Artur Jeż | Alexander Okhotin

Univariate equations over sets of natural numbers

Turku Centre *for* Computer Science

TUCS Technical Report
No 913, November 2008

Univariate equations over sets of natural numbers

Artur Jeż
Institute of Computer Science,
University of Wrocław,
50–383 Wroclaw, Poland
aje@ii.uni.wroc.pl
Alexander Okhotin
Academy of Finland, *and*
Department of Mathematics, University of Turku, *and*
Turku Centre for Computer Science
Turku FIN–20014, Finland
alexander.okhotin@utu.fi

TUCS Technical Report
No 913, November 2008

Abstract

It is shown that equations $X = \varphi(X)$, in which the unknown X is a set of natural numbers and φ uses operations of union, intersection and addition $S + T = \{ m + n \mid m \in S, n \in T\}$, can simulate systems of equations $X_i = \varphi_i(X_1, \ldots, X_n)$ with $1 \leqslant i \leqslant n$, in the sense that the solution of a system is encoded in the solution of an equation. This implies undecidability of some properties of one-nonterminal conjunctive grammars over a unary alphabet. In a relatively similar way, equations $\varphi(X) = \psi(X)$ can simulate systems of such equations with multiple variables, which implies computational universality of their least and greatest solutions, as well as undecidability of their basic decision problems. In both constructions it is sufficient to use only singleton constants.

Keywords: Language equations, conjunctive grammars, decision problems

TUCS Laboratory
Discrete Mathematics for Information Technology

1 Introduction

This paper is concerned with systems of equations, in which the unknowns are sets of natural numbers, while the left- and right-hand sides use Boolean operations, as well as element-wise addition of sets defined as $S + T = \{m + n \mid m \in S,\ n \in T\}$. On one hand, such equations can be regarded as a generalization of *integer expressions*, introduced in the seminal paper by Stockmeyer and Meyer [12] and later systematically studied by McKenzie and Wagner [7]. On the other hand, these equations are a particular case of *language equations* defined over a unary (one-letter) alphabet.

Language equations, which have formal languages as unknowns, have recently received much attention [6]. Their most well-known kind are systems of the form

$$\begin{cases} X_1 &= \varphi_1(X_1, \ldots, X_n) \\ & \vdots \\ X_n &= \varphi_n(X_1, \ldots, X_n) \end{cases} \tag{*}$$

in which the right-hand sides φ_i may contain union and concatenation of languages, as well as singleton constants. These equations, first proposed by Ginsburg and Rice [1], provide the most natural semantics for context-free grammars. If intersection is further allowed, then systems (*) represent *conjunctive grammars*, which are a natural extension of the context-free grammars introduced and studied by Okhotin [8, 10].

The expressive power of conjunctive grammars over a unary alphabet has been realised only recently, once Jeż [2] constructed a grammar for the non-regular language $\{a^{4^n} \mid n \geqslant 0\}$. This grammar can be equally regarded as a system of four equations over sets of numbers, using union, intersection and addition, and one of the questions raised by Jeż [2] was how many variables are necessary to obtain any non-periodic solution. This question was answered by Okhotin and Rondogiannis [11], who constructed a single univariate equation $X = \varphi(X)$ with a non-periodic solution, as well as presented a class of sets of numbers that are not representable by any such equations.

The first result of this paper, given in Section 3, generalizes the construction of Okhotin and Rondogiannis [11]. It will be shown that for every unary conjunctive grammar, the languages generated by all of its nonterminal symbols can be encoded together in a single unary language generated by a one-nonterminal conjunctive grammar. This construction implies that some undecidability and complexity results for unary conjunctive grammars due to Jeż and Okhotin [3, 4] hold already for one-variable grammars.

The other contribution of this paper concerns more general systems of equations of the form

$$\begin{cases} \varphi_1(X_1, \ldots, X_m) &= \psi_1(X_1, \ldots, X_m) \\ & \vdots \\ \varphi_\ell(X_1, \ldots, X_m) &= \psi_\ell(X_1, \ldots, X_m) \end{cases} \tag{**}$$

Language equations of this form using concatenation and union were shown to be computationally complete by Okhotin [9], essentially assuming that the alphabet contains at least two letters. This result has recently been remade for equations over sets of numbers (that is, for language equations over a unary alphabet) by Jeż and Okhotin [5].

It is natural to ask, how many variables and equations in (**) are necessary to attain computational universality. In Section 4 it will be shown that every system (**) can be encoded in a single univariate equation $\varphi(X) = \psi(X)$ using ultimately periodic constants. This construction is improved in Section 5 to use singleton constants only. It follows that all known undecidable properties of such systems [5] are possessed already by equations $\varphi(X) = \psi(X)$.

2 Conjunctive grammars and systems of equations

Definition 1 (Okhotin [8])**.** *A conjunctive grammar is a quadruple $G = (\Sigma, N, P, S)$, in which Σ and N are disjoint finite non-empty sets of terminal and nonterminal symbols respectively; P is a finite set of grammar rules, each of the form*

$$A \to \alpha_1 \& \ldots \& \alpha_n \quad (\text{where } A \in N,\ n \geqslant 1 \text{ and } \alpha_1, \ldots, \alpha_n \in (\Sigma \cup N)^*)$$

while $S \in N$ is a nonterminal designated as the start symbol.

The semantics of conjunctive grammars may be defined either by term rewriting [8], or, equivalently, by a system of language equations. This paper uses the latter approach:

Definition 2 ([10])**.** *Let $G = (\Sigma, N, P, S)$ be a conjunctive grammar. The associated system of language equations is the following system in variables N:*

$$A = \bigcup_{A \to \alpha_1 \& \ldots \& \alpha_m \in P} \bigcap_{i=1}^{m} \alpha_i \quad (\text{for all } A \in N)$$

Let $(\ldots, L_A, \ldots)$ be its least solution and denote $L_G(A) := L_A$ for each $A \in N$. Define $L(G) := L_G(S)$.

The existence of a least solution with respect to componentwise inclusion follows from the basic fixpoint theory.

The question of whether conjunctive grammars can generate any non-regular unary languages has been an open problem for some years [10], until recently solved by Jeż [2], who constructed a grammar for the language $\{a^{4^n} \mid n \geqslant 0\}$. Let us reformulate this grammar as the following resolved system of four equations over sets of numbers:

Example 1 (Jeż [2]). *The system*

$$\begin{cases} X_1 &= \big((X_2+X_2)\cap(X_1+X_3)\big) \ \cup \ \{1\} \\ X_2 &= \big((X_6+X_2)\cap(X_1+X_1)\big) \ \cup \ \{2\} \\ X_3 &= \big((X_6+X_6)\cap(X_1+X_2)\big) \ \cup \ \{3\} \\ X_6 &= \big((X_3+X_3)\cap(X_1+X_2)\big) \end{cases}$$

has least solution $X_i = \{i \cdot 4^n \mid n \geqslant 0\}$*, for* $i = 1, 2, 3, 6$.

Sets of this kind can be conveniently specified by regular expressions for the corresponding sets of base-k notations of numbers, which in this case are $\mathtt{10}^*$, $\mathtt{20}^*$, $\mathtt{30}^*$ and $\mathtt{120}^*$, respectively. In the following, some parentheses in the right-hand sides of equations shall be omitted, and the following default precedence of operations shall be assumed: addition has the highest precedence, followed by intersection, and then by union with the least precedence.

The construction in Example 1 essentially uses all four variables, and there seems to be no apparent way to replicate it using a single variable. However, this was achieved in the following example:

Example 2 (Okhotin, Rondogiannis [11]). *The univariate equation*

$$X = \big(11{+}X{+}X \ \cap \ 22{+}X{+}X\big) \cup \big(1{+}X{+}X \ \cap \ 9{+}X{+}X\big) \cup \\ \cup \big(7{+}X{+}X \ \cap \ 12{+}X{+}X\big) \cup \big(13{+}X{+}X \ \cap \ 14{+}X{+}X\big) \cup \{56, 113, 181\}$$

has the unique solution

$$S = \{4^n - 8 | n \geqslant 3\} \cup \{2 \cdot 4^n - 15 | n \geqslant 3\} \cup \{3 \cdot 4^n - 11 | n \geqslant 3\} \cup \{6 \cdot 4^n - 9 | n \geqslant 3\}.$$

This equation is actually derived from Example 1, and its solution encodes the values of all four sets in Example 1. Each of the four components in S represents one of the variables in Example 1 with a certain *offset* (8, 15, 11 and 9).

Note that the set from Example 2 is exponentially growing. It is known that unary conjunctive grammars can generate a set that grows faster than any given recursive set:

Proposition 1 (Jeż, Okhotin [3]). *For every recursively enumerable set of natural numbers* S *there exists a system* $X_i = \varphi_i(X_1, \ldots, X_n)$ *over sets of natural numbers with the least solution* $X_i = S_i$*, such that the growth function of* S_1 *is greater than that of* S *at any point.*

On the contrary, for univariate equations it has been proved that if a set grows faster than exponentially (for example, $\{n! \mid n \geqslant 1\}$), then it is not representable:

Proposition 2 (Okhotin, Rondogiannis [11]). *Let* $S = \{n_1, n_2, \ldots, n_i, \ldots\}$ *with* $0 \leqslant n_1 < n_2 < \ldots < n_i < \ldots$ *be an infinite set of numbers, for which* $\liminf_{i\to\infty} \frac{n_i}{n_{i+1}} = 0$. *Then* S *is not the least solution of any equation* $X = \varphi(X)$.

