_M\} \end{array}$$

The injection $\rho_M : x \mapsto \{x\}$ is a monoid homomorphism embedding M into the multiplicative monoid of 2^M.

Now let REG M denote the smallest Kleene subalgebra of 2^M containing the image of M under the map ρ_M. This is a *-continuous Kleene algebra and is called the *algebra of regular sets over* M.

For the free monoid Σ^* over the finite alphabet Σ, the Kleene algebra REG Σ^* is the family of regular sets of strings over Σ in the usual sense.

2.3 The Functor REG

The map $M \mapsto$ REG M, along with the map that associates with every monoid homomorphism $h : M \to M'$ the Kleene algebra homomorphism $\mathsf{REG}\,h : \mathsf{REG}\,M \to \mathsf{REG}\,M'$ defined by

$$\mathsf{REG}\,h\ (A) \quad \stackrel{\text{def}}{=} \quad \{h(x) \mid x \in A\},$$

constitute a functor REG from the category of monoids and monoid homomorphisms to the category KA^* of *-continuous Kleene algebras and Kleene algebra homomorphisms.

$$\begin{tikzcd} M \arrow[r, "h"] \arrow[d, "\rho_M"'] & M' \arrow[d, "\rho_{M'}"] \\ \mathsf{REG}\,M \arrow[r, "\mathsf{REG}\,h"] & \mathsf{REG}\,M' \end{tikzcd} \tag{3}$$

The functor REG is the left adjoint of the forgetful functor that takes a *-continuous Kleene algebra to its multiplicative monoid. This implies that any monoid homomorphism $h : M \to K$ from a monoid M to the multiplicative monoid of a *-continuous Kleene algebra K extends uniquely through ρ_M to a Kleene algebra homomorphism $\widehat{h} : \mathsf{REG}\,M \to K$:

$$\begin{tikzcd} M \arrow[r, "h"] \arrow[d, "\rho_M"'] & K \\ \mathsf{REG}\,M \arrow[ur, "\widehat{h}"'] & \end{tikzcd} \tag{4}$$

The homomorphism $\widehat{h}$ is defined as follows:

$$\widehat{h}(A) \quad \stackrel{\text{def}}{=} \quad \sup\,\{h(x) \mid x \in A\}.$$

This works for *-continuous Kleene algebras because of [20, Lemma 7.1, p. 35], which says that *suprema of all definable subsets of a *-continuous Kleene algebra exist.* It does not work for Kleene algebras in general.

3 Encoding Turing Machines

The lower bound proofs for E, F, and G depend partially on encoding Turing machine computations as monoid equations. This construction is standard. We sketch it here for completeness and because we need the equations in a particular form for the applications to follow. We follow the treatment of Davis [13].

Without loss of generality, we consider only deterministic Turing machines M that conform to the following restrictions.

- M has input alphabet $\{a\}$ and finite tape alphabet Γ containing a and a special blank symbol $\sqcup$ different from a (and possibly other symbols as well).
- It has a finite set of states Q disjoint from Γ containing a start state s. It may also contain one or more halt states.
- There are no transitions into the start state s and no transitions out of any halt state. Thus, once M enters a halt state, it cannot proceed.
- It has a single two-way-infinite read-write tape, padded on the left and right by infinitely many blanks $\sqcup$.
- M never writes a blank symbol between two non-blank symbols.

Let $\vdash$, $\dashv$ be two special symbols that are not in Γ or Q. Let

$$\Delta \quad \stackrel{\text{def}}{=} \quad \Gamma \cup Q \cup \{\vdash, \dashv\}.$$

A *configuration* is a string in Δ^* of the form $\vdash xqy \dashv$, where $x, y \in \Gamma^*$ and $q \in Q$. Configurations describe instantaneous global descriptions of M in the course of some computation. In the configuration $\vdash xqy \dashv$, the current state is q, the tape currently contains xy surrounded by infinitely many blanks $\sqcup$ on either side, and the machine is scanning the first symbol of y. If y is null, then the machine is assumed to be scanning the blank symbol immediately to the right of x, although that blank symbol need not be explicitly represented in the configuration.

