Solutions of Word Equations over Partially Commutative Structures

Volker Diekert[1]

Universität Stuttgart

ICALP 2016
Rome, July 15th, 2016

In memoriam Zoltán Ésik (1951 – 2016)

[1] Joint work with Artur Jeż (Wroclaw) and Manfred Kufleitner (Stuttgart)

Introduction to $\textsc{WordEquations}$

$$abX = Xba \iff X \in (ab)^*a.$$

$\textsc{WordEquations}$ over a monoid M: Given a pair (U, V) of strings over elements of M and variables. Is there a substitution of variables by elements in M such that $U = V$ in M?

- $M = \Sigma^*$ free monoid: $aX = bY$ no solution.
- $M = F(\Sigma)$ free group: $aX = bY$ infinitely many solutions.
- $M = M(\Sigma, I)$ free partially commutative monoid.
  $baX\bar{b}Y = aYX$ no solution due to length constraints.
- $M = G(\Sigma, I)$ free partially commutative group, $\bar{b} = b^{-1}$.
  $baX\bar{b}Y = aYX$ infinitely many solutions if $ab = ba$.

- 1900 HILBERT10. Given a polynomial $p(X_1, \ldots, X_k)$ with coefficients in $\mathbb{Z}$, is there an interger solution?
- 1960's WORDEQUATIONS special instance of HILBERT10
- 1970 Matiyasevich: HILBERT10 is undecidable based on previous work by Davis, Putnam, and Robinso
- 1977 Makanin: WORDEQUATIONS is decidable for Σ^*
- 1982/84 Makanin/Razborov: Existential and positive theories of free groups are decidable
- 1998–2006 Tarski's conjectures:
  Kharlampovich and Myasnikov: The theory of free groups is decidable.
  Kharlampovich/Myasnikov and Sela: The theories for free nonabelian groups are equivalent.

Complexity of Makanin's algorithms

- WordEquations. Complexity (first published estimation):

$$\mathsf{DTIME}\left(2^{2^{2^{2^{2^{\mathrm{poly}(n)}}}}}\right)$$

- Makanin's algorithm for solving equations in free groups is not primitive recursive. (Kościelski/Pacholski 1990)
- 1999 Plandowski: WordEquations is in PSPACE.
- 2000 Gutiérrez: WordEquations for free groups is in PSPACE.
- 2001 D., Gutiérrez, Hagenah: WordEquations for free groups with rational constraints is PSPACE-complete.

ICALP 1998. Plandowski and Rytter: Application of Lempel-Ziv Encodings to the Solution of Word Equations.

- **New conjecture:** WordEquations is NP-complete.
- Compression became a main tool in solving equations.

STACS 2013 and J. ACM 2016. Artur Jeż applied recompression to WordEquations and simplified all known proofs for decidability.

Free partially commutative monoids and groups

- Σ denotes a finite alphabet with involution $a \mapsto \overline{a}$ with $\overline{\overline{a}} = a$.
- $\rho : \Sigma \to 2^{\mathfrak{R}}$ where $\mathfrak{R}$ is a set of resources, $\rho(a) = \rho(\overline{a})$.
- $M(\Sigma, \rho) = \Sigma^* / \{\, ab = ba \mid \rho(a) \cap \rho(b) = \emptyset \,\}$ is a trace monoid with involution $\overline{a_1 \cdots a_\ell} = \overline{a_\ell} \cdots \overline{a_1}$.
- $\Sigma^* / \{\, ab = ba \mid \rho(a) \cap \rho(b) = \emptyset \,\} \cup \{\, a\overline{a} = 1 \mid a \in \Sigma \,\}$ is a RAAG $G(\Sigma, \rho)$ where $\overline{g} = g^{-1}$.
- RAAG = right angled Artin group = free partially commutative group = graph group.
- Our results hold more generally for graph products.

Task

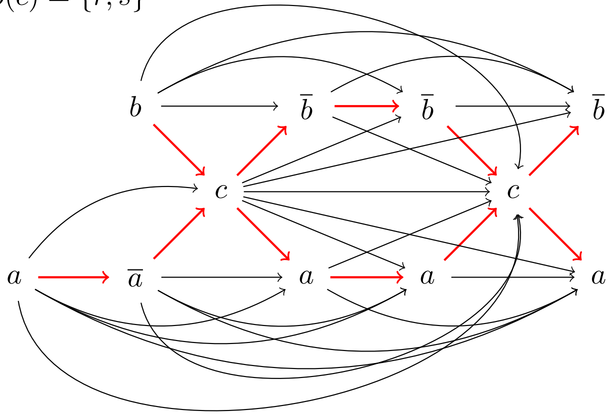
Solve equations over these partially commutative structures.

Traces are directed acyclic node labeled graphs

Example: $A = \{a, \overline{a}, b, \overline{b}, c, \overline{c}\}$ with

$\rho(a) = \rho(\overline{a}) = \{r\}$, $\rho(b) = \rho(\overline{b}) = \{s\}$, and

$\rho(c) = \rho(\overline{c}) = \{r, s\}$



Dependence graph (Hasse diagram in red) of $ab\overline{a}ca\overline{b}abca\overline{b}$

Overview

	free monoids	free groups
∃-theory	Makanin '77	Makanin '82
Pos. theory	undecidable	Makanin & Razborov '84
Theory	undecidable	Kharlampovich & Myasnikov 2004

	trace monoids	RAAGs	graph products
∃-theory	Matiyasevich '97	D. & Muscholl '02	D. & Lohrey '03
Pos. theory	undecidable	D. & Lohrey '03	D. & Lohrey '03[2]
Theory	undecidable	open	undec./open

Casals-Ruiz & Kazachkov 2011 define an analogue of Makanin-Razborov diagrams for RAAGS.