However, even though one-nonterminal conjunctive grammars cannot generate *all* unary conjunctive languages, it will now be demonstrated that they can represent a certain encoding of any conjunctive language.

3 One-nonterminal conjunctive grammars

The goal is to simulate an arbitrary conjunctive grammar over $\{a\}$ by a conjunctive grammar with a single nonterminal symbol. The construction formalizes and elaborates the intuitive idea of Example 2, making it provably work for any grammar.

The first step towards the construction is a small refinement of the known normal form for unary conjunctive grammars. It is known that every conjunctive language over every alphabet can be generated by a conjunctive grammar in the *binary normal form*, with all rules of the form $A \to B_1C_1 \& \ldots \& B_nC_n$ with $n \geqslant 1$ or $A \to a$. The following stronger form is required by the below construction.

Lemma 1. *For every conjunctive grammar $G = (\Sigma, N, P, S)$ there exists a conjunctive grammar $G' = (\Sigma, N', P', S')$ generating the same language, in which every rule is of the form $A \to a$ with $a \in \Sigma$ or*

$$A \to B_1C_1 \& \ldots \& B_nC_n \quad (\textit{with } n \geqslant 2),$$

in which the sets $\{B_1, C_1\}$, ..., $\{B_n, C_n\}$ are pairwise disjoint.

Proof. If there is a rule with no intersection, that is, $A \to \alpha$ for some nonterminal A and $\alpha \in (N \cup \Sigma)^*$, it can be replaced by a trivial intersection $A \to \alpha \& \alpha$.

Let m be the greatest number of conjuncts in the rules in P. Define m copies of every nonterminal: $N' = N \times \{1, \ldots, m\}$. Replace every rule

$$A \to B_1C_1 \& \ldots \& B_\ell C_\ell$$

with

$$(A, i) \to (B_1, 1)(C_1, 1) \& \ldots \& (B_\ell, \ell)(C_\ell, \ell).$$

For every rule $A \to a$ in the original grammar, define a new rule $(A, i) \to a$. Let $S' = (S, 1)$ be the new start symbol. The resulting grammar generates the same language. □

Theorem 1. *For every unary conjunctive grammar $G = (\{a\}, \{A_1, \ldots, A_m\}, P, A_1)$ of the form given in Lemma 1 there exist numbers $0 < d_1 < \ldots < d_m < p$ depending only on m and an equation*

$$X = \varphi(X)$$

over a set of natural numbers X, with a unique solution $S = \bigcup_{i=1}^{m} S_i$, where $S_i = \{np - d_i \mid a^n \in L_G(A_i)\}$. The size of φ is polynomial in the size of G, and this equation is associated to a certain one-nonterminal conjunctive grammar.

Let $p = 4^{m+2}$ and let $d_i = \frac{p}{4} + 4^i$ for every nonterminal A_i. For every number $t \in \{0, \ldots, p\}$, the set $\{np - t \mid n \geqslant 0\}$ is called *track number t*. The goal of the construction is to represent each set S_i in the track d_i. The rest of the tracks should be empty.

For every rule $A_i \to \alpha$, where $\alpha = A_{j_1}A_{k_1}\& \ldots \& A_{j_\ell}A_{k_\ell}$, consider the following expression over sets of numbers:

$$\varphi_{i,\alpha}(X) = \bigcap_{t=1}^{\ell} X + X + (d_{j_t} + d_{k_t} - d_i).$$

Define the following equation:

$$X = \bigcup_{A_i \to \alpha \in P} \varphi_{i,\alpha}(X) \cup \bigcup_{A_i \to a \in P} \{p - d_i\}$$

Now the task is to prove that the unique solution of this equation is $S = \bigcup_i S_i$, where $S_i = \{np - d_i \mid a^n \in L_G(A_i)\}$.

Each time X appears in the right-hand side of the equation, it is used in the context of an expression $\varphi_{i,\alpha}(X)$. The proof of the theorem is based upon the following property of these expressions.

Lemma 2. *Let $i, j, k, \ell \in \{1, \ldots, m\}$ with $\{i, j\} \cap \{k, \ell\} = \varnothing$. Then*

$$(S + S + d_i + d_j) \cap (S + S + d_k + d_\ell) = (S_i + S_j + d_i + d_j) \cap (S_k + S_\ell + d_k + d_\ell)\,.$$

Proof. As addition is distributive over union and union is distributive over intersection,

$$\begin{aligned}&(S + S + d_i + d_j) \cap (S + S + d_k + d_\ell) = \\ &\quad = \bigcup_{i',j'} (S_{i'} + S_{j'} + d_i + d_j) \cap \bigcup_{k',j'} (S_{k'} + S_{\ell'} + d_k + d_\ell) = \\ &\qquad = \bigcup_{i',j',k',j'} (S_{i'} + S_{j'} + d_i + d_j) \cap (S_{k'} + S_{\ell'} + d_k + d_\ell)\end{aligned}$$

It is sufficient to prove that if $\{i', j'\} \neq \{i, j\}$ or $\{k', \ell'\} \neq \{k, \ell\}$, then the intersection is empty. Consider any such intersection

$$\begin{aligned}&(S_{i'} + S_{j'} + d_i + d_j) \cap (S_{k'} + S_{\ell'} + d_k + d_\ell) = \\ &\quad (\{np \mid a^n \in L(A_{i'})\} - d_{i'} + \{np \mid a^n \in L(A_{j'})\} - d_{j'} + d_i + d_j) \cap \\ &\qquad (\{np \mid a^n \in L(A_{k'})\} - d_{k'} + \{np \mid a^n \in L(A_{\ell'})\} - d_{\ell'} + d_k + d_\ell)\,,\end{aligned}$$

and suppose it contains any number, which must consequently be equal to $d_i + d_j - d_{i'} - d_{j'}$ modulo p and to $d_k + d_\ell - d_{k'} - d_{\ell'}$ modulo p. As each d_t satisfies $\frac{p}{2} > d_t > \frac{p}{4}$, both offsets are between $-\frac{p}{2}$ and $\frac{p}{2}$, and therefore they must be equal to each other:

$$d_i + d_j - d_{i'} - d_{j'} = d_k + d_\ell - d_{k'} - d_{\ell'}.$$

Equivalently, $d_i + d_j + d_{k'} + d_{\ell'} = d_k + d_\ell + d_{i'} + d_{j'}$, and since each d_t is defined as $\frac{p}{4} + 4^t$, this holds if and only if

$$4^i + 4^j + 4^{k'} + 4^{\ell'} = 4^k + 4^\ell + 4^{i'} + 4^{j'}.$$

Consider the largest of these eight numbers, let its value be d. Without loss of generality, assume that it is on the left-hand side. Then the left-hand side is greater than d. On the other hand, if no number on the right-hand side is d, then the sum is at most $4 \cdot \frac{d}{4} = d$. Thus at least one number on the right-hand side must be equal to d as well. Removing those two numbers and giving the same argument for the sum of 3, 2 and 1 summands yields that

$$\{d_i, d_j, d_{k'}, d_{\ell'}\} = \{d_k, d_\ell, d_{i'}, d_{j'}\}.$$

Then, by the assumption that $\{i, j\} \cap \{k, \ell\} = \varnothing$,

$$\{d_i, d_j\} = \{d_{i'}, d_{j'}\} \quad \text{and} \quad \{d_{k'}, d_{\ell'}\} = \{d_k, d_\ell\},$$

and since the addition is commutative,

$$i = i', \quad j = j', \quad k = k' \quad \text{and} \quad \ell = \ell'.$$

Therefore,

$$\begin{aligned}(S + S + d_i + d_j) \cap (S + S + d_k + d_\ell) = \\ \bigcup_{i',j',k',j'} (S_{i'} + S_{j'} + d_i + d_j) \cap (S_{k'} + S_{\ell'} + d_k + d_\ell) = \\ (S_i + S_j + d_i + d_j) \cap (S_k + S_\ell + d_k + d_\ell),\end{aligned}$$

which completes the proof. □

Proof of Theorem 1. Let $P = P_1 \cup P_0$, where P_0 contains rules of the form $A_i \to a$, while P_1 consists of multiple-conjunct rules. The equation is strict and thus has a unique solution in the set of positive natural numbers, so it is enough to show that S is a solution, that is,

$$S = \bigcup_{A_i \to \alpha \in P_1} \varphi_{i,\alpha}(S) \cup \bigcup_{A_i \to a \in P_0} \{p - d_i\}$$

Consider each rule $A_i \to \alpha \in P_1$ with $\alpha = A_{j_1}A_{k_1}\& \ldots \& A_{j_t}A_{k_t}$. Then

$$\varphi_{i,\alpha}(S) = \bigcap_{t=1}^{\ell}(d_{j_t} + d_{k_t} - d_i) + S + S = \bigcap_{t=1}^{\ell}(d_{j_t} + d_{k_t} - d_i) + S_{j_t} + S_{k_t}$$

by Lemma 2, and it is easy to calculate that

$$\bigcap_{t=1}^{\ell}(d_{j_t} + d_{k_t} - d_i) + S_{j_t} + S_{k_t} = \{np - d_i \mid a^n \in L(\alpha)\}.$$