The symbols $\vdash$ and $\dashv$ are *not* part of M's tape alphabet, but only a device to mark the ends of configurations and to create extra blank symbols to the right and left of the input if required; more on this below.

Each transition of M is of the form $(p, a) \to (b, d, q)$, which means, "when in state p scanning symbol a, print b, move the tape head one cell in direction $d \in \{\text{left}, \text{right}\}$, and enter state q."

Now consider the following equations on Δ^*:

(E1) for each transition $(p, a) \to (b, \text{right}, q)$ of M, the equation $pa = bq$;

(E2) for each transition $(p, a) \to (b, \text{left}, q)$ of M and each $c \in \Gamma$, the equation $cpa = qcb$;

(E3) the equations $\vdash\, = \,\vdash \sqcup$ and $\dashv\, = \,\sqcup \dashv$.

Equations (E3) allow us to create extra blank symbols to the left and right of the input any time we need them and to destroy them if we do not.

For $x, y \in \Delta^*$, we write $x \approx y$ if x and y are congruent modulo (E1)–(E3), and we write $x \sim y$ if x and y are congruent modulo (E3) only.

Lemma 1 *If $x, y \in \Gamma$ and t is a halt state, then*

$$\begin{aligned} \vdash xsy \dashv \; &\approx \; \vdash ztw \dashv & (5) \\ \iff \quad \vdash xsy \dashv \; &\xrightarrow[M]{*} \; \vdash ztw \dashv . & (6) \end{aligned}$$

Proof. See [13, Theorem 4.3, p. 98]. The chief concern is that monoid equations are reversible, whereas computations are not; thus it is conceivable that (5) holds by some complicated sequence of substitutions modeling a zigzagging forwards-and-backwards computation even when (6) does not. It can be shown that since M is deterministic and there are no transitions out of state t, this cannot happen. $\square$

4 Monoid Equations

In this section we indicate how to take advantage of the universality property (4) of §2.3 to obtain results F and G.

Let Σ be a finite alphabet. Let E be a finite set of equations between words in Σ^*, the free monoid over Σ. Let s, t be regular expressions over Σ. Let Σ^*/E denote the quotient monoid. For $x \in \Sigma^*$, let $[x]$ denote the E-congruence class of x in Σ^*/E. The map $\iota : a \mapsto \{[a]\}$ constitutes an interpretation over the *-continuous Kleene algebra $\mathsf{REG}\,\Sigma^*/E$, called the *standard interpretation.*

Lemma 2 *The following are equivalent:*

(i) $\mathsf{KA}^* \vDash E \to s = t$*; that is, the Horn formula* $E \to s = t$ *is true in all *-continuous Kleene algebras under all interpretations;*

(ii) $\mathsf{REG}\,\Sigma^*/E,\ \iota \vDash s = t$.

Proof. It is easily verified that $\mathsf{REG}\,\Sigma^*/E$ satisfies E under the standard interpretation ι. The implication (i) $\Rightarrow$ (ii) follows.

Conversely, for (ii) $\Rightarrow$ (i), let I be any interpretation into a *-continuous Kleene algebra K satisfying E. The monoid homomorphism $I : \Sigma^* \to K$ factors as $I = I' \circ [\,]$, where $I' : \Sigma^*/E \to K$. The universality property (4) then implies that I', hence I, factors through $\mathsf{REG}\,\Sigma^*/E$.

$$\begin{tikzcd}
\Sigma^* \arrow[d, "{[\,]}"'] \arrow[dr, "I"] \arrow[dd, bend right=60, "\iota"'] & \\
\Sigma^*/E \arrow[r, "I'"] \arrow[d, "\rho_{\Sigma^*/E}"'] & K \\
\mathsf{REG}\,\Sigma^*/E \arrow[ur] &
\end{tikzcd}$$

Thus any equation true in $\mathsf{REG}\,\Sigma^*/E$ under interpretation ι is also true in K under I. $\square$

This result allows us to restrict our attention to $\mathsf{REG}\,\Sigma^*/E$ for the purpose of proving F and G.

