Equations over Sets of Natural Numbers with Addition Only

Artur Jez, Alexander Okhotin

Symposium on Theoretical Aspects of Computer Science 2009 (Freiburg), pp. 577–588
www.stacs-conf.org

EQUATIONS OVER SETS OF NATURAL NUMBERS WITH ADDITION ONLY

ARTUR JEŻ [1] AND ALEXANDER OKHOTIN [2]

[1] Institute of Computer Science, University of Wrocław, Poland
E-mail address: aje@ii.uni.wroc.pl

[2] Department of Mathematics, University of Turku, Finland; Academy of Finland
E-mail address: alexander.okhotin@utu.fi

Abstract. Systems of equations of the form $X = YZ$ and $X = C$ are considered, in which the unknowns are sets of natural numbers, "+" denotes pairwise sum of sets $S+T = \{m+n \mid m \in S,\ n \in T\}$, and C is an ultimately periodic constant. It is shown that such systems are computationally universal, in the sense that for every recursive (r.e., co-r.e.) set $S \subseteq \mathbb{N}$ there exists a system with a unique (least, greatest) solution containing a component T with $S = \{n \mid 16n+13 \in T\}$. This implies undecidability of basic properties of these equations. All results also apply to language equations over a one-letter alphabet with concatenation and regular constants.

1. Introduction

Language equations are equations of the form

$$\varphi(X_1, \ldots, X_n) = \psi(X_1, \ldots, X_n), \qquad (*)$$

in which the unknowns X_i are formal languages, while the expressions φ, ψ use language-theoretic operations, such as concatenation, Kleene star and Boolean operations, and constant languages. It is well-known that systems of the *resolved form* $X_i = \varphi_i(X_1, \ldots, X_n)$ $(1 \leqslant i \leqslant n)$ with union, concatenation and singleton constants define the semantics of the context-free grammars [1]. If intersection is also allowed, such equations characterize an extension of the context-free grammars known as *conjunctive grammars* [9] and notable for efficient parsing algorithms.

The expressive power of language equations of the general form (*) was determined by Okhotin [10, 11], who proved that a language is representable by a unique (least, greatest) solution of a system with concatenation, Boolean operations and singleton constants if and only if this language is recursive (recursively enumerable, co-r.e., respectively). The same expressive power is attained using concatenation with constants and union [11]. It was subsequently discovered that language equations can be computationally universal even without any Boolean operations: Kunc [6] constructed a finite language $L \subseteq \{a, b\}^*$, for

(A. Jeż) Supported by MNiSW grant number N206 024 31/3826, 2006–2008 and N206 259035 2008–2010.
(A. Okhotin) Supported by the Academy of Finland under grant 118540.

which the greatest solution of a language equation $LX = XL$ is Π_1-hard (that is, hard for co-r.e. sets). This paper establishes a similar result in the seemingly trivial case of a one-letter alphabet.

Unary languages, defined over an alphabet $\{a\}$, form an important special class of formal languages. It is well-known that context-free grammars over this alphabet generate only regular languages. The first example of a unary language equation with a non-regular unique solution was constructed by Leiss [7]: this was an equation $X = \varphi(X)$ with φ containing concatenation, complementation and constant $\{a\}$. The question of whether conjunctive grammars (in other words, systems of language equations with union, intersection and concatenation) can generate any non-regular languages had been a long-standing open problem [9], until Jeż [2] constructed a conjunctive grammar generating $\{a^{4^n} \mid n \geqslant 0\}$. The ideas of this example were used by Jeż and Okhotin [3] to establish some general results on the expressive power of these equations, as well as the EXPTIME-completeness of their solutions [4]. For systems of the general form (*) using concatenation and union, it has recently been shown by the authors [5] that they are computationally complete.

As unary languages can be regarded as sets of natural numbers, unary language equations are naturally viewed as *equations over sets of numbers*. Concatenation of languages accordingly turns into *addition* of sets $S + T = \{m + n \mid m \in S,\ n \in T\}$, an operation that has been a subject of much study in number theory and combinatorics [14]. For instance, if P is the set of primes, then the equation $P + P + P = \{6, 7, \ldots\}$ expresses the Goldbach conjecture. Computational complexity of expressions and circuits over sets of numbers with addition and different sets of Boolean operations has been studied by Stockmeyer and Meyer [13] and McKenzie and Wagner [8]. Equations over sets of numbers are a more general formalism, and its expressive power was related to the allowed Boolean operations in the aforementioned work on unary language equations [2, 3, 4, 5].