Calculating further,

$$\begin{aligned}
&\bigcap_{t=1}^{\ell}(d_{j_t} + d_{k_t} - d_i) + S_{j_t} + S_{k_t} = \\
&\bigcap_{t=1}^{\ell}(d_{j_t} + d_{k_t} - d_i) + \{pn_j - d_{j_t} \mid a^{n_j} \in L(A_{j_t})\} + \{pn_k - d_{k_t} \mid a^{n_k} \in L(A_{k_t})\} = \\
&\bigcap_{t=1}^{\ell}\{p(n_j + n_k) - d_i \mid a^{n_j} \in L(A_{j_t}), a^{n_k} \in L(A_{k_t})\} = \\
&\bigcap_{t=1}^{\ell}\{np - d_i \mid a^n \in L(A_{j_t}) \cdot L(A_{k_t})\} = \\
&\{np - d_i \mid a^n \in L(\alpha)\}.
\end{aligned}$$

Similarly for $A_i \to a \in P_0$,

$$\{p - d_i\} = \{np - d_i \mid a^n \in L(\{a\})\}.$$

Altogether,

$$\begin{aligned}
&\bigcup_{A_i \to \alpha \in P_1} \varphi_{i,\alpha}(S) \cup \bigcup_{A_i \to a \in P_0} \{p - d_i\} = \\
&\bigcup_i \Big(\bigcup_{A_i \to \alpha \in P_1} \varphi_{i,\alpha}(S) \cup \bigcup_{A_i \to a \in P_0} \{p - d_i\} \Big) = \\
&\bigcup_i \Big(\bigcup_{A_i \to \alpha \in P_1} \{np - d_i \mid a^n \in L(\alpha)\} \cup \{np - d_i \mid a^n \in L(a)\} \Big) = \\
&\bigcup_i \bigcup_{A_i \to \beta \in P} \{np - d_i \mid a^n \in L(\beta)\}.
\end{aligned}$$

Since $(\ldots, L(A_i), \ldots)$ is the solution of the associated system of language equations, $L(A_i) = \bigcup_{A_i \to \beta \in P} L(\beta)$, and hence the latter expression equals

$$\bigcup_i \{np - d_i \mid a^n \in L(A_i)\} = \bigcup_i S_i = S,$$

which completes the proof. □

Corollary 1. *For every unary conjunctive language $L \subseteq a^+$ there exist numbers $p \geqslant d \geqslant 1$ and a conjunctive grammar $G = (\{a\}, \{S\}, P, S)$, such that $L(G) \cap (a^p)^* a^{p-d} = \{a^{np-d} \mid a^n \in L\}$.*

The first implication of this result concerns the complexity of unary conjunctive grammars, as well as the complexity of the general membership problem. Assume that a number n is always described by its binary notation, that is, it uses space $\log n$. The following is known:

Proposition 3 (Jeż, Okhotin [4])**.** *There exists a EXPTIME-complete set of numbers $S \subseteq \mathbb{N}$, such that the language $L = \{a^n \mid n \in S\}$ of unary notations of numbers from S is generated by a conjunctive grammar.*

The problem stated as "Given a unary conjunctive grammar G and a number n in binary, determine whether $a^n \in L(G)$" is EXPTIME-complete.

To show that both results still hold for one-nonterminal unary conjunctive grammars, it is sufficient to take the grammar generating L and to transform it according to Theorem 1.

Theorem 2. *There exists a EXPTIME-complete set of numbers $S \subseteq \mathbb{N}$, such that the language $L = \{a^n | n \in S\}$ is generated by a one-nonterminal conjunctive grammar. The general membership problem for one-nonterminal unary conjunctive grammars with input encoded in binary is EXPTIME-complete.*

Proof. This problem clearly belongs to this complexity class, as it is decidable in exponential time in more general case of systems of such equations [4, Th. 4.1].

Hardness follows from Theorem 1 as follows. It is known [4, Th. 4.1] that there exists a system of equations $X_i = \varphi_i(X_1, \ldots X_k)$ with a least solution $(S_1, \ldots S_k)$, in which S_1 is an EXPTIME-hard set.

By Theorem 1 one can efficiently construct an equation $Y = \psi(Y)$ with a least solution S and numbers $p, d \geqslant 1$ such that $n \in S_1$ if and only if $pn - d \in S$. Hence the general membership problem for a system of equations polynomially reduces to the general membership problem for a single equation. □

Let us now consider the decidability of basic properties of one-nonterminal unary conjunctive grammars. In the case of multiple nonterminals, most basic problems are undecidable:

Proposition 4 (Jeż, Okhotin [3])**.** *For every fixed unary conjunctive language $L_0 \subseteq a^*$, the problem of whether a given conjunctive grammar over $\{a\}$ generates the language L_0 is co-RE-complete.*

However, for one-nonterminal grammars such a problem is decidable at least for $L_0 = \varnothing$ and $L_0 = a^*$. Decision procedures are given by the following necessary and sufficient conditions of emptiness and fullness.

Theorem 3. *Let $G = (\{a\}, \{S\}, P, S)$ be a one-nonterminal conjunctive grammar over a unary alphabet. Then:*

1. *$L(G) \neq \varnothing$ if and only if $S \to a^n \in P$ for some $n \geqslant 0$;*
2. *$L(G) = a^*$ if and only if there is a rule $S \to a^{\ell_1}S^{k_1}\& \ldots \& a^{\ell_m}S^{k_m}$ with $k_i \geqslant 1$ and $k_i + \ell_i \geqslant 2$ for all i, such that $a^{\leqslant \max_i \ell_i, 1} \subseteq L(G)$.*

Proof. The case of emptiness is clear; the criterion actually holds for one-nonterminal conjunctive grammars over any alphabet.

Let us check the characterization of grammars generating a^*.

$\Leftarrow$ Let the grammar contain such a rule. Then $a^n \in L(G)$ is proved by induction on n. The basis, $n \leqslant \ell_i$ and $n \leqslant 1$, is given. For the induction step, consider $n \geqslant 2$ with $n > \ell_i$ for all i. For every i-th conjunct of the selected rule, consider two cases. If $\ell_i = 0$, then $k_i \geqslant 2$ and $0, 1, n-1 \in L(G)$ by the induction hypothesis, so $a^n \in L_G(S^{k_i})$. If $\ell_i \geqslant 1$ and $k_i \geqslant 1$, then $0, n - \ell_i \in L(G)$ by the induction hypothesis and hence $a^n \in L_G(a^{\ell_i}S^{k_i})$. Therefore, the rule generates a^n.

$\Rightarrow$ Suppose the condition does not hold. If there is any rule of the given form, then $L(G)$ does not contain some string in $a^{\leqslant \max_i \ell_i, 1}$, and hence $L(G) \neq a^*$.

Consider the case when there are no such rules, that is, that every rule is of the form $S \to a^{\ell_1}S^{k_1}\& \ldots \& a^{\ell_m}S^{k_m}$ with $k_i = 0$ or $k_i + \ell_i \leqslant 1$ for some i. In other words, every rule is of the form $S \to a^{\ell}\& \ldots$ or of the form $S \to S\& \ldots$. All rules of the latter form may be eliminated without changing the language generated by the grammar, and each of the rest of the rules generates at most one string. Therefore, the grammar generates a finite language and hence $L(G) \neq a^*$. ☐

The same method can be elaborated to characterize equality to any given finite or co-finite language. By this characterization, both problems are clearly decidable. However, the more general problem of equivalence of two grammars is undecidable.

Theorem 4. *The equivalence problem for one-nonterminal unary conjunctive grammars is undecidable.*

Proof. The proof is by reduction from the equivalence problem for unary conjunctive grammars with multiple nonterminals. Two grammars are combined into one, the construction of Theorem 1 is applied, and then the start symbols of the two grammars are exchanged and the construction is applied again. The two resulting one-nonterminal grammars are equivalent if and only if the original grammars generate the same language.

Before approaching the equivalence problem for one-nonterminal conjunctive grammars, let us establish the undecidability of the following technical problem:

Claim 1. *The problem of testing whether for a given conjunctive grammar $G = (\{a\}, N, P, S)$ with two designated nonterminals S and S', $L_G(S) = L_G(S')$, is undecidable.*

It is known that the problem of whether two unary conjunctive grammars generate the same language is undecidable. Let $G_1 = (\{a\}, P_1, N_1, S_1)$ and $G_2 = (\{a\}, P_2, N_2, S_2)$ be any two conjunctive grammars over $\{a\}$. Assume, without loss of generality, that $N_1 \cap N_2 = \varnothing$. Construct a new conjunctive grammar $G = (\{a\}, P_1 \cup P_2, N_1 \cup N_2, S_1)$. Then $L_G(S_1) = L(G_1)$ and $L_G(S_2) = L(G_2)$, and therefore testing the equality of $L_G(S_1)$ and $L_G(S_2)$ solves the equivalence problem for G_1 and G_2.

