The theory of languages

Paul Brunet

University College London

paul@brunet-zamansky.fr

Abstract

We investigate the equational theory of algebras of formal languages with the constants empty language and unit language, the unary mirror image and Kleene star operators, and the binary operations of union, intersection and concatenation. We start by reducing the problem of deciding the validity of an equation over this signature to the equality of certain graph languages. This allows us to derive decidability, and to show that this problem is in fact ExpSpace-complete. Using this graph characterisation, and removing the Kleene star from our signature, we then propose a complete finite axiomatisation of this theory. This development was obtained using the proof assistant Coq.

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 ($\cap$), and concatenation ($\cdot$), and the unary operations of Kleene star ($\_^\star$) and mirror image, also called converse, ($\_^\smile$). 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 $e = f$ (respectively inequation $e \leq f$) is *valid* if the corresponding equality (resp. containment) holds universally. A *free representation* is a set $\mathcal{M}$ together with a map h from terms to elements of $\mathcal{M}$ such that $e = f$ is valid if and only if h maps e and f to the same element of $\mathcal{M}$.

It is well known that to any term over this syntax, one can associate a regular language, and that comparing regular languages is decidable. In fact, the problem of comparing regular expressions with intersection with respect to regular language equivalence is ExpSpace-complete [7]. The difference with the work presented here is that we are considering equations that are stable under substitution. Formally, that means that we do not interpret the letter a as the singleton language $\{a\}$, but rather as a universally quantified variable ranging over all languages. What is remarkable however is that testing the validity of equations in reversible Kleene lattices turns out to also be an ExpSpace-complete problem.

Fragments of this algebra have been studied:

Kleene algebra (KA, [6]). If we restrict ourselves to the operators of regular expressions (0, 1, $+$, $\cdot$, and $\_^\star$), then the free representation is the set of regular languages, with the usual definition of the language of an expression. Testing the validity of equations in KA is thus a PSpace-complete problem [8].

Kleene algebra with converse (KAC, [2]). If we add to KA the converse operation, then the free representation consists of regular languages over a duplicated alphabet, with a letter a' denoting the converse of the letter a. The associated decision problem is still in PSpace.

Identity-free Kleene lattices (KL$^-$, [1]). This algebra stems from the operators 0, $+$, $\cdot$, $\cap$ and $\_^+$, where the latter is the non-zero iteration. Andréka Mikulás and Németi studied this fragment, and showed that the free representation of this algebra consists of languages of series-parallel graphs, downward closed with respect to some graph preorder. We reformulated their results with Pous [5], and introduced a new class of automata, called Petri automata, able to recognise these languages of graphs. We provided a decision procedure to compare these automata, thus yielding an ExpSpace decision procedure for this theory. We prove ExpSpace-hardness by adapting a proof from [7].

The present work is then an extension of identity-free Kleene lattices, by adding unit and mirror image. The addition of mirror image is fairly simple, relying mainly on ideas from [2]. However, the seemingly small addition of 1 yields some complications. In fact, in [1, 5] there is a free representation of Kleene allegories, an algebra over the same signature as reversible Kleene lattices, but whose intended model is binary relations rather than languages. In that context, adding 1 means moving from *series-parallel* graphs to graphs of tree-width 2, that might have cycles. This is a significant problem for automata based decision procedures.

In the context of languages, adding 1 yields other problems. However, the free representation we get for reversible Kleene lattices remains more tractable than that of Kleene allegories. In particular we do not create cycles in series parallel graphs, but rather have to collect additional information. Let us illustrate this with the following inequation:

$$c \cdot (1 \cap a) \leq a \cdot c. \tag{1}$$

On the left hand side (LHS), the term $1 \cap a$ appears. This term is either equal to 1 if the empty word belongs to language a, or 0 otherwise. In the first case, the LHS is equal to c and we have $1 \leq a$, meaning that $c = 1 \cdot c \leq a \cdot c$. In the second case the LHS is equal to 0, which is contained in $a \cdot c$ as well. The key observation here is that the second case does not really matter: in a term build out of concatenations, intersections, converse, variables and units, if 0 appears somewhere then the term will always evaluate to 0 and thus be contained in any other term. The free representations we develop for union-free terms consist of pairs of a representation of a 1-free term and a set of language variables that are assumed to contain the empty word. This allows us to make the reasoning we used to study (1) automatic.

Following an approach similar to [5], we first construct the free representation of reversible Kleene lattices, and introducing a new Petri net-based automata model we show that testing the validity of equations is a decidable problem, and in fact an ExpSpace-complete one [3, 4]. Using the fact that without the Kleene star these free representations are finite, we build on ideas from [1] to find an prove correct a complete finite axiomatisation of the theory of languages over the signature $\langle 0, 1, +, \cdot, \cap, \_^{\smile} \rangle$, using the proof assistant Coq.

References

[1] H. Andréka, S. Mikulás, and I. Németi. The equational theory of Kleene lattices. *TCS*, 412(52):7099–7108, 2011. doi:10.1016/j.tcs.2011.09.024.

[2] S. L. Bloom, Z. Ésik, and G. Stefanescu. Notes on equational theories of relations. *Alg. Univ.*, 33(1):98–126, 1995. doi:10.1007/BF01190768.

[3] P. Brunet. Reversible Kleene lattices. In *MFCS*, 2017. to appear.

[4] P. Brunet. Reversible Kleene lattices. working paper, 2017. URL: https://hal.archives-ouvertes.fr/hal-01474911.

[5] P. Brunet and D. Pous. Petri Automata for Kleene Allegories. In *Proc. LICS*, pages 68–79, July 2015. doi:10.1109/LICS.2015.17.

[6] J. H. Conway. *Regular algebra and finite machines*. Chapman and Hall Mathematics Series, 1971.

[7] M. Fürer. The complexity of the inequivalence problem for regular expressions with intersection. In *Proc. ICALP*, pages 234–245, 1980. doi:10.1007/3-540-10003-2_74.

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