Theorem 3 *Consider the problem of deciding whether a given Horn formula $E \to s = t$ is true in all *-continuous Kleene algebras.*

(i) When E consists of commutativity conditions (or for that matter, any monoid equations $x = y$ such that $|x| = |y|$), the problem is Π^0_1-complete.

(ii) When E consists of arbitrary monoid equations $x = y$, the problem is Π^0_2-complete.

Proof. Using Lemma 2 and expressing an equation as the conjunction of two inequalities, we can reduce the problem to the conjunction of two instances of

$$\mathsf{REG}\,\Sigma^*/E,\ \iota \quad \vDash \quad s \le t. \tag{7}$$

The upper bounds for both (i) and (ii) are obtained by expressing (7) as a first-order formula with the appropriate quantifier prefix. Let $\equiv$ denote congruence modulo E on Σ^*. Applying (3) with $M = \Sigma^*$ and $M' = \Sigma^*/E$, (7) can be expressed

$$\forall x\ x \in \rho_{\Sigma^*}(s) \to \exists y\ y \equiv x \wedge y \in \rho_{\Sigma^*}(t). \tag{8}$$

The predicates $x \in \rho_{\Sigma^*}(s)$ and $y \in \rho_{\Sigma^*}(t)$ are decidable, and efficiently so: this is just string matching with regular expressions. Thus the formula (8) is Π^0_2 formula. Moreover, if all equations in E are length-preserving, then the existential subformula

$$\exists y\ y \equiv x \wedge y \in \rho_{\Sigma^*}(t)$$

is decidable, so (8) is equivalent to a Π^0_1 formula.

The lower bound for (i) uses the characterization of Lemma 2 and the result of Berstel [5] (see also [14, 22]) that (7) is undecidable. The reductions given in the cited references show that (7) is Π^0_1-hard. This result holds even when E consists only of commutativity conditions of the form $pq = qp$ for atomic p and q.

We prove the lower bound for (ii) by encoding the totality problem for Turing machines; that is, whether a given Turing machine halts on all inputs. Let M be a Turing machine of the form described in §3 with a single halt state t. Assume without loss of generality that M erases its tape before halting. The totality problem is to decide whether

$$\vdash sa^n \dashv \quad \xrightarrow[M]{*} \quad \vdash t \dashv, \quad n \ge 0.$$

This is a well-known Π^0_2-complete problem. By Lemma 1, this is true iff

$$\mathsf{REG}\,\Delta^*/E,\ \iota \quad \vDash \quad \vdash sa^n \dashv \; = \; \vdash t \dashv, \quad n \ge 0,$$

where E consists of equations (E1)–(E3) of §3. This is equivalent to

$$\mathsf{REG}\,\Delta^*/E,\ \iota \quad \models \quad \vdash sa^n \dashv\ \leq\ \vdash t \dashv, \quad n \geq 0,$$

since $\{x\} \subseteq \{y\}$ iff $x = y$. By the *-continuity condition (1), this is true iff

$$\mathsf{REG}\,\Delta^*/E,\ \iota \quad \models \quad \vdash sa^* \dashv\ \leq\ \vdash t \dashv,$$

and by Lemma 2, this is true iff

$$\mathsf{KA}^* \quad \models \quad E\ \rightarrow\ \vdash sa^* \dashv\ \leq\ \vdash t \dashv\ .$$

□

5 Π^1_1-Completeness of $\mathcal{H}\,\mathsf{KA}^*$

In this section we prove that the universal Horn theory of the *-continuous Kleene algebras is Π^1_1-complete.