Now this technical problem may be easily reduced to the equivalence problem for one-nonterminal conjunctive grammars over $\{a\}$. Let a grammar $G = (\{a\}, \{A_1, A_2, \ldots, A_m\}, P, A_1)$ be given, and assume without loss of generality that it is of the form required in Lemma 1; it is asked whether $L_G(A_1) = L_G(A_2)$. Construct a one-nonterminal unary conjunctive grammar G' that encodes G according to Theorem 1, with

$$L(G') = \{a^{np-d_1} \mid a^n \in L_G(A_1)\} \cup \{a^{np-d_2} \mid a^n \in L_G(A_2)\} \cup \\ \cup \bigcup_{i \geqslant 3} \{a^{np-d_i} \mid a^n \in L_G(A_i)\}.$$

Next, the same transformation is applied to the grammar $G = (\{a\}, \{A_2, A_1, A_3, \ldots, A_m\}, P, A_2)$, with nonterminals A_1 and A_2 exchanged. The values of $p, d_1, \ldots, d_m$ are the same, as they depend only on m, so the generated language is

$$L(G'') = \{a^{np-d_2} \mid a^n \in L_G(A_1)\} \cup \{a^{np-d_1} \mid a^n \in L_G(A_2)\} \cup \\ \cup \bigcup_{i \geqslant 3} \{a^{np-d_i} \mid a^n \in L_G(A_i)\}.$$

Clearly, the two languages are the same if and only if $L_G(A_1) = L_G(A_2)$. $\square$

4 Equations $\varphi(X) = \psi(X)$ with ultimately periodic constants

Now consider *unresolved equations* over sets of numbers in which both the left- and right-hand side may contain any expressions. It has recently been established that such equations are computationally complete:

Theorem 5 (Jeż, Okhotin [5]). *The family of sets of natural numbers representable by unique (least, greatest) solutions of systems of equations of the form $\varphi_i(X_1, \ldots, X_m) = \psi_i(X_1, \ldots, X_m)$ with union, intersection and addition, is exactly the family of recursive sets (r.e. sets, co-r.e. sets, respectively).*

The goal of this section is to replicate this result using a unique equation with a unique variable. This is achieved by taking an arbitrary system of equations and encoding it in the way similar to Theorem 1:

Theorem 6. *For every system of equations over sets of numbers $E_j(Y_1, \ldots, Y_m) = F_j(Y_1, \ldots, Y_m)$ with $j \in \{1, \ldots, \ell\}$, in which every expression E_j and F_j is of the form*

$$Y_i \cap Y_{i'} \quad \text{or} \quad Y_i \cup Y_{i'} \quad \text{or} \quad Y_i + Y_{i'} \quad \text{or} \quad \{1\},$$

there exist numbers $0 \leqslant d_1 < \ldots < d_m < p$ and an equation

$$\varphi(X) = \psi(X)$$

using singleton constants and the constant $\{kp \mid k \geqslant 0\}$, such that a set S is its solution if and only if $S = \bigcup_{i=1}^{\ell} \{kp - d_i \mid k \in S_i\}$ for some solution $(S_1, \ldots, S_m)$ of the original system.

The numbers d_i are offsets of tracks for X, and the statement of the lemma already specifies that S is split into tracks like in the proof of Theorem 1. That is, each set S_i is represented in a track $S \cap \{kp - d_i \mid k \geqslant 1\}$. For each variable Y_i, a unique offset d_i is assigned. Define $T = \{0, \ldots, p-1\} \setminus \{d_1, \ldots, d_m\}$; each t-th track with $t \in T$ should be empty.

The set $\varphi(S) = \psi(S)$ is as well split into tracks of its own, which are not directly related to the tracks of S. The tracks of $\varphi(S) = \psi(S)$ correspond to equations of the original system. A track $\{kp - e_j \mid k \geqslant 1\}$ with a unique offset e_j is assigned to an equation number j.

Let $p = 2(m + \ell + 3)$. Then d_i and e_j can be defined so that

- $1 < d_i < \frac{p}{2} - 1$ and $1 < e_j < \frac{p}{2} - 1$ for all i and j;
- variable offsets are greater than equation offsets, that is, $d_i > e_j$ for all i, j.

Define the following expression used to extract the track t from the set X:

$$f_t(X) = \begin{cases} X \cap \{kp \mid k \geqslant 0\}, & \text{if} \quad t = 0 \\ X \cap \big(\{kp \mid k \geqslant 0\} + (p - t)\big), & \text{if} \quad 1 \leqslant t \leqslant p - 1 \end{cases}$$

Provided that $X \subseteq \mathbb{N}$, this definition is equivalent to $f_t(X) = X \cap \{kp - t | k \geqslant 0\}$.

Define the encoding of the system into a single equation. First the left- and right-hand sides of each equation are to be translated. For every expression E as in the statement, define φ_E as follows:

$$\begin{aligned} \varphi_{j,\{1\}}(X) &= \{p - e_j\} \\ \varphi_{j,X_i + X_{i'}}(X) &= f_{d_i}(X) + f_{d_{i'}}(X) + (d_i + d_{i'} - e_j) \\ \varphi_{j,X_i \cup X_{i'}}(X) &= \big(f_{d_i}(X) + (d_i - e_j)\big) \cup \big(f_{d_{i'}}(X) + (d_{i'} - e_j)\big) \\ \varphi_{j,X_i \cap X_{i'}}(X) &= \big(f_{d_i}(X) + (d_i - e_j)\big) \cap \big(f_{d_{i'}}(X) + (d_{i'} - e_j)\big) \end{aligned}$$

Next, these translated expressions are used to define a single equation:

$$\bigcup_{j=1}^{\ell} \varphi_{j,E_j}(X) = \bigcup_{j=1}^{\ell} \varphi_{j,F_j}(X) \cup \bigcup_{\substack{t \in \{0,\ldots,p-1\} \\ t \notin \{d_1,\ldots,d_m\}}} \big(f_t(X) + t + 1\big) \tag{1}$$

Its left-hand side and the first big union on its right-hand side encode the equations of the original system, while the second big union on the right ensures that there is no “garbage” on any other track of X.

Before proceeding with the main proof, let us state some technical properties of this construction. The constructed expressions φ_{j,E_j} simulate the original expressions E_j as follows:

Lemma 3. *Let $S \subseteq \mathbb{N}$ and $S_1, \ldots, S_m \subseteq \mathbb{N}$ satisfy*

$$S = \bigcup_{i=1}^{m} \{kp - d_i \mid k \in S_i\} \tag{2}$$

Let E_j be an expression. Then $\varphi_{j,E_j}(S) = \{kp - e_j \mid k \in E_j(S_1, \ldots, S_m)\}$. The same correspondence holds for F_j.

Proof. Since the defnitions of E_j and F_j are similar, it is enough to consider E_j. Suppose that $E_j(X_1, \ldots, X_m) = X_i \cup X_{i'}$. The value of $\varphi_{j,E_j}(S)$ is calculated as follows:

$$\begin{gathered}\varphi_{j,X_i \cup X_{i'}}(S) = [f_{d_i}(S) + d_i - e_j] \cup [f_{d_{i'}}(S) + d_{i'} - e_j] = \\ \big[(S \cap \{kp - d_i \mid k \geqslant 0\}) + d_i - e_j\big] \cup \big[(S \cap \{kp - d_{i'} \mid k \geqslant 0\}) + d_{i'} - e_j\big] = \\ \{kp - d_i + d_i - e_j \mid k \in S_i\} \cup \{kp - d_{i'} + d_{i'} - e_j \mid k \in S_{i'}\} = \\ \{kp - e_j \mid k \in S_i\} \cup \{kp - e_j \mid k \in S_{i'}\} = \{kp - e_j \mid k \in S_i \cup S_{i'}\}.\end{gathered}$$

Similar calculations yield

$$\begin{aligned}\varphi_{j,\{1\}}(S) &= \{p - e_j\}, \\ \varphi_{j,X_i \cap X_{i'}}(S) &= \{kp - e_j \mid k \in S_i \cap S_{i'}\}, \\ \varphi_{j,X_i + X_{i'}}(S) &= \{kp - e_j \mid k \in S_i + S_{i'}\}.\end{aligned}$$

□

It follows that the equations formed from the new expressions simulate the original equations:

Lemma 4. *Let $S \subseteq \mathbb{N}$ and $S_1, \ldots, S_m \subseteq \mathbb{N}$ satisfy (2). Then, for every j, $E_j(S_1, \ldots, S_m) = F_j(S_1, \ldots, S_m)$ if and only if $\varphi_{j,E_j}(S) = \varphi_{j,F_j}(S)$.*

Proof. By Lemma 3,

$$\varphi_{j,E_j}(S) = \{pk - e_j \mid k \in E_j(S_1, \dots, S_m)\}$$

and

$$\varphi_{j,F_j}(S) = \{pk - e_j \mid k \in F_j(S_1, \dots, S_m)\}.$$

Thus an equality

$$E_j(S_1, \dots, S_m) = F_j(S_1, \dots, S_m)$$

holds if and only if

$$\varphi_{j,E_j}(S) = \varphi_{j,F_j}(S).$$

□

Then the original system of equations is represented by the following system of $m+1$ equations:

Lemma 5. *Let a system of equations*

$$E_j(X_1, \dots, X_m) = F_j(X_1, \dots, X_m) \quad \text{for } j = 1, \dots, \ell$$

be as in Theorem 6, denote $T = \{0, \dots, p-1\} \setminus \{d_1, \dots, d_m\}$, and let $C_t \subseteq \{kp - t \mid k \geqslant 0\}$, for all $t \in T$, be any constant sets. Then a system of equations

$$X \cap \{kp - t \mid k \geqslant 0\} = C_t \qquad \text{for } t \in T \tag{3}$$

$$\varphi_{j,E_j}(X) = \varphi_{j,F_j}(X) \qquad \text{for } j = 1, \dots, \ell \tag{4}$$

has a solution S if and only if

$$S = \bigcup_{t \in T} C_t \cup \bigcup_{i=1}^{m} \{kp - d_i \mid k \in S_i\} \tag{5}$$

for some solution $(S_1, \dots, S_m)$ of the original system.