[2] There is a reduction from the graph product to the factors.

All Solutions

Main contribution: high level version

- We describe the set of all solution by an EDT0L language: This is, we construct an NFA where the labels are endomorphisms, the accepted language is a rational set $\mathcal{R}$ of endomorphisms over a free monoid C^*. If $X_1, \ldots, X_k$ denote the variables, then we obtain all solutions by:

$$\{ (h(c_1), \ldots, h(c_k)) \mid h \in \mathcal{R} \}.$$

- New decidability results: Finiteness of solution sets for equations over trace monoids.
- Improved complexity: NSPACE$(n \log n)$.
- Simplified proofs.

Our theorem for trace monoids with involution

Input. A resource alphabet $(A \cup \mathcal{X}, \rho)$ with involution, a trace equation (U, V) in constants A and variables $\mathcal{X} = \{X_1, \ldots, X_k\}$.

Output. An "extended" alphabet C with involution. An NFA $\mathcal{A}$ of singly exponential size accepting a rational set $\mathcal{R}$ of A-endomorphisms on C^* such that under the canonical projection $\pi : A^* \to M(A, \rho)$ we obtain:

$$
\begin{aligned}
&\{(\pi h(c_1), \ldots, \pi h(c_k)) \mid h \in \mathcal{R}\} \\
&= \{(\sigma(X_1), \ldots, \sigma(X_k)) \mid \sigma \text{ solves } U = V \text{ in } M(A, \rho)\}.
\end{aligned}
$$

Furthermore, (U, V) has a solution if and only if $\mathcal{A}$ accepts a nonempty set; (U, V) has infinitely many solutions if and only if $\mathcal{A}$ has a directed cycle. These conditions can be tested in $\mathsf{NSPACE}(n \log n)$ where $n = |UV|$.

[Group version] The same, but solutions σ satisfy $\sigma(U) = \sigma(V)$ in the RAAG $G(A, \rho)$ and for a variable X the solution $\sigma(X)$ is restricted to be a reduced trace (= no factors aa^{-1}).

All solutions of a trace equation as an EDT0L language

1. Construct the NFA $\mathcal{A}$ using simple rules.
2. The overall strategy is an induction: remove first letters that use the least set of resources. Repeat.
3. States are equations over certain quotients of resource monoids: these intermediate structures use partial commutation beyond trace monoids.
4. Transitions are labeled by endomorphisms.
5. Prove soundness.
6. Prove completeness using (a modified) Jeż compression.

The NFA

States (for equations without constraints) are tuples $(W, B, \mathcal{X}, \rho, \theta)$

$W = (U, V)$	equation
B	constants with $A \subseteq B \subseteq C$
$\mathcal{X}$	variables in W
$\rho : B \cup \mathcal{X} \to 2^{\mathfrak{R}}$	resources
θ	additional commutation rules

Transitions change these parameters.

A B-solution is given by $\sigma : \mathcal{X} \to M(B, \rho, \theta)$ such that $\sigma(U) = \sigma(V)$. A solution is given by a pair (σ, α) where $\alpha : M(B, \rho, \theta) \to M(A, \rho_A, \emptyset)$ transforms the B-solution to a solution over the original trace monoid.

ε-transitions: substitutions of variables

There is no change in constants: the label is the identity id_{C^*}.

1. $\tau(X) = 1$, remove X from the equation. Potentially removes partial commutation.
2. $\tau(X) = aX$, where a is a constant.
3. $\tau(X) = YaX$ where Y is a fresh variable such that $\rho(Y) \subsetneq \rho(X)$. Prevent that such a splitting occurs for X more than a constant number of times. (This is crucial.)
4. Define types $\theta(x) = u$ to express that $xu = ux$. Here, x is a variable or a word of length at most 2 and u is a word of length at most 2.
5. There are symmetric rules for the right side.

Compression: transitions that modify constants

Choose fresh letters c and $\overline{c}$.

1. Rename a as c. The label is the morphism defined by

$$h(c) = a \text{ and } h(\overline{c}) = \overline{a}.$$

2. Compress some word u into a single letter c. This includes compressions $ab \to a$, $ab \to b$, $aa \to a$. The label is the morphism defined by

$$h(c) = u \text{ and } h(\overline{c}) = \overline{u}.$$

If we wish to compress a factor ab into a fresh letter c, then we must uncross the factor first. Consider

$$\cdots bXaXu\overline{X}\overline{a}\cdots$$

with $\sigma(X) = vbw$ where $\rho(a) \cap \rho(v) = \emptyset$ and $\rho(b) = \rho(a) \cup \rho(v)$.

Then we split X by $\tau(X) = YbX$ where Y is a new variable with $\rho(Y) = S$. The new solution is $\sigma(Y) = v$ and $\sigma(X) = w$.

We obtain

$$\cdots bYbXaYbXu\overline{X}\,\overline{b}\,\overline{Y}\overline{a}\cdots = \cdots bYbXYabXu\overline{X}\,\overline{b}\overline{a}\overline{Y}\cdots$$

and compression yields

$$\cdots bYbXYcXu\overline{X}\,\overline{c}\overline{Y}\cdots$$

Conclusion

- The algorithm is greedy: it tries nondeterministically everything within a given space bound.
- The "tricky part" is to prove completeness: every solution can be recovered by some path in the NFA $\mathcal{A}$ if "the extended alphabet C is large enough."
- **Open problem.** Can we construct an NFA for endomorphisms over some free group $F(C)$ if there are elements of order 2? The answer is "yes" for free products.
- **Challenge.** Prove NP-completeness for $\textsc{WordEquations}$.

Thank you