Let $G = (\omega,\ R)$ be a recursive directed graph on vertices ω, the natural numbers. For $m \in \omega$, denote by $R(m)$ the set of R-successors of m:

$$R(m) \quad = \quad \{n \mid (m, n) \in R\}.$$

Let $\mathsf{WF} \subseteq \omega$ be the set of all m such that all R-paths out of m are finite. Alternatively, we could define WF as the least fixpoint of the following recursive equation:

$$\mathsf{WF} \quad = \quad \{m \mid R(m) \subseteq \mathsf{WF}\}.$$

Let us call G *well-founded* if $0 \in \mathsf{WF}$; that is, if all R-paths out of 0 are finite.

A well-known Π^1_1-complete problem is:

> *Given a recursive graph (say by a total Turing machine accepting the set of encodings of edges $(m, n) \in R$), is it well-founded?*

We reduce this problem to $\mathcal{H}\,\mathsf{KA}^*$, thereby showing that the latter problem is Π^1_1-hard.

By assumption, R is a recursive set, thus there is a total deterministic Turing machine M that decides whether $(m, n) \in R$. We can assume without loss of generality that M satisfies the restrictions of §3 and operates as follows.

In addition to its start state s, M has three halt states t, r, u. When started in configuration $\vdash a^m sa^n \dashv$, it first performs a check that the tape initially contains a contiguous string of a's surrounded by blanks and enters halt state u if not. It then determines whether $(m, n) \in R$. If so, it halts in configuration $\vdash a^n t \dashv$, and if not, it halts in configuration $\vdash r \dashv$. Thus

$$\vdash a^m sa^n \dashv \quad \xrightarrow[M]{*} \quad \begin{cases} \vdash a^n t \dashv & \text{if } (m, n) \in R \\ \vdash r \dashv & \text{if } (m, n) \notin R. \end{cases}$$

By Lemma 1, we have

$$\begin{aligned} \vdash a^m sa^n \dashv\ \approx\ \vdash a^n t \dashv \quad &\iff \quad (m, n) \in R, \\ \vdash a^m sa^n \dashv\ \approx\ \vdash r \dashv \quad &\iff \quad (m, n) \notin R, \end{aligned}$$

where $\approx$ denotes congruence modulo equations (E1)–(E3) of §3.

Now consider the Kleene algebra equation

$$t \quad \leq \quad sa^*. \tag{9}$$

Let E be the set of equations (E1)–(E3) together with (9).

The following is our main lemma.

Lemma 4 *For all $m \geq 0$,*

$$\mathsf{KA}^* \quad \models \quad E \quad \rightarrow \quad \vdash a^m t \dashv\ \leq\ \vdash r \dashv$$

if and only if $m \in \mathsf{WF}$.

Proof. The reverse implication ($\Leftarrow$) is proved by transfinite induction on the stages of the inductive definition of WF. Suppose that $m \in \mathsf{WF}$. Let $\tau : 2^\omega \rightarrow 2^\omega$ be the monotone map

$$\tau(A) \quad = \quad \{m \mid R(m) \subseteq A\}$$

and define

$$\begin{aligned} \tau^0(A) \quad &= \quad A \\ \tau^{\alpha+1}(A) \quad &= \quad \tau(\tau^\alpha(A)) \\ \tau^\lambda(A) \quad &= \quad \bigcup_{\alpha<\lambda} \tau^\alpha(A), \quad \lambda \text{ a limit ordinal.} \end{aligned}$$