Proof. Let $(S_1, \dots, S_m)$ be a solution of the original system, construct the set S as in (5). Then S clearly satisfies (3), and also it meets the assumptions of Lemma 4. Then, since every j-th equation $E_j = F_j$ is satisfied by $(S_1, \dots, S_m)$, it follows by Lemma 4 that S satisfies the j-th equation (4).

Suppose now that S satisfies (3)–(4). Then, for every $t \in T$,

$$S \cap \{kp - t \mid k \geqslant 0\} = C_t,$$

by (3). Let $S_i = \{k \mid kp - d_i \in S\}$. Then S is obtained from S_i as in (5). It remains to show that $(S_1, \dots, S_m)$ is a solution of the first system, that is,

$$E_j(S_1, \dots, S_m) = F_j(S_1, \dots, S_m).$$

for each j. This follows from Lemma 4, as by (4)

$$\varphi_{j,E_j}(X) = \varphi_{j,F_j}(X) \quad \text{for } j = 1, \dots, \ell$$

and by Lemma 4 this implies

$$E_j(S_1, \dots, S_m) = F_j(S_1, \dots, S_m) \quad \text{for } j \in \{1, \dots, \ell\}.$$

□

Note that the actual equation (1) mixes the statements (3,4) in a single equality. Showing that they are indeed equivalent yields the proof of the theorem.

Proof of Theorem 6. Assume that S is a solution of (1) and consider intersections of both sides of (1) with $\{kp + 1 \mid k \geqslant 0\}$. Since by Lemma 3 $\varphi_{j,E_j}(S) \subseteq \{kp - j \mid k \geqslant 0\}$, there is $\varnothing$ on the left-hand side, and for the same reason the first big union in the right-hand side vanishes. Thus the equation turns into

$$\varnothing = \varnothing \cup \bigcup_{t \in T} \big(S \cap \{kp - t \mid k \geqslant 0\}\big) + t + 1,$$

and therefore

$$S \cap \{kp - t \mid k \geqslant 0\} = \varnothing \quad \text{for } t \in T. \tag{6}$$

Consider the intersection of (1) with the set $\{kp - e_j \mid k \geqslant 0\}$. As by Lemma 3, $\varphi_{j',E}(S) \subseteq \{kp - e_{j'} \mid k \geqslant 1\}$ then on the left-hand side only $\varphi_{j,E_j}(S)$ remains and on the right-hand side only $\varphi_{j,F_j}(S)$, as for each $t \in T$: $f_t(S) + t + 1 \subseteq \{kp + 1 \mid k \geqslant 0\}$. Thus we obtain a system of equations

$$\varphi_{j,E_j}(S) = \varphi_{j,F_j}(S) \quad \text{for } j = 1, \dots, \ell \tag{7}$$

Hence every solution S also satisfies this system (6)–(7).

Conversely, consider any S satisfying both (6) and (7). Then S clearly satisfies the original equation, as it is obtained as a union of sides of (6) (with additional $+t+1$ at both sides of the equation for t) and (7).

As (6) and (7) satisfy the assumptions of Lemma 5 with constants

$$C_t = \varnothing \quad \text{for } t \in T,$$

every solution of the original equation is of the form

$$S = \bigcup_{i=1}^{m} \{kp - d_i \mid k \in S_i\}$$

for some solution $S_1, \dots, S_m$ of the system

$$E_j(X_1, \dots, X_m) = F_j(X_1, \dots, X_m) \quad \text{for } j = 1, \dots, \ell,$$

which completes the proof of the theorem. □

5 Equations $\varphi(X) = \psi(X)$ with singleton constants

The construction from the previous section will now be refined by eliminating infinite periodic constants from the equation: that is, only singleton constants will be used.

Theorem 7. *For every system of equations $E_j(Y_1, \dots, Y_m) = F_j(Y_1, \dots, Y_m)$ with $j \in \{1, \dots, \ell\}$, in which every expression E_j and F_j is as in Theorem 6, there exist numbers $0 \leqslant d_1 < \dots < d_m < p$ and an equation $\lambda(X) = \rho(X)$ using singleton constants, such that a set $S \subseteq \mathbb{N}$ is its solution if and only if*

$$S = \{kp, kp + \tfrac{p}{2} + 1 \mid k \geqslant 0\} \cup \bigcup_{i=1}^{m} \{kp - d_i \mid k \in S_i\}$$

for some solution $(S_1, \dots, S_m)$ of the original system.

The definitions and the assumptions on d_i and e_j are as in Theorem 6. Similarly to previous section, denote the set of offsets of unused tracks by

$$T = \{0, \dots, p-1\} \setminus \{0, d_1, \dots, d_m, \tfrac{p}{2} - 1\}.$$

The equation (1) defined there actually uses only one infinite constant, $\{kp | k \geqslant 0\}$. In the new construction it is possible to extract infinite constants from any solution S using the following expressions:

$$\pi(X) = X \cap (X + \tfrac{p}{2} - 1)$$
$$\pi'(X) = X \cap (X + \tfrac{p}{2} + 1)$$

Indeed, from the intended form of solutions stated in the theorem, $\pi(S) = \{kp \mid k \geqslant 1\}$ and $\pi'(S) = \{kp - (\frac{p}{2} - 1) \mid k \geqslant 1\}$.

Using these subexpressions, the expressions $f_t(X)$ from the previous construction will be replaced by the following:

$$f'_t(X) = \begin{cases} X \cap (\pi'(X) + \frac{p}{2} - 1 - t) & \text{for } 0 \leqslant t \leqslant \frac{p}{2} - 1 \\ \Big(X \cap (\pi(X) + p - t)\Big) \cup \Big(X \cap \{p - t\}\Big) & \text{for } \frac{p}{2} \leqslant t \leqslant p - 1 \end{cases}$$

The goal is to construct such an equation that $f'_i(S) = f_i(S)$ for each of its solutions S, which will allow reusing parts of the construction and the proof from Theorem 6. In particular, the expressions $\varphi'_{j,E}$ are defined in the same way as $\varphi_{j,E}$ in Theorem 6, this time using $f'_i(X)$ instead of $f_i(X)$. Furthermore, define the following three new expressions:

$$\psi(X) = \{\tfrac{p}{2} + 1\} \cup (\pi'(X) + \tfrac{p}{2} - 1) \cup (\pi(X) + \tfrac{p}{2} + 1)$$
$$\psi'(X) = \pi(X) \cup \pi'(X)$$
$$\theta(X) = \bigcup_{j=1}^{\ell} \big(f'_{e_j}(X) + e_j - 1\big) \cup \bigcup_{t \in T \setminus \{e_1, \dots, e_\ell\}} f'_t(X)$$

The expressions $\psi(X)$ and $\psi'(X)$ are used to generate the sets $\{kp \mid k \geqslant 1\} \cup \{kp - (\frac{p}{2} - 1) \mid k \geqslant 1\}$. The expression $\theta(X)$ deals with the "garbage" in the same way as the second part of the right-hand side of (1).

Now the new equation is constructed the follows:

$$\psi(X) \cup \bigcup_{j=1}^{\ell} \big(\varphi'_{j,E_j}(X) + p\big) = \psi'(X) \cup \bigcup_{j=1}^{\ell} \big(\varphi'_{j,F_j}(X) + p\big) \cup \theta(X) \tag{8}$$

The main technical property of this equation is that in each of its solutions the tracks 0 and $\frac{p}{2}-1$ are full, while all tracks besides these two and $d_1, \ldots, d_m$ are empty.

Lemma 6. *If S is a solution of* (8)*, then*

$$\begin{aligned} \{kp \mid k \geqslant 0\} &\subseteq S, \\ \{kp - (\tfrac{p}{2} - 1) \mid k \geqslant 1\} &\subseteq S, \\ S \cap \{kp - t \mid k \geqslant 1\} &= \varnothing \quad (\textit{for all } t \in T) \end{aligned}$$

Proof. Let us adopt the following terminology for any deviations from this rule. For any set of numbers S_0, a number $n \in S_0$ is said to be *extra* if $n = kp - t$ for $t \in T$. A number n is *missing* if $n \notin S_0$ and $n = kp - t$ for $t \in \{0, \frac{p}{2} - 1\}$. Then it has to be proved that there cannot be any extra or missing numbers in any solution of the equation (6).

The proof begins with the following technical claim:

Claim 1. *Let $S_0 \subseteq \mathbb{N}$ be any set that has no extra numbers. Then, for every expression $E \in \{E_j, F_j\}$ in every j-th equation, it holds that $\varphi'_{j,E}(S_0) \subseteq \{kp - e_j \mid k \geqslant 1\}$.*

Proof. The main step towards establishing the claim is showing that, under the assumptions, $\pi'(S_0) \subseteq \{kp - (\frac{p}{2} - 1) \mid k \geqslant 1\}$.

