Tommi Lehtinen | Alexander Okhotin

On equations $X + X + C = X + X + D$ and $X + E = F$ with unknown $X \subseteq \mathbb{N}$

Turku Centre *for* Computer Science

TUCS Technical Report
No 952, May 2009

On equations $X + X + C = X + X + D$ and $X + E = F$ with unknown $X \subseteq \mathbb{N}$

Tommi Lehtinen
Department of Mathematics, University of Turku
Turku FIN–20014, Finland
tojleht@utu.fi

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 952, May 2009

Abstract

It is shown that the recently discovered computational universality in systems of equations over sets of numbers occurs already in systems of the simplest form, with one unknown X and two equations $X + X + C = X + X + D$ and $X + E = F$, where C, D, E, $F \subseteq \mathbb{N}$ are four ultimately periodic constants and $+$ denotes the operation of elementwise addition of sets, $S+T = \{m+n \mid m \in S, n \in T\}$.

Keywords: Language equations, equations over sets of numbers, computational univesality.

TUCS Laboratory
Discrete Mathematics for Information Technology

1 Introduction

Equations over sets of natural numbers are a particular case of language equations [9], with an alphabet consisting of a single letter. Until recently, nothing was known about these equations beyond the fact that they are nontrivial: this was demonstrated by Leiss [11], who constructed an equation with a non-periodic solution.

Recently Jeż [2] has introduced a new method of constructing systems of equations of the form

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

where X_i are unknown sets of numbers and the right-hand sides φ_i may contain the following operations: union, intersection, *addition* of sets defined as $S + T = \{m + n \mid m \in S,\ n \in T\}$, and singleton constant sets. These equations correspond to conjunctive grammars [12] over a one-letter alphabet. The method of Jeż [2] was further explored by Jeż and Okhotin [3, 4], who have subsequently used it [5] to establish computational completeness of equations over sets of numbers of a more general form

$$\begin{cases} \varphi_1(X_1, \ldots, X_n) &= \psi_1(X_1, \ldots, X_n) \\ & \vdots \\ \varphi_m(X_1, \ldots, X_n) &= \psi_m(X_1, \ldots, X_n) \end{cases}$$

where both left-hand and right-hand sides may use union, addition and singleton constants. To be precise, it was proved that a set of numbers is represented by a unique (least, greatest) solution of such a system if and only if it is recursive (r.e., co-r.e., respectively).

The next result due by Jeż and Okhotin [6] was a simulation of a system with union and addition by a system using addition only, such that every solution $X_i = S_i$ of the original system is represented by a solution $X_i = S'_i = \sigma(S_i)$ of the new system, with $16n + 13 \in S'_i$ if and only if $n \in S_i$. This, in particular, leads to a representation of a set $\sigma(S)$ by unique (least, greatest) solutions of systems , for every recursive (r.e., co-r.e., respectively) set S. On the other hand, it was proved by Lehtinen and Okhotin [10] that some quite simple sets cannot be specified by equations using only addition without any encoding, and therefore equations with addition only are slightly less powerful that equations with union and addition.

This paper continues to explore simple cases of equations over sets of numbers by demonstrating computational universality of systems of two equations of the form

$$\begin{cases} X + X + C &= X + X + D \\ X + E &= F \end{cases}$$

where X is the unique unknown and C, D, E and F are ultimately periodic constant sets. The final result is stated as follows: for every recursive (r.e., co-r.e.) set S there is a system of two equations of the above form with a unique (least, greatest, respectively) solution S', satisfying $np+t \in S'$ if and only if $n \in S_i$, for some constants $p \geqslant 1$ and $t \geqslant 0$. At the same time, some limitations of the expressive power of univariate equations are exposed, and thus these systems are again a little less powerful than systems with multiple variables.

2 Existing construction

Let $\mathbb{N} = \{0, 1, 2, \ldots\}$ be the set of natural numbers including 0, and let $S, T \subseteq \mathbb{N}$ be its subsets. The *sum* of these sets is the set $S+T = \{m+n \,|\, m \in S,\ n \in T\}$.

Define a partial order of componentwise inclusion on vectors of sets of numbers by $(S_1, \ldots, S_n) \sqsubseteq (S'_1, \ldots, S'_n)$ if $S_i \subseteq S'_i$ for all i. For systems with multiple solutions, there is sometimes the *least* or the *greatest* solution with respect to this order.

Equations over sets of numbers using two operations, union and addition, have recently been proved computationally complete.