Then

$$\mathsf{WF} \quad = \quad \bigcup_\alpha \tau^\alpha(\varnothing).$$

Let α be the smallest ordinal such that $m \in \tau^\alpha(\varnothing)$. Then α must be a successor ordinal $\beta + 1$, therefore $m \in \tau(\tau^\beta(\varnothing))$, so $R(m) \subseteq \tau^\beta(\varnothing)$. By the induction hypothesis, if $n \in R(m)$, then

$$\mathsf{KA}^* \quad \models \quad E\ \rightarrow\ \vdash a^n t \dashv\ \leq\ \vdash r \dashv,$$

and $\vdash a^m sa^n \dashv\ \approx\ \vdash a^n t \dashv$, therefore

$$\mathsf{KA}^* \quad \models \quad E\ \rightarrow\ \vdash a^m sa^n \dashv\ \leq\ \vdash r \dashv\ .$$

For $n \notin R(m)$, $\vdash a^m sa^n \dashv\ \approx\ \vdash r \dashv$. Thus for all n,

$$\mathsf{KA}^* \quad \models \quad E\ \rightarrow\ \vdash a^m sa^n \dashv\ \leq\ \vdash r \dashv\ .$$

By *-continuity,

$$\mathsf{KA}^* \quad \models \quad E\ \rightarrow\ \vdash a^m sa^* \dashv\ \leq\ \vdash r \dashv,$$

and by (9),

$$\mathsf{KA}^* \quad \models \quad E \;\to\; \vdash a^m t \dashv \;\leq\; \vdash r \dashv\;.$$

Conversely, for the forward implication ($\Rightarrow$), we construct a particular interpretation satisfying E in which for all $m \in \omega$, $\vdash a^m t \dashv \;\leq\; \vdash r \dashv$ implies $m \in \mathsf{WF}$.

For $A \subseteq \Delta^*$, define the monotone map

$$\begin{aligned} \sigma(A) \quad &= \quad A \cup \{x \mid \exists y \in A\ x \approx y\} \\ &\qquad \cup \{utv \mid \forall n\ usa^n v \in A\}. \end{aligned} \tag{10}$$

Call a subset of Δ^* *closed* if it is closed under the operation σ. The *closure* of A is the smallest closed set containing A and is denoted $\overline{A}$. Build a Kleene algebra consisting of the closed sets with operations

$$\begin{aligned} A \oplus B \quad &= \quad \overline{A \cup B} \\ A \odot B \quad &= \quad \overline{AB} \\ A^{\circledast} \quad &= \quad \overline{\bigcup_n A^n} \\ 0 \quad &= \quad \varnothing \\ 1 \quad &= \quad \{\epsilon\}, \end{aligned}$$

where ϵ is the null string and A^n is the n^{th} power of A under the operation $\odot$. It is not difficult to show that the family of closed sets forms a *-continuous Kleene algebra under these operations.

We show now that under the interpretation $a \mapsto \overline{\{a\}}$, the equations E are satisfied. For an equation $x = y$ of type (E1)–(E3), we need to show that $\overline{\{x\}} = \overline{\{y\}}$. It suffices to show that $x \in \overline{\{y\}}$ and $y \in \overline{\{x\}}$. But since $x \approx y$, this follows immediately from (10).

For the equation $t \leq sa^*$, we need to show that

$$t \in \overline{\{s\}} \odot \overline{\bigcup_n \overline{\{a\}}^n}.$$

It suffices to show $t \in \overline{\{sa^n \mid n \geq 0\}}$. Again, this follows immediately from (10).

Finally, we show that for $x \in \overline{\{\vdash r \dashv\}}$, either

(i) $x \xrightarrow[M]{*} \;\vdash r \dashv$;

(ii) $x \xrightarrow[M]{*} \;\vdash a^n t \dashv$ for some $n \in \mathsf{WF}$; or

(iii) $x \;\sim\; \vdash a^n t a^k \dashv$ for some $k \geq 1$.