Consider any $n \in \pi'(S_0) = S_0 \cap (S_0 + (\frac{p}{2} + 1))$. Then $n \in S_0$ and $n' = n - \frac{p}{2} - 1 \in S_0$. Let $n = kp - t$, then, as there are no extra numbers in S_0, t and $(t + \frac{p}{2} + 1 \mod p)$ must be in $\{0, d_1, \ldots, d_m, \frac{p}{2} - 1\}$, which is only possible if $t = \frac{p}{2} - 1$. Hence $n = kp - (\frac{p}{2} - 1)$ and $n' = (k-1)p$, which shows that

$$\pi'(S_0) = S_0 \cap (S_0 + (\tfrac{p}{2} + 1)) \subseteq \{kp - (\tfrac{p}{2} - 1) \mid k \geqslant 0\}.$$

Then the definition of f'_{d_i} can be expanded as

$$\begin{aligned} f'_{d_i}(S_0) = S_0 \cap (\pi'(S_0) + \tfrac{p}{2} - 1 - d_i) \subseteq \pi'(S_0) + \tfrac{p}{2} - 1 - d_i \subseteq \\ \subseteq \{kp - (\tfrac{p}{2} - 1) \mid k \geqslant 1\} + \tfrac{p}{2} - 1 - d_i = \{kp - d_i \mid k \geqslant 1\}, \end{aligned}$$

and therefore, for an expression $E = X_i \cap X_{i'}$,

$$\begin{aligned} \varphi'_{j,E}(S_0) = f'_{d_i}(S_0) + d_i - e_j \cap f'_{d_{i'}}(S_0) + d_{i'} - e_j \subseteq \\ f'_{d_i}(S_0) + d_i - e_j = \{kp - d_i \mid k \geqslant 1\} + d_i - e_j = \{kp - e_j \mid k \geqslant 1\}. \end{aligned}$$

Similar calculations can be made for $E = X_i \cup X_{i'}$, $E = X_i + X_{i'}$ and $E = \{1\}$. $\square$

Let S be any solution of the equation. The first claim is that there are no extra or missing numbers in S that are smaller than p.

Claim 2. *It holds that* $0, \frac{p}{2}+1 \in S$ *and* $p - t \notin S$ *for all* $t \in T$.

Proof. Consider a number $n < p$ appearing on the left-hand side of (8) under the substitution $X = S$. Then $n \in \psi(S)$, as the rest of the left-hand side cannot produce any number less than p. As $\psi(S)$ is a union of three subexpressions, consider each of them. First suppose that $n \in \pi(S) + \frac{p}{2} + 1$, that is, $n - \frac{p}{2} - 1 \in \pi(S)$. Then, by the definition of π, $n - \frac{p}{2} - 1 \in S + \frac{p}{2} - 1$, and so $n \geqslant p$. Next, suppose that $n \in \pi'(S) + \frac{p}{2} - 1$. Then $n - \frac{p}{2} + 1 \in \pi'(S)$ and thus, by the definition of π', $n - \frac{p}{2} + 1 \in S + \frac{p}{2} + 1$, and again $n \geqslant p$. The only remaining possibility is $n = \frac{p}{2} + 1$.

Therefore, the only number smaller than p that appears on the right-hand side is $\frac{p}{2} + 1$. Based on this fact, it will be shown which small numbers must belong to S in order to obtain $\frac{p}{2} + 1$ on the right-hand side, and which may not belong to S, as they would produce other small numbers on the right-hand side.

First, note that the only number $n' \leqslant \frac{p}{2}$ that may be in S is 0, as otherwise n' would be in $S \cap (p - t)$ for some $\frac{p}{2} \leqslant t \leqslant p - 1$, which is a part of $\theta(S)$, and clearly n' does not occur on the left-hand side.

Let us now consider how the number $\frac{p}{2} + 1$ is obtained on the right-hand side. Every number in $\varphi'_{j,F_j}(S) + p$ for any j is at least p and hence is of no concern. Consider now $\theta(S)$. Suppose that $\frac{p}{2} + 1 \in \theta(S)$. Then it belongs to one of the subexpressions in $\theta(S)$. Suppose that $\frac{p}{2} + 1 \in f'_{e_j}(S) + e_j - 1$ for some j. Let $n' \in f'_{e_j}(S)$ be such that $\frac{p}{2}+1 = n'+(e_j-1)$. Then $n' \geqslant p-e_j$ by the definition of $f'_{e_j}(S)$, and hence $\frac{p}{2}-1 = n'+(e_j-1) \geqslant p-1$, a contradition.

Suppose $\frac{p}{2}+1 \in f'_t(S)$ for $t \in T \setminus \{e_1, \ldots, e_\ell\}$. According to the definition of $f'_t(S)$, there are two cases. If $0 \leqslant t \leqslant \frac{p}{2} - 1$, then $\frac{p}{2} + 1 \in \pi'(X) + \frac{p}{2} - 1 - t$ and hence $t + 2 \in \pi'(X)$, which implies that $t - \frac{p}{2} + 1 \in S$. This is only possible if $t = \frac{p}{2} - 1$, which is not an appropriate value of t. In the second case of $\frac{p}{2} \leqslant t \leqslant p - 1$, $\frac{p}{2} + 1 \in f'_t(S)$ means that $\frac{p}{2} + 1 \in (\pi(S) \cup \{0\}) + p - t$ and hence $t \in (\pi(S) \cup \{0\}) + \frac{p}{2} - 1$. As t is not $\frac{p}{2} - 1$, $t - \frac{p}{2} + 1$ must be in $\pi(S)$ and hence in S, which is impossible because $1 \leqslant t - \frac{p}{2} + 1 \leqslant \frac{p}{2}$.

The number $\frac{p}{2}+1$ also cannot be in $S \cap (p-t)$ for any $t \in T \setminus \{e_1, \ldots, e_\ell\}$, because $t = \frac{p}{2} - 1$ is excluded from the set.

The only remaining possibility is $\frac{p}{2} + 1 \in \psi'(S)$, that is, $\frac{p}{2} + 1 \in \pi(S) \cup \pi'(S)$. If $\frac{p}{2}+1 \in \pi(S) = S \cap (S+\frac{p}{2}-1)$, then $2 \in S$, which is not the case, as it is smaller than $\frac{p}{2}+1$. Thus $\frac{p}{2}+1 \in \pi'(S) = S \cap (S + \frac{p}{2} + 1)$, and therefore $0, \frac{p}{2} + 1 \in S$, as desired.

The only thing left to show is that for each $t \in T$, the number $p-t$ is not in S. If $t \geqslant \frac{p}{2}$, this has already been proved above. Suppose that $p - t \in S$

for any $t \in T$ with $t < \frac{p}{2}$ and $t \notin \{e_1, \ldots, e_\ell\}$. Then $p - t \in \pi'(S) + \frac{p}{2} - 1 - t$ because $\frac{p}{2} + 1 \in \pi'(S)$, and hence $p - t \in f'_t(S) \subseteq \theta(S)$ is on the right-hand side. However, there is no corresponding number on the left-hand side, which is a contradiction.

Finally, suppose $p - e_j \in S$ for some j. Then, as $0, \frac{p}{2}+1 \in S$, $\frac{p}{2}+1 \in \pi'(S)$ and accordingly

$$p - e_j = (\tfrac{p}{2} + 1) + (\tfrac{p}{2} - 1 - e_j) \in \pi'(S) + \tfrac{p}{2} - 1 - e_j = f'_{e_j}(S).$$

Therefore, $p - e_j + e_j - 1 = p - 1 \in \theta(S)$, and thus, again, $p - 1$ must appear on the right-hand side, which is a contradiction. □

In order to show that there are no missing or extra numbers, suppose there are any such numbers. Then there is the smallest among them. It will now be proved that for every such number there is a smaller missing or extra number, that is, there cannot be the smallest among them.

Claim 3. *Let $n \notin S$ be the least missing number. Then there exists a number $n' < n$ that is extra.*

Proof. Let n be the least missing number. If it is of the form $n = kp$, then $n \geqslant p$, since $0 \in S$ by Claim 2. The numbers $n - p$ and $n - \frac{p}{2} + 1$ are in S, because, by assumption, there are no missing numbers less than n. Therefore, $n - \frac{p}{2} + 1 \in \pi'(S)$ and hence $n = (n - \frac{p}{2} + 1) + \frac{p}{2} - 1 \in \psi(S)$, that is, n belongs to the left-hand side.

A similar analysis applies for $n = kp - \frac{p}{2} + 1$. By Claim 2, $n \geqslant \frac{3p}{2} + 1$, since there are no missing numbers less than p. The numbers $n - p$ and $n - \frac{p}{2} - 1$ must be in S, because they are smaller than the least missing number. Then $n - \frac{p}{2} - 1 \in \pi(S)$ and, accordingly, $n = (n - \frac{p}{2} - 1) + \frac{p}{2} + 1 \in \psi(S)$.

In both cases, since n appears on the left-hand side, it should also appear on the right-hand side. Consider the subexpression in which n is obtained.