Theorem 1 (Jeż, Okhotin [5])**.** *For every recursive (r.e., co-r.e.) set $S \subseteq \mathbb{N}$ there exists a system of equations*

$$\begin{cases} \varphi_1(X_1, \ldots, X_n) &= \psi_1(X_1, \ldots, X_n) \\ & \vdots \\ \varphi_m(X_1, \ldots, X_n) &= \psi_m(X_1, \ldots, X_n) \end{cases}$$

with φ_j, ψ_j using singleton constants and the operations of union and addition, which has a unique (least, greatest, respectively) solution with $X_1 = S$.

As a matching upper bound, unique (least, greatest) solutions of equations over sets of numbers with any Boolean operations and addition are known to be recursive (r.e., co-r.e., respectively), so this result precisely characterizes the families of sets representable by solutions of such equations.

For systems of equations over sets of numbers with addition as the only operation, a computational universality result was recently established by Jeż and Okhotin [5]. The idea was to take any recursive (r.e., co-r.e.) set S and consider its *encoding*: another set S' with $16n + 13 \in S'$ if and only if $n \in S$. Then it was proved that any system as in Theorem 1 (that is, with the operations of union and addition) can be simulated by another system using addition only, which manipulates such encodings of sets instead of the sets in their original form.

This encoding of sets requires the following notation: for each $S \subseteq \mathbb{N}$, $p \geqslant 1$ and $i \in \{0, 1, \ldots, p-1\}$, define

$$\tau_i^p(S) = \{pn + i \mid n \in S\}.$$

A set of this form will be called a *track*, and it will be said that the set S is encoded *on the i-th track*. Tracks of the form $\tau_i^p(\mathbb{N})$ are called *full tracks*, and whenever a set S' has $S' \cap \tau_i^p(\mathbb{N}) = \varnothing$, it will be said that S' has an empty i-th track.

Now, setting $p = 16$, the encoding of a set is defined as follows:

Definition 1 (Jeż, Okhotin [6])**.** *For every set $S \subseteq \mathbb{N}$, its* encoding *is the set*

$$\sigma(S) = \{0\} \cup \tau_6^{16}(\mathbb{N}) \cup \tau_8^{16}(\mathbb{N}) \cup \tau_9^{16}(\mathbb{N}) \cup \tau_{12}^{16}(\mathbb{N}) \cup \tau_{13}^{16}(S).$$

Theorem 2 (Jeż, Okhotin [6])**.** *For every recursive (r.e., co-r.e.) set $S \subseteq \mathbb{N}_0$ there exists a system of equations*

$$\begin{cases} \varphi_1(X_1, \ldots, X_n) & = & \psi_1(X_1, \ldots, X_n) \\ & \vdots & \\ \varphi_m(X_1, \ldots, X_n) & = & \psi_m(X_1, \ldots, X_n) \end{cases}$$

with φ_j, ψ_j using the operation of addition and ultimately periodic constants, which has a unique (least, greatest, respectively) solution with $X_i = \sigma(S_i)$ for some $S_i \subseteq \mathbb{N}$, of which $S_1 = S$.

Given a Turing machine recognizing S (the complement of S in the case of a greatest solution), such a system can be effectively constructed.

Though every set can be represented in an encoded form, not all sets can be represented as they are. A class of non-representable sets has been found. These are sets satisfying the following two conditions. First, they must be *prime* in the sense of having no nontrivial representation as a sum of two sets:

Definition 2. *A set $S \subseteq \mathbb{N}$ is prime if $S = S_1 + S_2$ implies $S_1 = \{0\}$ or $S_2 = \{0\}$.*

Second, they are *fragile*, which means that the sum of this set with any set containing at least two elements is co-finite.

Definition 3. *A set $S \subseteq \mathbb{N}$ is fragile if $S + \{n_1, n_2\}$ is co-finite for all $n_1, n_2 \in \mathbb{N}$ with $n_1 \neq n_2$.*

This definition is equivalent to the statement that for every $k \geqslant 1$ there are only finitely many numbers $n \in \mathbb{N}$ with $n, n + k \notin S$.

Theorem 3 (Lehtinen, Okhotin [10])**.** *No set that is prime and fragile is representable by systems of equation over natural numbers with operation of addition and ultimately periodic constants.*

There exist computationally easy sets that are prime and fragile.

3 Encoding into one variable

In this section a system of equations over multiple variables is simulated by a system with only one variable. The constructed system will then be further simulated by a system with one variable and only two equations.

Assume that the simulated system has m variables $X_1, X_2, \ldots, X_m$. The equations with constants are $X_i = E_i$ for $i = 1, 2, \ldots, c$, where E_i is an ultimately periodic constant containing zero. The rest of the equations contain only variables and are of the form $X_k + X_\ell = X_{k'} + X_{\ell'}$. The system will be simulated in such a way that solutions $X_i = S_i$ for $i = 1, \ldots, m$ of the original system are encoded into tracks of solutions of the new equation. Only solutions that have zero in all the sets S_i are considered, other solutions are not represented in the new system.