The argument proceeds by transfinite induction on the inductive definition of closure:

$$\begin{aligned} \sigma^0(A) \quad &= \quad A \\ \sigma^{\alpha+1}(A) \quad &= \quad \sigma(\sigma^\alpha(A)) \\ \sigma^\lambda(A) \quad &= \quad \bigcup_{\alpha<\lambda} \sigma^\alpha(A), \qquad \lambda \text{ a limit ordinal} \\ \overline{A} \quad &= \quad \bigcup_\alpha \sigma^\alpha(A). \end{aligned}$$

Let α be the least ordinal such that

$$x \quad \in \quad \sigma^\alpha(\{\vdash r \dashv\}).$$

Then α must be a successor ordinal $\beta + 1$, thus

$$x \quad \in \quad \sigma(\sigma^\beta(\{\vdash r \dashv\})).$$

There are two cases, one for each clause in the definition (10) of σ.

If there exists $y \in \sigma^\beta(\{\vdash r \dashv\})$ such that $x \approx y$, then by the induction hypothesis, y satisfies one of (i)–(iii), therefore so does x; the argument here is similar to [13, Theorem 4.3, p. 98].

Otherwise, $x = utv$ and

$$usa^n v \quad \in \quad \sigma^\beta(\{\vdash r \dashv\})$$

for all n. By the induction hypothesis, one of (i)–(iii) holds for each $usa^n v$. But (iii) is impossible because of the form of (E3). Moreover, by construction of M, each of (i) and (ii) implies that $u \sim\; \vdash a^m$ and $v \sim a^k \dashv$ for some k, m. Thus $x \sim\; \vdash a^m t a^k \dashv$. If $k \geq 1$, then x satisfies (iii). Otherwise, $x \sim\; \vdash a^m t \dashv$ and

$$\vdash a^m s a^n \dashv \quad \in \quad \sigma^\beta(\{\vdash r \dashv\})$$

for all n, therefore either (i) or (ii) holds for $\vdash a^m s a^n \dashv$. If (i), then $(m, n) \notin R$. If (ii), then $(m, n) \in R$ and $n \in \mathsf{WF}$. Thus $R(m) \subseteq \mathsf{WF}$ and $m \in \mathsf{WF}$. □

Theorem 5 *$\mathcal{H}\,\mathsf{KA}^*$ is Π^1_1-complete.*

Proof. Taking $m = 0$ in Lemma 4, we have

$$\mathsf{KA}^* \quad \models \quad E \;\to\; \vdash t \dashv \;\leq\; \vdash r \dashv$$

if and only if G is well-founded. This gives the desired lower bound. The upper bound follows from the form of the infinitary axiomatization of *-continuous Kleene algebra (2); validity is equivalent to the existence of a well-founded proof tree. □

References

[1] A. V. Aho, J. E. Hopcroft, and J. D. Ullman. *The Design and Analysis of Computer Algorithms.* Addison-Wesley, 1975.

[2] S. Anderaa. On the algebra of regular expressions. *Appl. Math.*, pages 1–18, January 1965.

[3] K. V. Archangelsky. A new finite complete solvable quasiequational calculus for algebra of regular languages. Manuscript, Kiev State University, 1992.

[4] R. C. Backhouse. *Closure Algorithms and the Star-Height Problem of Regular Languages.* PhD thesis, Imperial College, 1975.

[5] J. Berstel. *Transductions and Context-free Languages.* Teubner, 1979.

[6] S. L. Bloom and Z. Ésik. Equational axioms for regular sets. *Mathematical Structures in Computer Science*, 3:1–24, 1993.

[7] M. Boffa. Une remarque sur les systèmes complets d'identités rationnelles. *Informatique Théoretique et Applications/Theoretical Informatics and Applications*, 24(4):419–423, 1990.

[8] E. Cohen. Hypotheses in Kleene algebra. Unpublished, April 1994.

[9] E. Cohen. Lazy caching. Unpublished, 1994.

[10] E. Cohen. Using Kleene algebra to reason about concurrency control. Unpublished, 1994.

[11] E. Cohen, D. C. Kozen, and F. Smith. The complexity of Kleene algebra with tests. Technical Report TR96-1598, Cornell University, July 1996.