This paper is concerned with equations over sets of numbers that use *only addition and no Boolean operations*. These are systems of equations of the form

$$X_{i_1} + \ldots + X_{i_k} + C = X_{j_1} + \ldots + X_{j_\ell} + D$$

in variables $(X_1, \ldots, X_n)$, where $C, D \subseteq \mathbb{N}$ are ultimately periodic constants. In terms of language equations over $\{a\}$, these are equations

$$X_{i_1} \ldots X_{i_k} K = X_{j_1} \ldots X_{j_\ell} L,$$

with regular constants $K, L \subseteq a^*$. This is the ultimately simplest case of language equations, and at the first glance it seems out of question that such equations could have any non-trivial unique (least, greatest) solutions. Probably for that reason no one has ever proclaimed their expressive power to be an open problem. However, as proved in this paper, these equations can have not only non-periodic unique solutions, but in fact are computationally universal. Furthermore, their main decision problems are as hard as similar problems for language equations over multiple-letter alphabets and using all Boolean operations [10, 11].

The new results are directly based on the authors' recent proof of the computational completeness of equations over sets of numbers with addition and union [5], though it is established using completely different techniques. The idea is to take an arbitrary system using addition and union and *encode* it in another system using addition only. The solutions of the two systems will not be identical, but there will be a bijection between solutions based upon an encoding of sets of numbers.

This encoding of sets, defined in Section 3, is an injection $\sigma : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$, which represents every number n of the encoded set as the number $16n + 13$ in the encoding. The given encoding has two key properties. First of all, its form can be *checked* by an equation, which is satisfied exactly by those sets that are valid encodings; such an equation is constructed in Section 3. Second, the sum of any two valid encodings, encodes *both the sum and the union* of the encoded sets of numbers, and furthermore, adding a certain constant to such a sum of encodings produces a set that encodes only the sum of the original sets, while adding another constant allows representing only the union of the original sets. In overall, as shown in Section 4, the sum and the union of any two sets is represented by summing their encodings.

Finally, on the basis of this encoding, in Section 5 it is demonstrated how an arbitrary system of equations over sets of numbers with union and addition can be simulated using addition only. Each variable X_i of the original system will be represented in the new system by a variable X'_i, and the solutions of the new system will be of the form $X'_i = \sigma(S_i)$ for all variables X'_i, where $X_i = S_i$ is a solution of the original system.

All constants in the construction are ultimately periodic; some of them are finite and some are infinite. The last question is whether infinite constants are necessary to specify any non-periodic sets, and an affirmative answer is given in Section 6.

2. Equations over sets of numbers

Throughout this paper, the set of natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$ is assumed to contain zero. A set of numbers $S \subseteq \mathbb{N}$ is *ultimately periodic* if there exist numbers $d \geqslant 0$ and $p \geqslant 1$, such that $n \in S$ if and only if $n + p \in S$ for every $n \geqslant d$. Otherwise, S is *non-periodic*. Note that S is ultimately periodic if and only if the corresponding language $L = \{a^n \mid n \in S\} \subseteq a^*$ is regular.

For every two subsets of natural numbers $S, T \subseteq \mathbb{N}$, their *sum* is the set $\{m + n \mid m \in S,\ n \in T\}$. Other typical operations on sets are the Boolean operations, such as union, intersection and complementation. Using complementation and addition, the first example of an equation with a non-periodic unique solution was constructed:

Example 2.1 (Leiss [7])**.** For every expression φ, denote $2\varphi = \varphi + \varphi$. Then the unique solution of the equation

$$X = 2\Big(\overline{2\big(\overline{2\overline{X}}\big)}\Big) + \{1\}$$

is $\{n \mid \exists i \geqslant 0 :\ 2^{3i} \leqslant n < 2^{3i+2}\} = \{n \mid \text{base-8 notation of } n \text{ starts with } \texttt{1}, \texttt{2} \text{ or } \texttt{3}\}$.

However, the expressive power of this family of equations is still quite limited [12], with some simple languages being non-representable.

