
Abstract of PhD Thesis

Author: Artur Jeż
Title: Conjunctive Grammars and Equations
over Sets of Natural Numbers
Language: English
Supervisor: Krzysztof Loryś and Alexander Okhotin
Institute: University of Wrocław, Institute of Computer Science, Poland
Date: 14 September 2010

Abstract

This thesis is devoted to the study of systems of equations

$$\varphi_i(X_1, \dots, X_n) = \psi_i(X_1, \dots, X_n) \quad \text{for } i = 1, \dots, m,$$

in which the variables X_1, \dots, X_n represent sets of natural numbers. The allowed operations are union, intersection and addition, which is defined as

$$X + Y = \{x + y \mid x \in X, y \in Y\}.$$

Such systems can be equally interpreted as systems of language equations over a single-letter alphabet and operations of union, intersection and concatenation.

The study begins with considering the subclass of systems of equations over sets of numbers, consisting of systems of the *resolved form*, i.e.,

$$X_i = \varphi_i(X_1, \dots, X_n) \quad \text{for } i = 1, \dots, n.$$

The counterparts of such systems among the language equations are the resolved systems of language equations over a single-letter alphabet. These can be also viewed as appropriate grammars: when the allowed operations are union and concatenation, they correspond to context free grammars; when also intersection is allowed, to the conjunctive grammars.

It is shown that the resolved systems of equations over sets of natural numbers can have non-ultimately periodic sets as the least solutions. Equivalently, conjunctive grammars over a single-letter alphabet can generate non-regular languages, as opposed to context-free grammars. This claim is firstly demonstrated by giving a simple system, which has the least solution with $\{4^n : n \in \mathbb{N}\}$ as its first component. This system exploits the properties of the base-4 positional notation of sets, in particular, the aforementioned set should be viewed as the set of numbers with base-4 notation 10^ℓ , for some $\ell \geq 0$.

Then, this example is generalised. For every set S of numbers, such that base- k positional notations of numbers from S are recognised by a certain type of a real-time cellular automaton, an explicit construction of a resolved system with S as the first component of the least solution is given.

It is then shown that the systems with no restrictions on the form of the equations imposed are computationally universal: the class of unique solutions of such systems coincides with the class of recursive sets. Similar characterisations are shown for the class of least (greatest) solutions: this class coincides with the class of recursively enumerable sets (co-recursively enumerable sets, respectively). This generalises the known result for systems of language equations over a multiple-letter alphabet. These results hold even when only union and addition (or only intersection and addition) are allowed in the system.

Systems with addition as the only allowed operation are also considered; they correspond to systems of language equations over a single-letter alphabet with concatenation as the only operation. It is shown that the obtained class of sets is computationally universal, in the sense that their unique (least, greatest) solutions can represent encodings of all recursive (recursively enumerable, co-recursively enumerable, respectively) sets.

The computational complexity of decision problems for both formalisms is investigated. It is shown that the membership problem for the resolved systems of equations is EXPTIME-complete. Many other decision problems for both types of systems are proved to be undecidable, and their exact undecidability level is settled. Most of these results hold even when the systems are restricted to the use of one equation with one variable.

Table of Contents

1 Introduction 1

2 Basic terminology and notation 9

3 Sets with regular positional notation 21

 3.1 Toy example 21

 3.2 Regular notation 24

4 Sets with trellis positional notation 45

5 Transforming resolved to unresolved 69

 5.1 Two general translation lemmata 70

 5.2 Sets with regular positional notation 74

5.3 Sets trellis positional notation	79
6 Completeness of systems of equations	97
6.1 Sketch	98
6.2 Unresolved systems with $\{\cup, \cap, +\}$	99
6.3 Equations with $\cup, +$ or $\cap, +$	107
7 Equations with addition only	113
7.1 Overview of the argument	114
7.2 Encoding of sets	114
7.3 Simulating operations	118
7.4 Simulating a system of equations	121
7.5 Systems with finite constants	123
7.6 Afterthought	125
8 Single nonterminal grammars	131
9 Equation with one variable	139
10 Membership for resolved equations	155
11 Conjunctive grammars	175
11.1 Membership	176
11.2 Equivalence	178
11.3 Finiteness	182
11.4 Co-finiteness	184
12 General systems of equations	189

Author's correspondence address Artur Jez
Institute of Computer Science
University of Wrocław
ul. Joliot-Curie 15
PL-50-383 Wrocław
Poland