First suppose that $n \in \varphi'_{j,F_j}(S) + p$ for some j, and define the finite set $S_0 = S \cap \{n' \mid n' \leqslant n - p\}$. Then $n \in \varphi'_{j,F_j}(S_0) + p$, because the membership of numbers larger than $n - p$ in the argument does not influence the value of this expression. If S_0 contains an extra number n', this establishes the claim, as $n' < n$. So suppose, for the sake of a contradiction, that S_0 contains no extra numbers. Thus $\varphi'_{j,F_j}(S_0) + p \subseteq \{kp - e_j \mid k \geqslant 2\}$ by Claim 1. It follows that $n \in \{kp - e_j \mid k \geqslant 2\}$, which contradicts the form of missing numbers. Hence S_0 contains an extra number.

Suppose that $n \in \psi'(S)$, then $n \in \pi(S) \cup \pi'(S)$, but as $\pi(S) \subseteq S$ and $\pi'(S) \subseteq S$ then $n \in S$ and this is not possible, as n is a missing number. For the same reason n cannot belong to the second part of $\theta(S)$, as $f'_t(S) \subset S$ by the definition of $f'_t(X)$.

Therefore, n must belong to the first part of $\theta(S)$. Then there exists equation number j, such that $n \in f'_{e_j}(S) + e_j - 1$. This implies $n \in \pi'(S) + \frac{p}{2} - 2$, from whence it follows that $n - \frac{p}{2} + 2 \in S$. To see that this is the promised extra number in S, consider two cases of n: if $n = kp$, then $n - \frac{p}{2} + 2$ belongs to track $\frac{p}{2} - 2 \in T$, and if $n = kp - \frac{p}{2} + 1$, then $n - \frac{p}{2} + 2 = kp - p + 3$ is in track $p - 3 \in T$. $\square$

Claim 4. *If $n \in S$ is the least extra number, then there exists a number $n' < n$ that is missing.*

Proof. As it has already been shown that there are no extra numbers smaller than p, and p cannot be an extra number, assume $n > p$.

Let $n = kp - t$ and suppose there are no missing numbers smaller than n. Then it can be inferred that $n \in f'_t(S)$. If $0 < t < \frac{p}{2} - 1$, then $n + t - (\frac{p}{2} - 1), n + t - p \in S$ (as they are smaller than n). Hence $n + t - (\frac{p}{2} - 1) \in \pi'(S)$ and thus $n = \big(n + t - (\frac{p}{2} - 1)\big) + \frac{p}{2} - 1 - t \in f'_t(S)$. The second case is that $\frac{p}{2} \leqslant t < p$. Since there are no missing numbers smaller than n, then $n + t - p, n + t - p - (\frac{p}{2} - 1) \in S$. Thus $n + t - p \in \pi(S)$ and hence $n = n + t - p + (p - t) \in f'_t(S)$. The rest of the proof is split into two cases depending on t.

Let $t \in T \backslash \{e_1, \ldots, e_\ell\}$. Then $n \in f'_t(S) \subseteq \theta(S)$. Therefore, n is present on the left-hand side, and so it should appear on the right-hand side. Consider the expressions on the right-hand side from which n is obtained.

If $n \in \psi(S)$, then, in particular, $n \in S + p$. But this means that $n - p \in S$, which is a contradiction, as n was supposed to be the smallest extra number.

If $n \in f'_{e_j}(S) + p$ for some j. Let $S_0 = S \cap \{n'' \mid n'' \leqslant n - p\}$. Then $n \in f'_{e_j}(S_0) + p$. Since there are no extra numbers in S_0, by Claim 1, $f'_{e_j}(S_0) + p$ contains only numbers on the equation tracks, and thus $n \notin f'_{e_j}(S_0) + p$. Contradiction.

Consider the other case of $n = kp - e_j$ for some j. Since $n \in f'_{e_j}(S)$, the number $n' = n + e_j - 1$ is in $\theta(S)$, hence n' is on the right-hand side. Note, that there is no extra number smaller than $n' - \frac{p}{2}$.

Suppose that $n' \in \psi(S)$. Then, in particular $n' \in S + p$. But this means that $n' - p$ is an extra number, which is a contradiction, there is no extra number smaller than $n' - \frac{p}{2}$. Thus n' is not in $\psi(S)$.

Suppose that $n' \in f'_{e_j}(S) + p$ for some e_j. Let $S_0 = S \cap \{n'' \mid n'' \leqslant n' - p\}$. Then $n' \in f'_{e_j}(S_0) + p$. There are no extra numbers in S_0. Then by Claim 1 $f'_{e_j}(S_0) + p$ contains only numbers on the equation tracks, and thus $n' \notin f'_{e_j}(S_0) + p$. Contradiction. $\square$

Hence, by Claim 2, Claim 3 and Claim 4, there are no missing and extra numbers. Then the first and second inclusions in the statement of Lemma 6 hold, as they state that there are no missing numbers. The third inclusion states that there is no extra numbers, which completes the proof of Lemma 6. $\square$

The above lemma has established the basic structure of any solution S, which must contain all numbers in tracks 0 and $\frac{p}{2} - 1$ and no elements of any tracks besides these two and the tracks $d_1, \ldots, d_m \in \{2, \ldots, \frac{p}{2} - 2\}$. Because of this, S can be shifted and intersected with itself to obtain a certain periodic set. This is what is done in the expressions π and π':

Lemma 7. *If S is a solution of (8), then*

$$\pi(S) = \{kp \mid k \geqslant 1\}$$
$$\pi'(S) = \{kp - (\tfrac{p}{2} - 1) \mid k \geqslant 1\}$$

Proof. The proof is based upon Lemma 6, which assures that if $n = kp - t \in S$, then $t \in \{0, d_1, \ldots, d_m, \frac{p}{2} - 1\}$.

Consider any $n \in \pi(S) = S \cap (S + (\frac{p}{2} - 1))$. Then $n \in S$ and $n - \frac{p}{2} + 1 \in S$. Let $n = kp - t$, then t and $(t + \frac{p}{2} - 1 \mod p) \in \{0, d_1, \ldots, d_m, \frac{p}{2} - 1\}$, which is only possible if $t = 0$. Hence $n = kp$ and $n' = kp - (\frac{p}{2} - 1)$, which shows that $S \cap (S + (\frac{p}{2} - 1)) \subseteq \{kp \mid k \geqslant 0\}$.

Conversely, let $n = kp$ for some $k \geqslant 1$. Then $n, n - (\frac{p}{2} - 1) \in S$ by Lemma 6, and thus $n \in \pi(S)$.

A similar calculation can be done for the second equality. If $n \in \pi'(S) = S \cap (S + (\frac{p}{2} + 1))$, then $n \in S$ and $n - (\frac{p}{2} + 1) \in S$. By the same argument, $n = -(\frac{p}{2} - 1) \pmod p$, and hence it belongs to the given set. In the other direction, if $n = kp - (\frac{p}{2} - 1)$ for some $k \geqslant 1$, then $n \in S$ and $n - (\frac{p}{2} + 1) = (k-1)p \in S$ by Lemma 6, which shows that $n \in \pi'(S)$. □

Now the values of the auxiliary expressions $\psi(X)$, $\psi'(X)$ and $\theta(X)$ can be determined by direct calculations based on the result of Lemma 7:

Lemma 8. *If $\pi(S) = \{kp \mid k \geqslant 1\}$ and $\pi'(S) = \{kp - (\frac{p}{2} - 1) \mid k \geqslant 1\}$, then*

$$\begin{aligned}
\psi(S) = \psi'(S) &= \{kp \mid k \geqslant 0\} \cup \{kp - (\tfrac{p}{2} - 1) \mid k \geqslant 1\}, \\
f'_t(S) &= f_t(S) \quad (\textit{for all } t \in \{0, \ldots, p-1\}), \\
\theta(S) &= \varnothing, \\
\varphi'_{j,E}(S) &= \varphi_{j,E}(S) \quad (1 \leqslant j \leqslant \ell,\ E \in \{E_j, F_j\}).
\end{aligned}$$

Proof. Using Lemma 7, the following direct calculations are carried out:

$$\begin{aligned}
\psi(S) &= \{\tfrac{p}{2} + 1\} \cup (\pi'(S) + \tfrac{p}{2} - 1) \cup (\pi(S) + \tfrac{p}{2} + 1) = \\
&= \{\tfrac{p}{2} + 1\} \cup (\{kp - (\tfrac{p}{2} - 1) \mid k \geqslant 1\} + \tfrac{p}{2} - 1) \cup (\{kp \mid k \geqslant 1\} + \tfrac{p}{2} + 1) = \\
&= \{kp \mid k \geqslant 1\} \cup \{kp + \tfrac{p}{2} + 1 \mid k \geqslant 0\}.
\end{aligned}$$

Similarly,

$$\psi'(S) = \pi(S) \cup \pi'(S) = \{kp - \tfrac{p}{2} - 1 \mid k \geqslant 1\} \cup \{kp \mid k \geqslant 1\}.$$

Consider now $f'_t(S)$. For $0 \leqslant t \leqslant \frac{p}{2} - 1$,

$$\begin{aligned}
f'_t(S) &= S \cap (\pi'(S) + \tfrac{p}{2} - 1 - t) = \\
&= S \cap (\{kp - (\tfrac{p}{2} - 1) \mid k \geqslant 1\} + \tfrac{p}{2} - 1 - t) = \\
&= S \cap (\{kp - t \mid k \geqslant 1\} = f_t(S),
\end{aligned}$$

and for $\frac{p}{2} \leqslant t \leqslant p-1$, similarly,

$$f'_t(S) = S \cap (\pi(S) + p - t) \cup S \cap (\{p-t\}) = \\ = \big(S \cap (\{kp \mid k \geqslant 1\} + p - t)\big) \cup \big(S \cap \{p-t\}\big) = \\ = S \cap \{kp - t \mid k \geqslant 1\} = f_t(S).$$