The second example of non-periodic solutions of equations over sets of numbers was constructed by Jeż [2] as a conjunctive grammar [9] generating the language $\{a^{4^n} \mid n \geqslant 0\}$. In terms of equations it is stated as follows:

Example 2.2 (Jeż [2])**.** The least solution of the system

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

is $X_1 = \{4^n \mid n \geqslant 0\}$, $X_2 = \{2 \cdot 4^n \mid n \geqslant 0\}$, $X_3 = \{3 \cdot 4^n \mid n \geqslant 0\}$, $X_6 = \{6 \cdot 4^n \mid n \geqslant 0\}$.

The idea behind this example is to manipulate positional notations of numbers, and this idea was subsequently used to establish the following general result on the expressive power of such equations. The statement refers to the family of *linear conjunctive languages* [9], which properly contains the Boolean closure of linear context-free languages.

Proposition 2.3 (Jeż, Okhotin [3])**.** *For every $k \geqslant 2$ and for every linear conjunctive language $L \subseteq \{\mathtt{0}, \mathtt{1}, \ldots, k-1\}^+$ there exists a resolved system of equations*

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

with φ_i using singleton constants and the operations of union, intersection and addition, which has a least solution with $X_1 = \{n \mid$ the base-k notation of n is in $L\}$.

On the basis of this result, it was shown that systems of the general form with the same operations are computationally complete.

Theorem 2.4 (Jeż, Okhotin [5])**.** *For every recursive (r.e., co-r.e.) set $S \subseteq \mathbb{N}$ there exists an unresolved system*

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

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

Exactly the same results hold for unresolved systems with intersection, sum and singleton constants [5], though they will not be used in this paper.

The goal is now to take any system of equations with union and addition, such as those constructed in Theorem 2.4, and to simulate it by another system using addition only. The solutions of the new system will *encode* the solutions of the original system as described in the next section.

3. Encoding of sets

An arbitrary set of numbers $\widehat{S} \subseteq \mathbb{N}$ will be represented by another set $S \subseteq \mathbb{N}$, which contains a number $16n+13$ if and only if n is in $\widehat{S}$. The membership of numbers i with $i \neq 13 \pmod{16}$ in S does not depend on $\widehat{S}$ and will be defined below. Since many constructions in the following will be done modulo 16, the following notation shall be adopted:

Definition 3.1. For each $i \in \{0, 1, \ldots, 15\}$,

$$\textsc{track}_i(S) = \{n \mid 16n + i \in S\},$$
$$\tau_i(S') = \{16n + i \mid n \in S'\}.$$

The subset $S \cap \{16n + i \mid n \geqslant 0\}$ is called the i^{th} *track of* S. A set S is said to have an *empty (full) track* i if $\textsc{track}_i(S) = \varnothing$ ($\textsc{track}_i(S) = \mathbb{N}$, respectively).

In these terms, it can be said that a set $\widehat{S}$ shall be encoded in the 13^{th} track of a set S. The rest of the tracks of S contain technical information needed for the below constructions to work: track 0 contains a singleton $\{0\}$, tracks 6, 8, 9 and 12 are full and the rest of the tracks are empty.

Definition 3.2. For every set $\widehat{S} \subseteq \mathbb{N}$, its *encoding* is the set

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

The first property of the encoding announced in the introduction is that there exists an equation with the set of all valid encodings as its set of solutions. Such an equation will now be constructed.

Lemma 3.3. *A set $X \subseteq \mathbb{N}$ satisfies an equation*

$$X + \{0,4,11\} = \bigcup_{\substack{i \in \{0,4,6,8,9,\\ 10,12,13\}}} \tau_i(\mathbb{N}) \cup \bigcup_{i \in \{1,3,7\}} \tau_i(\mathbb{N}+1) \cup \{11\}$$

if and only if $X = \sigma(\widehat{X})$ for some $\widehat{X} \subseteq \mathbb{N}$.