The sets $S_1, \ldots, S_m$ are encoded with $p = 2^{m+2}$ tracks so that the set S_j will be on the track $d_j = \frac{3p}{8} + 2^{j-1} - 1$. The encoding is given by:

$$\pi(S_1, \ldots, S_m) = \bigcup_{i=0}^{\frac{p}{4}-1} \tau_i^p(\mathbb{N}) \cup \bigcup_{j=1}^{m} \tau_{d_j}^p(S_j). \tag{1}$$

The idea of the encoding is that in the sum $\pi(S_1, \ldots, S_m) + \pi(S_1, \ldots, S_m)$ the encoding of the sum $S_k + S_\ell$ is on track $d_k + d_\ell$. Note that the numbers $d_k + d_\ell$ are pairwise different (cf. the encoding for conjunctive grammars due to Jeż and Okhotin [4]).

The simulation represents three different conditions. One equation guarantees that all solutions of the new system are valid π-encodings of some m sets: this equation is formulated in Lemma 1. All equations $X_i = E_i$ are verified by a single equation given in Lemma 2. And for every equation $X_k + X_\ell = X_{k'} + X_{\ell'}$ there is a corresponding equation stated in Lemma 3.

The following Lemma formalizes an equation that has only correct π-encodings as a solution.

Lemma 1. *A set $S \subseteq \mathbb{N}$ is of the form $S = \pi(S_1, \ldots, S_m)$ for some $S_1, \ldots, S_m \subseteq \mathbb{N}$ with $0 \in S_i$ for all i if and only if it satisfies the equation*

$$X + (\tau_0^p(\mathbb{N}) \cup \{\tfrac{3p}{4}\}) = \bigcup_{i=0}^{\frac{p}{4}-1} \tau_i^p(\mathbb{N}) \cup \bigcup_{j=1}^{m} \tau_{d_j}^p(\mathbb{N}) \cup \bigcup_{k=\frac{3p}{4}}^{p-1} \tau_k^p(\mathbb{N})$$

Proof. Let $0 \in S_i \subseteq \mathbb{N}$ and $S = \pi(S_1, \ldots, S_m)$. Then

$$S + \tau_0^p(\mathbb{N}) = \bigcup_{i=0}^{\frac{p}{4}-1} \tau_i^p(\mathbb{N}) \cup \bigcup_{j=1}^{m} \tau_{d_j}^p(\mathbb{N})$$

and

$$S + \{\tfrac{3p}{4}\} = \bigcup_{k=\frac{3p}{4}}^{p-1} \tau_k^p(\mathbb{N}) \cup \bigcup_{j=1}^{m} \tau_{d_j - \frac{p}{4}}^p(S_j + \{1\}).$$

The union of these sets gives the right-hand side of the equation and thus S is a solution.

Conversely, let $X = S$ be a solution of the equation. All tracks that are empty in the right-hand side are empty in S. To see that tracks $\frac{3p}{4}$, $\frac{3p}{4}+1$, $\ldots$, $p-1$ are empty in S, suppose that there is $n \in S$ with $n \equiv \frac{3p}{4}+i \mod p$ for some $i \in \{0, \ldots, \frac{p}{4}-1\}$; but then the number $n + \frac{3p}{4}$ cannot be in the right-hand side, because it belongs to track $\frac{p}{2}+i$ that is empty in the right-hand side, which is a contradiction. Since these tracks $\frac{3p}{4}, \frac{3p}{4}+1, \ldots, p-1$ are full on the right-hand side, they can only be obtained by shifting full tracks in S by $\frac{3p}{4}$; that is, tracks 0, 1, $\ldots$, $\frac{p}{4}-1$ are full in S. It remains to show that each data track d_j, for $j = 1, 2, \ldots, m$, contains the encoding of 0, that is, the number d_j. Since the right-hand side contains a full track d_j and $d_j - \frac{3p}{4} < 0$, this number can be obtained only by having $d_j \in S$. It follows that $S = \pi(S_1, \ldots, S_m)$, for some $S_1, \ldots, S_m \subseteq \mathbb{N}$. □

The constant equations of the original system are checked by the following equation:

Lemma 2. *The sets $0 \in S_i \subseteq \mathbb{N}$ for $i = 1, \ldots, c$ satisfy the equations $X_\ell = E_\ell$ for $\ell = 1, \ldots, c$ if and only if $\pi(S_1, \ldots, S_m)$ satisfies the equation*

$$X + (\tau_0^p(\mathbb{N}) \cup \{p-1-d_c\}) = \\ = \bigcup_{i=0}^{\frac{p}{4}-1} \tau_i^p(\mathbb{N}) \cup \bigcup_{j=1}^{m} \tau_{d_j}^p(\mathbb{N}) \cup \bigcup_{k=p-1-d_c}^{p-1-d_c+\frac{p}{4}-1} \tau_k^p(\mathbb{N}) \cup \bigcup_{l=1}^{c} \tau_{p-1-d_c+d_\ell}^p(E_\ell) \quad (2)$$

Proof. The sum with $\tau_0^p(\mathbb{N})$ gives:

$$\pi(S_1, \ldots, S_m) + \tau_0^p(\mathbb{N}) = \bigcup_{i=0}^{\frac{p}{4}-1} \tau_i^p(\mathbb{N}) \cup \bigcup_{j=1}^{m} \tau_{d_j}^p(\mathbb{N}).$$

When summed to $\{p-1-d_c\}$ the tracks are just moved, the full tracks to tracks $p-1-d_c, \ldots, p-1-d_c+\frac{p}{4}-1$:

$$\bigcup_{i=0}^{\frac{p}{4}-1} \tau_i^p(\mathbb{N}) + \{p-1-d_c\} = \bigcup_{k=p-1-d_c}^{p-1-d_c+\frac{p}{4}-1} \tau_k^p(\mathbb{N}).$$

And the sums $\tau_{d_j}^p(S_j) + \{p-1-d_c\})$ yield $\tau_{p-1-d_c+d_j}^p(S_j)$ for $j = 1, \ldots, c$, which are reflected in the right-hand side of (2), while $\tau_{d_c+d_j-1}^p(S_j + \{1\}) \subseteq \bigcup_{i=0}^{\frac{p}{4}-1} \tau_i^p(\mathbb{N})$ for $j = (c+1), \ldots, m$ is overwritten by $\bigcup_{i=0}^{\frac{p}{4}-1} \tau_i^p(\mathbb{N})$. □

Lemma 3. *The sets $S_1, \ldots, S_m \subseteq \mathbb{N}$ satisfy the equation $X_k + X_\ell = X_{k'} + X_{\ell'}$ if and only if $\pi(S_1, \ldots, S_m)$ satisfies the equation*

$$X + X + \{0, \tfrac{p}{4}, p-1-d_k-d_\ell\} = X + X + \{0, \tfrac{p}{4}, p-1-d_{k'}-d_{\ell'}\}$$

Proof. The sum $S_k + S_\ell$ will be on the track $d_k + d_\ell = \frac{3p}{4} + 2^{k-1} + 2^{\ell-1} - 2$ of $\pi(S_1, \ldots, S_m) + \pi(S_1, \ldots, S_m)$:

$$\pi(S_1, \ldots, S_m) + \pi(S_1, \ldots, S_m) = \bigcup_{i=0}^{\frac{3p}{4}-2} \tau_i^p(\mathbb{N}) \cup \bigcup_{k \leqslant \ell} \tau_{d_k+d_\ell}(S_k + S_\ell).$$

By adding the constant $\{0, \frac{p}{4}, p-1-d_k-d_\ell\}$ to $\pi(S_1, \ldots, S_m) + \pi(S_1, \ldots, S_m)$ the sum $S_k + S_\ell$ ends up to track $p-1$, while all other tracks are full:

$$\pi(S_1, \ldots, S_m)+\pi(S_1, \ldots, S_m)+\{\tfrac{p}{4}, p-1-d_k-d_\ell\} = \textstyle\bigcup_{i=0}^{p-2} \tau_i^p(\mathbb{N}) \cup \tau_{p-1}(S_k+S_\ell)$$

The claim of the lemma is a direct consequence of this. □

To sum up the construction, the original system has variables $X_1, \ldots, X_m$, equations with constants

$$X_i = E_i, \quad \text{for } i = 1, \ldots, c$$

and equations without constants

$$X_k + X_\ell = X_{k'} + X_{\ell'}, \quad (k, \ell, k', \ell') \in V.$$

The new system has only one variable X, equations $X + E_1 = F_1$, $X + E_2 = F_2$, where E_i and F_i are given in Lemmas 1 and 2 and equations

$$X + X + \{0, \tfrac{p}{4}, p-1-d_k-d_\ell\} = X + X + \{0, \tfrac{p}{4}, p-1-d_{k'}-d_{\ell'}\}$$

for all $(k, \ell, k', \ell') \in V$. The correctness of the construction is stated as follows:

Proposition 1. *A set S is a solution of the constructed system if and only if there are sets $0 \in S_i \subseteq \mathbb{N}$ for $i = 1, \ldots, m$ such that $S = \pi(S_1, \ldots, S_m)$ and $X_i = S_i$ is a solution of the original system.*

4 Encoding into two equations

The equations produced in the previous section are defined uniformly and differ only by the constants. They can be merged into four constants as follows:

Lemma 4. *Consider a system of equations of the form $\{X + X + C_i = X + X + D_i \mid 0 \leqslant i \leqslant m-1\} \cup \{X + E_i = F_i \mid 0 \leqslant i \leqslant m'-1\}$ with*

$m \geqslant 0$, $m' \geqslant 1$ and $C_i, D_i, E_i, F_i \subseteq \mathbb{N}$, and assume $E_i \neq \varnothing$ for all i. Let $p = \max(m, m'+1)$ and define

$$C = \bigcup_{i=0}^{m-1} \{pn + i \mid n \in C_i\}, \qquad D = \bigcup_{i=0}^{m-1} \{pn + i \mid n \in D_i\},$$
$$E = \bigcup_{i=0}^{m'-1} \{pn + i \mid n \in E_i\}, \qquad F = \bigcup_{i=0}^{m'-1} \{pn + i \mid n \in E_i\}.$$

Then $X = S$ is a solution of the above system if and only if $X = S' = \{np \mid n \in S\}$ is a solution of the following system of two equations:

$$X + X + C = X + X + D$$
$$X + E = F,$$

and all solutions of the latter system are of this form.

Proof. The first claim is that every solution $X = S'$ of the second system contains only numbers equivalent to 0 modulo p, that is, $S' \subseteq \{np \mid n \geqslant 0\}$. Suppose there is a number $np + i \in S'$, where $1 \leqslant i \leqslant p-1$, and let $j = \min(p - 1 - i, m' - 1)$. Then $0 \leqslant j \leqslant m' - 1$ and $m' \leqslant j + i \leqslant p - 1$. Since it is known that $E_j \neq \varnothing$, there exists a number n' with $n'p + j \in E$. Then $(np + i) + (n'p + j) \in S' + E$, and thus $(n + n')p + (i + j)$ must be in F, which is a contradiction, as all numbers in F are between 0 and $m' - 1$ modulo p.

Let $S \subseteq \mathbb{N}$ and define $S' = \{np \mid n \in S\}$. It should be proved that S is a solution of the original system if and only if S' is a solution of the new system.

Claim 1. *For all i and n, $n \in S + S + C_i$ if and only if $pn + i \in S' + S' + C$*

If $n \in S + S + C_i$, then $n = n_1 + n_2 + n_3$ with $n_1, n_2 \in S$ and $n_3 \in C_i$. Then $pn_1, pn_2 \in S'$ and $pn_3 + i \in C$, and therefore $pn + i \in S' + S' + C$.

Conversely, if $pn + i \in S' + S' + C$, then $pn + i = pn_1 + pn_2 + pn_3 + i$ for some numbers $pn_1, pn_2 \in S'$ and $pn_3 + i \in C$, since all numbers in S' are equivalent to 0 modulo p, and so the number taken from C can only be equal to i modulo p. Then, by the definition of S' and C $n_1, n_2 \in S$ and $n_3 \in C_i$, and thus $n \in S + S + C_i$.

Claim 2. *For each $i \in \{0, \ldots, m-1\}$, $S + S + C_i = S + S + D_i$ if and only if $(S' + S' + C) \cap \tau_i^p(\mathbb{N}) = (S' + S' + D) \cap \tau_i^p(\mathbb{N})$*

If $S + S + C_i = S + S + D_i$, then, by Claim 1, a number $np + i$ is in $S' + S' + C$ if and only if $n \in S + S + C_i$, which equivalent to $n \in S + S + D_i$ by the equation. The latter holds if and only if $np + i \in S' + S' + D$, by Claim 1 again. Thus the new equation holds modulo intersection with $\tau_i^p(\mathbb{N})$. Conversely, assuming that $(S' + S' + C) \cap \tau_i^p(\mathbb{N}) = (S' + S' + D) \cap \tau_i^p(\mathbb{N})$,

consider that $n \in S + S + C_i$ if and only if $np + i \in S' + S' + C$ by Claim 1, which is the same as $np + i \in S' + S' + D$ by the equation, and then which holds if and only if $n \in S + S + D_i$.