[12] J. H. Conway. *Regular Algebra and Finite Machines.* Chapman and Hall, 1971.

[13] M. Davis. *Computability and Unsolvability.* McGraw-Hill, 1958.

[14] A. Gibbons and W. Rytter. On the decidability of some problems about rational subsets of free partially commutative monoids. *Theoretical Computer Science*, 48:329–337, 1986.

[15] K. Iwano and K. Steiglitz. A semiring on convex polygons and zero-sum cycle problems. *SIAM J. Comput.*, 19(5):883–901, 1990.

[16] S. C. Kleene. Representation of events in nerve nets and finite automata. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, pages 3–41. Princeton University Press, 1956.

[17] D. C. Kozen. On induction vs. *-continuity. In Kozen, editor, *Proc. Workshop on Logic of Programs*, volume 131 of *Lect. Notes in Comput. Sci.*, pages 167–176. Springer-Verlag, 1981.

[18] D. C. Kozen. On Kleene algebras and closed semirings. In Rovan, editor, *Proc. Math. Found. Comput. Sci.*, volume 452 of *Lect. Notes in Comput. Sci.*, pages 26–47. Springer-Verlag, 1990.

[19] D. C. Kozen. A completeness theorem for Kleene algebras and the algebra of regular events. In *Proc. 6th Symp. Logic in Comput. Sci.*, pages 214–225. IEEE, July 1991.

[20] D. C. Kozen. *The Design and Analysis of Algorithms.* Springer-Verlag, 1991.

[21] D. C. Kozen. A completeness theorem for Kleene algebras and the algebra of regular events. *Infor. and Comput.*, 110(2):366–390, May 1994.

[22] D. C. Kozen. Kleene algebra with tests and commutativity conditions. In T. Margaria and B. Steffen, editors, *Proc. Second Int. Workshop Tools and Algorithms for the Construction and Analysis of Systems (TACAS'96)*, volume 1055 of *Lect. Notes in Comput. Sci.*, pages 14–33. Springer-Verlag, March 1996.

[23] D. Krob. A complete system of B-rational identities. *Theoretical Computer Science*, 89(2):207–343, October 1991.

[24] W. Kuich. The Kleene and Parikh theorem in complete semirings. In T. Ottmann, editor, *Proc. 14th Colloq. Automata, Languages, and Programming*, volume 267 of *Lect. Notes in Comput. Sci.*, pages 212–225, New York, 1987. EATCS, Springer-Verlag.

[25] W. Kuich and A. Salomaa. *Semirings, Automata, and Languages.* Springer-Verlag, 1986.

[26] E. W. Mayr and A. Meyer. The complexity of the word problems for commutative semigroups and polynomial ideals. *Adv. Math.*, 46(3):305–329, December 1982.

[27] K. C. Ng. *Relation Algebras with Transitive Closure.* PhD thesis, University of California, Berkeley, 1984.

[28] V. Pratt. Dynamic algebras as a well-behaved fragment of relation algebras. In D. Pigozzi, editor, *Proc. Conf. on Algebra and Computer Science*, volume 425 of *Lect. Notes in Comput. Sci.*, pages 77–110. Springer-Verlag, June 1988.

[29] V. N. Redko. On defining relations for the algebra of regular events. *Ukrain. Mat. Z.*, 16:120–126, 1964. In Russian.

[30] J. Sakarovitch. Kleene's Theorem revisited: a formal path from Kleene to Chomsky. In A. Kelemenova and J. Keleman, editors, *Trends, Techniques, and Problems in Theoretical Computer Science*, volume 281 of *Lect. Notes in Comput. Sci.*, pages 39–50. Springer-Verlag, 1987.

[31] A. Salomaa. Two complete axiom systems for the algebra of regular events. *J. Assoc. Comput. Mach.*, 13(1):158–169, January 1966.

[32] A. Tarski. On the calculus of relations. *J. Symb. Logic*, 6(3):65–106, 1941.