Introduction to Kleene Algebra	**Lecture 26**
CS786 Spring 2002	**Tuesday, April 30, 2002**

Complexity of Reasoning under Assumptions

In this lecture and the next we study the complexity of reasoning in Kleene algebra and star-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 star-continuous Kleene algebra, (i) if E contains only commutativity assumptions $pq = qp$, the problem is Π^0_1-complete; (ii) if E contains only monoid equations, the problem is Π^0_2-complete; and (iii) for arbitrary equations E, the problem is Π^1_1-complete. The last problem is the universal Horn theory of the star-continuous Kleene algebras. These results are from [8].

Reasoning with Assumptions

The equational theory of KA alone is *PSPACE*-complete, and this is as efficient as one could expect. As we have argued, 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. In the presence of tests, the commutativity condition $pb = bp$ models the fact that the execution of the program p does not affect the value of the test b. Such assumptions are needed to reason about basic program transformations such as constant propagation and moving static computations out of loops.

As shown in Lecture **??**, assumptions of the form $pb = bp$ where b is a test do not increase the complexity of KAT. 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. This result can also be shown to follow from a 1979 result of Berstel [1] with a little extra work; we will give this argument this below.

These considerations bring up the general question:

How hard is it to reason in Kleene algebra under equational assumptions?

Equivalently and more formally,

What is the complexity of deciding the validity of universal Horn formulas of the form $E \to s = t$, 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 star-continuity (KA^*). Recall that a Kleene algebra is *star-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 star-continuous, but all known naturally occurring ones are. Moreover, although star-continuity often provides a convenient shortcut in equational proofs, there are no more equations provable with it than without it, as we have shown.

Because of these considerations, it has become common practice to adopt star-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.

The known results about the complexity of the Horn theories of KA and KA^* are summarized in Table 1.

Table 1: Main Results

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

The results D and H apply to Kleene algebras with tests and were proved in Lecture **??**. The decision problems in the column labeled KA are all r.e. because of the finitary axiomatization of KA. 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 [3, Theorem 4.3, p. 98]). The *EXPSPACE*-hardness of C follows from the *EXPSPACE*-hardness of the word problem for commutative monoids [9]. It is not known whether C is decidable.

Perhaps the most remarkable of these results is E. This is the general question of the complexity of the universal Horn theory of the star-continuous Kleene algebras. This question is related to a conjecture of Conway [2, p. 103], who asked for an axiomatization of the universal Horn theory of the regular sets. The phrasing of Conway's conjecture is somewhat ambiguous, and a literal interpretation is relatively easy to refute [6].

That the universal Horn theory of KA^* should be so highly complex is quite surprising in light of the relative simplicity of the axiomatization. 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.

This lecture and the next assume a basic knowledge of complexity of abstract data types and recursion theoretic hierarchies. Good introductory references on these topics are [10] and [5], respectively.

Regular Sets over a Monoid

A cornerstone of our approach is the universality of $\mathsf{Reg}\,M$, the regular subsets of an arbitrary monoid M, discussed earlier in Lecture **??**. Recall that any monoid homomorphism $h : M \to K$ from a monoid M to the multiplicative monoid of a star-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 star-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. Intuitively, we can think in terms of regular sets of equivalence classes of words modulo E.

More formally, let M be a monoid with identity 1_M. Recall that the powerset 2^M forms a natural star-continuous Kleene algebra under the operations

$$\begin{array}{rclcrcl} A+B & \stackrel{\mathrm{def}}{=} & A \cup B & \qquad & 0 & \stackrel{\mathrm{def}}{=} & \varnothing \\ AB & \stackrel{\mathrm{def}}{=} & \{xy \mid x \in A,\ y \in B\} & & 1 & \stackrel{\mathrm{def}}{=} & \{1_M\} \\ A^* & \stackrel{\mathrm{def}}{=} & \bigcup_{n \ge 0} A^n & & & & \end{array}$$

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

Let $\mathsf{Reg}\,M$ denote the smallest Kleene subalgebra of 2^M containing the image of M under the map ρ_M. This is a star-continuous Kleene algebra and is called the *algebra of regular sets over M*.

The map $M \mapsto \mathsf{Reg}\,M$, 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

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

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

$$\begin{tikzcd} M \arrow[r, "h"] \arrow[d, "\rho_M"'] & M' \arrow[d, "\rho_{M'}"] \\ \mathsf{Reg}\ M \arrow[r, "\mathsf{Reg}\ h"] & \mathsf{Reg}\ M' \end{tikzcd} \tag{1}$$

The functor Reg is the left adjoint of the forgetful functor that takes a star-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 star-continuous Kleene algebra K extends uniquely through ρ_M to a Kleene algebra homomorphism $\widehat{h} : \mathsf{Reg} M \to K$:

$$\begin{tikzcd} M \arrow[r, "h"] \arrow[d, "\rho_M"'] & K \\ \mathsf{Reg}\ M \arrow[ur, "\widehat{h}"'] & \end{tikzcd} \tag{2}$$

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

$$\widehat{h}(A) \ \stackrel{\text{def}}{=}\ \sup\ \{h(x) \mid x \in A\}. \tag{3}$$

This makes sense for star-continuous Kleene algebras because of Theorem **??** of Lecture **??**, which says that *suprema of all definable subsets of a star-continuous Kleene algebra exist.* It does not work for Kleene algebras in general, since the supremum on the right-hand side of (3) may not exist.