Claim 3. *For every $i \in \{0, \ldots, m' - 1\}$, $S + E_i = F_i$ if and only if $(S' + E) \cap \tau_i^p(\mathbb{N}) = F \cap \tau_i^p(\mathbb{N})$.*

Proved in the same way as the previous claim. □

5 Representable sets and decision problems

Theorem 4. *For every recursive (r.e., co-r.e.) set $S \subseteq \mathbb{N}_0$ there exist numbers $p, d \geqslant 1$, finite sets $C, D \subseteq \mathbb{N}_0$ and ultimately periodic sets $E, F \subseteq \mathbb{N}_0$, such that the system of two equations*

$$\begin{cases} X + X + C &= X + X + D \\ X + E &= F \end{cases}$$

with an unknown $X \subseteq \mathbb{N}_0$ has a unique (least, greatest, respectively) solution $X = S'$, such that $n \in S$ if and only if $pn + d \in S'$.

Given a Turing machine recognizing S (the complement of S in the case of a greatest solution), such p, d, C, D, E and F can be effectively constructed.

Proof. Consider the system given by Theorem 2 for S, which has variables $X_1, \ldots, X_{\widetilde{m}}$ and all equations of the form $X_{k_1} + \ldots + X_{k_s} + C = X_{k_{s+1}} + \ldots + X_{k_t} + D$. The unique (least, greatest) solution of this system is $(S_1, \ldots, S_{\widetilde{m}})$, where $p_0 n + d_0 \in S_1$ if and only if $n \in S$, where $p_0 = 16$ and $d_0 = 13$. Furthermore, since each S_i is a σ-encoding of some set, it is known that $0 \in S_i$.

This system can be transformed to a system in variables $X_1, \ldots, X_m$ for some $m \geqslant \widetilde{m}$, and with equations of the form $X_{k_1} + X_{k_2} = X_{k_3} + X_{k_4}$ and $X_1 = C_i, \ldots, X_{m'} = C_{m'}$ for some $1 \leqslant m' \leqslant m$. This is done by separating subexpressions into extra variables, and by permuting the variables so that variables with equations of the form $X_i = C$ have smaller numbers. The unique (least, greatest) solution of the latter system is $(S'_1, \ldots, S'_m)$, with all S'_i containing zero, and there is an index i_0 with $S'_{i_0} = S_1$.

By the constructions in Section 3, there exists a system with a unique variable X and with all equations of the form $X + X + C_i = X + X + D_i$ and $X + E_i = F_i$, and its solutions correspond to the solutions of the previous system in variables $(X_1, \ldots X_m)$ as stated in Proposition 1. In particular, if the previous system has a unique (least, greatest) solution $X_i = S_i$, then $X = \widehat{S} = \pi(S'_1, \ldots, S'_m)$ will be the unique (least, greatest) solution of the constructed system, with $p_1 n + d_1 \in \widehat{S}$ if and only if $n \in S'_{i_0}$, for some $p_1 \geqslant 1$ and $d_1 \geqslant 0$.

Finally, applying Lemma 4 to the latter system gives the system of two equations $X + X + C = X + X + D$ and $X + E = F$, which has a unique (least, greatest) solution $X = S'$, with $p_2 n + d_2 \in S'$ if and only if $n \in \widehat{S}$.

Now $n \in S$ if and only if $p_0 n + d_0 \in S_1$ if and only if $p_1(p_0 n + d_0) + d_1 \in \widehat{S}$ if and only ig $p_2(p_1(p_0 n + d_0) + d_1) + d_2 \in S'$, and setting $p = p_0 p_1 p_2$ and $d = d_0 p_1 p_2 + d_1 p_2 + d_2$ proves the theorem. □

Since the constructions preserve the cardinality of the set of solutions, the decision problems about this cardinality maintain their level of undecidability:

Theorem 5. *The problem of whether a given system of two equations*

$$\begin{cases} X + X + C &= X + X + D \\ X + E &= F \end{cases}$$

with an unknown $X \subseteq \mathbb{N}_0$ has a solution (unique solution, finitely many solutions) is Π_1-complete (Π_2-complete, Σ_3-complete), respectively.

6 Limitations of one variable

It was recently proved by the authors [10] that systems of equations with multiple variables using addition only cannot represent sets that are both prime and fragile. At the same time, some fragile sets that are not prime can be represented [10]. It turns out that none of these sets are representable using a single variable.

Lemma 5. *If a fragile set is the greatest solution of a univariate system, then it is co-finite.*

Proof. Let S be this fragile set.