The value of $\theta(S)$ is calculated as

$$\theta(S) = \bigcup_{j=1}^{\ell} \big(f'_{e_j}(S) + e_j - 1\big) \cup \bigcup_{t \in T \setminus \{e_1,\dots,e_\ell\}} f'_t(S) = \\ \bigcup_{j=1}^{\ell} \big(f_{e_j}(S) + e_j - 1\big) \cup \bigcup_{t \in T \setminus \{e_1,\dots,e_\ell\}} f_t(S) = \\ \bigcup_{j=1}^{\ell} \varnothing \cup \bigcup_{t \in T \setminus \{e_1,\dots,e_\ell\}} \varnothing = \varnothing.$$

As φ' is defined analogously to φ, with $f_t(X)$ replaced by $f'_t(X)$, and as it has already been proved that $f'_t(S) = f_t(S)$, it follows that $\varphi'_{j,E}(S) = \varphi_{j,E}(S)$ for $E \in \{E_j, F_j\}$. □

The proof of Theorem 7 is generally similar to the proof of Theorem 6, though there are more details to consider. Roughly speaking, once Lemma 7 Lemma 8 determine the values of the auxiliary expressions and establish the equality of φ' with the earlier expression φ, Lemma 5 from the previous section becomes applicable, and it yields the equivalence.

Theorem 7. Suppose S satisfies the equation (8). Then, by Lemma 6, it satisfies the following system of equations as well:

$$X \cap \{kp, kp + \tfrac{p}{2} + 1 \mid k \geqslant 0\} = \{kp, kp + \tfrac{p}{2} + 1 \mid k \geqslant 0\} \tag{9}$$
$$X \cap \{kp - t \mid k \geqslant 0\} = \varnothing \quad \text{for } t \in T \tag{10}$$

Let us substitute S into (8) and intersect both of its sides with the set $\{kp - e_j \mid k \geqslant 0\}$. According to Lemma 8,

$$\psi(S) \cap \{kp - e_j \mid k \geqslant 0\} = \psi(S)' \cap \{kp - e_j \mid k \geqslant 0\} = \theta(S) = \varnothing.$$

On the other hand, by Lemma 8 and by Lemma 3,

$$\varphi'_{j,E}(S) = \varphi_{j,E}(S) \subseteq \{kp - e_j \mid k \geqslant 1\}. \tag{11}$$

This gives the following equality:

$$\varphi_{j,E_j}(S) + p = \varphi_{j,F_j}(S) + p \quad (\text{for } j = 1, \dots, \ell). \tag{12}$$

Then (9,10,12) satisfy the assumptions of Lemma 5 with constants

$$C_0 = \{kp \mid k \geqslant 0\}$$
$$C_{\frac{p}{2}-1} = \{kp - (\tfrac{p}{2} - 1) \mid k \geqslant 0\}$$
$$C_t = \varnothing \quad \text{for } t \in T,$$

and the lemma states that S is of the form

$$S = \{kp, kp + \tfrac{p}{2} + 1 \mid k \geqslant 0\} \cup \bigcup_{i=1}^{m} \{kp - d_i \mid k \in S_i\},$$

where $(S_1, \ldots, S_m)$ is a solution of the original system.

Conversely, assume that $(S_1, \ldots, S_m)$ is a solution of the original system, and let

$$S = \{kp, kp + \tfrac{p}{2} + 1 \mid k \geqslant 0\} \cup \bigcup_{i=1}^{m} \{kp - d_i \mid k \in S_i\}.$$

Then S satisfies (9,10,12) by Lemma 5. Under these premises, $\pi(S)$ and $\pi'(S)$ can be calculated in the same way as in Lemma 7, resulting in

$$\pi(S) = \{kp \mid k \geqslant 1\} \quad \text{and} \quad \pi'(S) = \{kp - (\tfrac{p}{2} - 1) \mid k \geqslant 1\}.$$

Then, by Lemma 8,

$$\psi(S) = \psi'(S) \tag{13}$$
$$\theta(S) = \varnothing, \tag{14}$$
$$\varphi'_{j,E}(S) = \varphi_{j,E}(S) \quad (1 \leqslant j \leqslant \ell,\ E \in \{E_j, F_j\}). \tag{15}$$

Now it can be verified that a substitution $X = S$ turns (8) into an equality:

$$\psi(S) \cup \bigcup_{j=1}^{\ell} (\varphi'_{j,E_j}(S)) = \psi(S) \cup \bigcup_{j=1}^{\ell} (\varphi_{j,E_j}(S)) \cup \varnothing =$$
$$= \psi'(S) \cup \bigcup_{j=1}^{\ell} (\varphi_{j,F_j}(S)) \cup \theta(S) = \psi'(S) \cup \bigcup_{j=1}^{\ell} (\varphi'_{j,F_j}(S)) \cup \theta(S)$$

by (11), (13), (14) and (15). Hence S is a solution. □

Now the known constructions of systems of equations reprensenting recursive, r.e. and co-r.e. sets by their unique, least and greatest solutions are immediately extended to univariate equations:

Corollary 2. *For every recursive (r.e., co-r.e.) set* $S_0 \subseteq \mathbb{N}$ *there exist numbers* $0 \leqslant d < p$ *and an equation* $\varphi(X) = \psi(X)$ *using union, intersection, addition and singleton constants, such that its unique (least, greatest) solution* S *satisfies* $S \cap \{kp - d \mid k \geqslant 1\} = \{kp - d \mid k \in S_0\}$.

In particular, there exists an equation $\varphi(X) = \psi(X)$ with an r.e.-complete least solution, as well as one with a co-r.e.-complete greatest solution.

Another implication is that all decision problems about the cardinality of the set of solutions have the same complexity for systems and for univariate equations.

Corollary 3. *The problem of whether an equation $\varphi(X) = \psi(X)$ using union, intersection, addition and singleton constants has solutions (a unique solution, finitely many solutions) is co-r.e.-complete (Π_2-complete, Σ_3-complete, respectively).*

Acknowledgements

Research supported by the Polish Ministry of Science and Higher Education under grant MNiSW N206 024 31/3826, 2006–2008, and by the Academy of Finland under grant 118540.

References

[1] S. Ginsburg, H. G. Rice, "Two families of languages related to ALGOL", *Journal of the ACM*, 9 (1962), 350–371.

[2] A. Jeż, "Conjunctive grammars can generate non-regular unary languages", *DLT 2007* (Turku, Finland, July 3–6, 2007), LNCS 4588, 1242–253.

[3] A. Jeż, A. Okhotin, "Conjunctive grammars over a unary alphabet: undecidability and unbounded growth", *CSR 2007* (Ekaterinburg, Russia, September 3-7 2007), LNCS 4649, 168–181.

[4] A. Jeż, A. Okhotin, "Complexity of equations over sets of natural numbers", *25th Annual Symposium on Theoretical Aspects of Computer Science* (STACS 2008, Bordeaux, France, 21–23 February, 2008), 373–383.

[5] A. Jeż, A. Okhotin, "On the computational completeness of equations over sets of natural numbers" *35th International Colloquium on Automata, Languages and Programming* (ICALP 2008, Reykjavik, Iceland, July 7–11, 2008), 63–74.

[6] M. Kunc, "What do we know about language equations?", *Developments in Language Theory* (DLT 2007, Turku, Finland, July 3–6, 2007), LNCS 4588, 23–27.

[7] P. McKenzie, K. W. Wagner, "The complexity of membership problems for circuits over sets of natural numbers", *Computational Complexity*, 16 (2007), 211–244.

[8] A. Okhotin, “Conjunctive grammars”, *Journal of Automata, Languages and Combinatorics*, 6:4 (2001), 519–535.

[9] A. Okhotin, “Unresolved systems of language equations: expressive power and decision problems”, *Theoretical Computer Science*, 349:3 (2005), 283–308.

[10] A. Okhotin, “Nine open problems for conjunctive and Boolean grammars”, *Bulletin of the EATCS*, 91 (2007), 96–119.

[11] A. Okhotin, P. Rondogiannis, “On the expressive power of univariate equations over sets of natural numbers”, *IFIP Intl. Conf. on Theoretical Computer Science* (TCS 2008, Milan, Italy, 8–10 September, 2008), IFIP vol. 273, 215–227.

[12] L. J. Stockmeyer, A. R. Meyer, “Word problems requiring exponential time”, *STOC 1973*, 1–9.

TURKU

CENTRE *for*

COMPUTER

SCIENCE

Lemminkäisenkatu 14 A, 20520 Turku, Finland | www.tucs.fi

University of Turku
- Department of Information Technology
- Department of Mathematical Sciences

Åbo Akademi University
- Department of Computer Science
- Institute for Advanced Management Systems Research

Turku School of Economics and Business Administration
- Institute of Information Systems Sciences

ISBN 978–952–12–2149–1
ISSN 1239–1891