Monoid Equations

Now we indicate how to take advantage of the universality property (2) to obtain the results F and G in Table 1.

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 star-continuous Kleene algebra $\mathsf{Reg}\ \Sigma^*/E$, called the *standard interpretation.*

Lemma 0.1 *The following are equivalent:*

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

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

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 star-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 (2) 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 will allow us to restrict our attention to $\mathsf{Reg}\,\Sigma^*/E$ for the purpose of proving F and G in Table 1.

Encoding Turing Machines

The lower bound proofs for E, F, and G in Table 1 depend partially on encoding Turing machine computations as monoid equations. We follow the treatment of Davis [3].

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. The alphabet Γ may contain other symbols as well.
- It has a finite set of states Q disjoint from Γ containing a start state s and one or more halt states distinct from s.
- 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 nonblank symbols.

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

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

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 $d \in$ {left, 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 0.2 *If $x, y \in \Gamma$ and t is a halt state, then*

$$\vdash xsy \dashv \; \approx \; \vdash ztw \dashv \quad \Leftrightarrow \quad \vdash xsy \dashv \; \xrightarrow[M]{*} \; \vdash ztw \dashv \, . \tag{4}$$

Proof. See [3, Theorem 4.3, p. 98]. The chief concern is that monoid equations are reversible, whereas computations are not; thus it is conceivable that the left-hand side of (4) holds by some complicated sequence of substitutions modeling a zigzagging forwards-and-backwards computation even when the right-hand side of (4) does not. It can be shown that since M is deterministic and there are no transitions out of state t, this cannot happen. $\square$

Main Results

Theorem 0.3 *The following complexity results hold for the problem of deciding whether a given Horn formula $E \to s = t$ is true in all star-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 0.1 and expressing an equation as the conjunction of two inequalities, we can reduce the problem to the conjunction of two instances of

$$\mathsf{Reg}\,\Sigma^*/E,\ \iota\ \vDash\ s \le t. \tag{5}$$

The upper bounds for both (i) and (ii) are obtained by expressing (5) as a first-order formula with the appropriate quantifier prefix. Let $\equiv$ denote congruence modulo E on Σ^*. Applying (1) with $M = \Sigma^*$ and $M' = \Sigma^*/E$, (5) can be expressed

$$\forall x\ x \in \rho_{\Sigma^*}(s) \ \to\ \exists y\ y \equiv x \wedge y \in \rho_{\Sigma^*}(t). \tag{6}$$

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 (6) is a 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 (6) is equivalent to a Π^0_1 formula.

The lower bound for (i) uses the characterization of Lemma 0.1 and the result of Berstel [1] (see also [4, 7]) that (5) is undecidable. The reductions given in the cited references show that (5) 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 above 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

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

This is a well-known Π^0_2-complete problem. By Lemma 0.2, this is true iff

$$\mathsf{Reg}\,\Delta^*/E,\ \iota\ \vDash\ \vdash s a^n \dashv\ =\ \vdash t \dashv, \quad n \ge 0,$$

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

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

since $\{x\} \subseteq \{y\}$ iff $x = y$. By star-continuity, this is true iff

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

and by Lemma 0.1, this is true iff

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

□

References

[1] Jean Berstel. *Transductions and Context-free Languages.* Teubner, Stuttgart, Germany, 1979.

[2] John Horton Conway. *Regular Algebra and Finite Machines.* Chapman and Hall, London, 1971.

[3] Martin Davis. *Computability and Unsolvability.* McGraw-Hill, New York, 1958.

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

[5] Peter G. Hinman. *Recursion Theoretic Hierarchies.* Springer-Verlag, New York, 1977.

[6] Dexter Kozen. On Kleene algebras and closed semirings. In Rovan, editor, *Proc. Math. Found. Comput. Sci.*, volume 452 of *Lecture Notes in Computer Science*, pages 26–47, Banska-Bystrica, Slovakia, 1990. Springer-Verlag.

[7] Dexter 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 *Lecture Notes in Computer Science*, pages 14–33, Passau, Germany, March 1996. Springer-Verlag.

[8] Dexter Kozen. On the complexity of reasoning in Kleene algebra. In *Proc. 12th Symp. Logic in Comput. Sci.*, pages 195–202, Los Alamitos, Ca., June 1997. IEEE.

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

[10] J. Meseguer and J. A. Goguen. Initiality, induction and computability. In M. Nivat and J. C. Reynolds, editors, *Algebraic Methods in Semantics*, pages 460–541. Cambridge University Press, 1985.