For every equation of the form $X + \ldots + X + C = X + \ldots + X + D$, if there are multiple X's in either side or $|C| > 1$ or $|D| > 1$, then the substitution of S yields a co-finite set. In this case, consider the number, starting from which all elements are there. Otherwise the equation is of the form $X + \{m\} = X + \{m\}$ and let the number be 0.

For each equation of the form $X + \ldots + X + E = F$, if there are multiple X's or $|E| > 1$, then F is co-finite and consider the number, starting from which all elements are there. Otherwise the equation is $X + \{m\} = F$, which immediately proves that S is co-finite.

Let n_0 be the maximum of the above numbers. Then $\mathbb{N} + \{n_0\} \subseteq S + \ldots + S + E$ for all sums appearing in the system of equations. This yields

$$\begin{aligned} \big(S \cup (\mathbb{N} + \{n_0\})\big) + \ldots + \big(S \cup (\mathbb{N} + \{n_0\})\big) + E = \\ = S + \ldots + S + E \cup \underbrace{\big((\mathbb{N} + \{n_0\}) + (\ldots)\big)}_{\subseteq \mathbb{N} + \{n_0\}} = S + \ldots + S + E, \end{aligned}$$

and all equations in the system are satisfied. Since S is the greatest solution, it follows that $\mathbb{N} + \{n_0\} \subseteq S$ and thus S is co-finite. □

Lemma 6. *If a fragile set is the least solution of a univariate system, then it is co-finite.*

$$\pi(S_1, \ldots, S_m) = \bigcup_{i=0}^{\frac{p}{4}-1} \tau_i^p(\mathbb{N}) \cup \bigcup_{j=1}^{m} \tau_{d_j}^p(S_j). \tag{3}$$

Proof. Let S be a fragile solution of the system.

If the system contains an equation of the form $X + \{m\} = F$, then it has a unique solution, which is co-finite. Every set satisfies an equation of the form $X + \{m\} = X + \{m\}$ and they don't have to be considered. Equations $X + \{m\} = X + \{n\}$ with $m \neq n$ have the emtpty set as the only solution so they cannot appear.

Suppose then there are only equations of the forms $X + \ldots + X + C = X + \ldots + X + D$ and $X + \ldots + X + E = F$, where the sides have multiple occurrences of X or a constant with at least two elements.

Let k be bigger than the differences of the two smallest numbers in S or in any constant with at least two elements appearing in the system.

Since S is fragile, there is a number ℓ, such that every pair of missing numbers $m, n \notin S$ with $n > m \geqslant \ell$ satisfies $n - m > k$.

Let $\ell' \geqslant \ell$ be such that $\{\ell', \ell' + 1, \ldots, \ell' + 2k\} \subseteq S$. Such ℓ' exists by the fragility of S. Now $S \setminus \{\ell' + k\}$ is a solution also. Consider the sum $(S \setminus \{\ell' + k\}) + \ldots + (S \setminus \{\ell' + k\}) + E$. It is obviously a subset of $S + \ldots + S + E$, to prove the lemma it is to be shown that also the converse inclusion holds. So let $m_1 + m_2 + \cdots + m_n + e \in S + \ldots + S + E$ with $m_i \in S$ and $e \in E$. If all m_i are different from $\ell' + k$, then $m_i \in S \setminus \{\ell' + k\}$ and $m_1 + m_2 + \cdots + m_n + e \in (S \setminus \{\ell' + k\}) + \ldots + (S \setminus \{\ell' + k\}) + E$. Suppose then that $m_i = \ell' + k$ for some set of indexes i. For every pair of such i and i' the numbers can be replaced by $m_i + 1$ and $m_{i'} - 1$. So it can be assumed that there is only one such i, say $i = 1$. Let $a_1 < a_2$ be the two smallest numbers in S, (or in the case that the equation only has one X the two smallest numbers of E). Then one of $m_1 + m_2 - a_1$ or $m_1 + m_2 - a_2$ ($m_1 + e - a_1$ or $m_1 + e - a_2$ in the second case) is in $S \setminus \{\ell' + k\}$, since they are bigger than ℓ and their difference is less than k. Suppose it is $m_1 + m_2 - a_1$. Now

$$\begin{aligned}(m_1 + m_2 - a_1) + a_1 + m_3 + \ldots + m_n + e = \\ = m_1 + m_2 + \cdots + m_n + e \in (S \setminus \{\ell' + k\}) + \ldots + (S \setminus \{\ell' + k\}) + E.\end{aligned}$$

Thus $S + \ldots + S + E \subseteq (S \setminus \{\ell' + k\}) + \ldots + (S \setminus \{\ell' + k\}) + E$ and the sets are equal. It follows that $S \setminus \{\ell' + k\}$ is a solution of the system, and S cannot be a minimal solution. □

Theorem 6. *There exists a set of natural numbers representable by a unique solution of a multivariate system of equations with addition, which, however, is not a unique (least, greatest) solution of any univariate system.*

7 Limitations of one equation

Systems of two equations

$$\begin{cases} X + X + C &= X + X + D \\ X + E &= F \end{cases}$$

constructed in Theorems 4 and 5 have one equation a constant side one equation without a constant side. It turns out that systems with all equations of the same type (that is, either all with constant sides or all without constant sides) have quite limitated expressive power.

If all equations in a system are without constant sides, their least solution is trivial: $X_i = \varnothing$. Greatest solutions are bound to be trivial as well: as shown in the next lemma, they are always co-finite.

Lemma 7. *If a system of equations of the form* $X_{i_1} + \ldots + X_{i_\ell} + C = X_{j_1} + \ldots + X_{j_m} + D$ *has a solution* $(S_1, \ldots, S_n)$*, then* $(S_1 + \mathbb{N}, \ldots, S_n + \mathbb{N})$ *is a solution as well.*

Proof. Consider the smallest number in $S_{i_1} + \ldots + S_{i_\ell} + C$. Then $(S_{i_1} + \mathbb{N}) + \ldots + (S_{i_\ell} + \mathbb{N}) + C$ contains this number and all numbers that are greater. The claim follows, since all equations have the same smallest numbers in the both sides. □

The other type of systems have all equations with constant sides. It can be shown that every non-periodic solution can be extended to a greater periodic solution, which will have the same period as the common period of the constant sides.

Lemma 8. *If a system of equations of the form* $X_1 + \ldots + X_m + E = F$ *has a solution* $(S_1, \ldots, S_n)$*, then it has an ultimately periodic solution* $(S'_1, \ldots, S'_n)$ *with* $S_i \subseteq S'_i$*.*

Such a statement holds in a much more general case of language equations, and can be inferred from the syntactic monoid and Conway's [1] results.

Proof. Let p and d be numbers, such that for every equation $X_1 + \ldots + X_m + E = F$, F has period p starting from d. Define $S'_i = S_i \cup (S_i \cap \mathbb{N}_{+d}) + \{np | n \geqslant 0\}$. Then it is easy to check that $X_i = S'_i$ is still a solution. □

Such an argument does not work for least solutions. In fact, it is easy to construct an equation with uncountably many pairwise incomparable solutions [7]:

$$X + \{0, 1\} = \mathbb{N}$$

In case there exists a least solution (or at least countably many solutions), no examples of nontrivial expressive power are known. Though it cannot be ruled out that these equations might be able to represent some nonperiodic set, constructing such a representation is beyond the current knowledge on equations over sets of numbers. It is more likely that equations with constant right-hand sides cannot, after all, express anything beyond ultimately periodic sets.

References

[1] J. H. Conway, *Regular Algebra and Finite Machines*, Chapman and Hall, 1971.

[2] A. Jeż, "Conjunctive grammars can generate non-regular unary languages", *International Journal of Foundations of Computer Science*, 19:3 (2008), 597–615.

[3] A. Jeż, A. Okhotin, "Conjunctive grammars over a unary alphabet: undecidability and unbounded growth", *Theory of Computing Systems*, to appear.

[4] A. Jeż, A. Okhotin, "One-nonterminal conjunctive grammars over a unary alphabet", *Computer Science in Russia* (CSR 2009, Novosibirsk, Russia, 18–23 August, 2009), to appear.

[5] A. Jeż, A. Okhotin, "On the computational completeness of equations over sets of natural numbers", *Automata, Languages and Programming* (*ICALP 2008*, Reykjavík, Iceland, July 6–13, 2008), part II, LNCS 5126, 63–74.

[6] A. Jeż, A. Okhotin, "Equations over sets of natural numbers with addition only", *STACS 2009* (Freiburg, Germany, 26–28 February, 2009), 577–588.

[7] J. Karhumäki, M. Kunc, personal communication, September 2005.

[8] M. Kunc, "The power of commuting with finite sets of words", *Theory of Computing Systems*, 40:4 (2007), 521–551.

[9] 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.

[10] T. Lehtinen, A. Okhotin, "On equations over sets of numbers and their limitations", *Developments in Language Theory* (DLT 2009, Stuttgart, Germany, 30 June–3 July, 2009), LNCS 5583, to appear.

[11] E. L. Leiss, "Unrestricted complementation in language equations over a one-letter alphabet", *Theoretical Computer Science*, 132 (1994), 71–93.

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

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 978-952-12-2329-7
ISSN 1239-1891