Proof. $\Leftarrow$ Let X be any set that satisfies the equation. Then the sum $X + \{0,4,11\}$ has empty tracks 2, 5, 14 and 15:

$$\begin{aligned}\textsc{track}_2(X + \{0,4,11\}) = \textsc{track}_5(X + \{0,4,11\}) = \\ = \textsc{track}_{14}(X + \{0,4,11\}) = \textsc{track}_{15}(X + \{0,4,11\}) = \varnothing\end{aligned}$$

For this condition to hold, X must have many empty tracks as well. To be precise, each track t with t, $t+4$ or $t+11 \pmod{16}$ being in $\{2,5,14,15\}$ must be an empty track in X. Calculating such set of tracks, $\{2,5,14,15\} - \{0,4,11\} \pmod{16} = \{1,2,3,4,5,7,10,11,14,15\}$ are the numbers of tracks that must be empty in X.

Similar considerations apply to track 11, as $\textsc{track}_{11}(X + \{0,4,11\}) = \{0\}$. For every track t with $t = 11$, $t + 4 = 11$ or $t + 11 = 11 \pmod{16}$, it must hold that the t^{th} track of X is either an empty track or $\textsc{track}_t(X) = \{0\}$. The latter must hold for at least one such t. Let us calculate all such tracks t: these are tracks with numbers $\{11\} - \{0,4,11\} \pmod{16} = \{0,7,11\}$. Since tracks number 7 and 11 are already known to be empty, it follows that $\textsc{track}_0(X) = \{0\}$.

In order to prove that X is a valid encoding of some set, it remains to prove that tracks number $6,8,9,12$ in X are full. Consider first that $\textsc{track}_3(X + \{0,4,11\}) = \mathbb{N} + 1$. Let us calculate the track numbers t such that there is $t' \in \{0,4,11\}$ with $(t + t') \pmod{16} = 3$: these are $\{3\} - \{0,4,11\} \pmod{16} = \{3,8,15\}$. Since tracks $3, 15$ are known to be empty, then

$$\begin{aligned}\mathbb{N} + 1 = \textsc{track}_3(X + \{0,4,11\}) = \\ = \textsc{track}_3(X) \cup (\textsc{track}_{15}(X) + 1) \cup (\textsc{track}_8(X) + 1) = \\ = \varnothing \cup \varnothing \cup (\textsc{track}_8(X) + 1) = \textsc{track}_8(X) + 1,\end{aligned}$$

and thus track 8 of X is full. The analogous argument is used to prove that tracks $12, 9, 6$ are full. Consider $\textsc{track}_7(X + \{0,4,11\}) = \mathbb{N} + 1$. Then $\{7\} - \{0,4,11\} \pmod{16} = \{7,3,12\}$.

Since it is already known that tracks $3, 7$ are empty, the track 12 is full:

$$\begin{aligned}\mathbb{N}+1 = \textsc{track}_7(X+\{0,4,11\}) = \\ = \textsc{track}_7(X) \cup \textsc{track}_3(X) \cup (\textsc{track}_{12}(X)+1) = \\ = \varnothing \cup \varnothing \cup (\textsc{track}_{12}(X)+1) = \textsc{track}_{12}(X)+1.\end{aligned}$$

In the same way consider $\textsc{track}_9(X+\{0,4,11\}) = \mathbb{N}$. Then $\{9\} - \{0,4,11\} \pmod{16} = \{9,5,14\}$ and tracks $5, 14$ are empty, thus track 9 is full:

$$\begin{aligned}\mathbb{N} = \textsc{track}_9(X+\{0,4,11\}) = \\ = \textsc{track}_9(X) \cup \textsc{track}_5(X) \cup (\textsc{track}_{14}(X)+1) = \\ = \textsc{track}_9(X) \cup \varnothing \cup \varnothing = \textsc{track}_9(X).\end{aligned}$$

Now let us inspect $\textsc{track}_{10}(X+\{0,4,11\})$. Then $\{10\} - \{0,4,11\} \pmod{16} = \{10,6,15\}$. Since the tracks $10, 15$ are empty, then the 6^{th} track is full:

$$\begin{aligned}\mathbb{N} = \textsc{track}_{10}(X+\{0,4,11\}) = \\ = \textsc{track}_{10}(X) \cup \textsc{track}_6(X) \cup (1+\textsc{track}_{15}(X)) = \\ = \varnothing \cup \textsc{track}_6(X) \cup \varnothing = \textsc{track}_6(X).\end{aligned}$$

Thus it has been proved that $X = \sigma(\textsc{track}_{13}(X))$.

$\Leftarrow$ It remains to show the converse, that is, that if $X = \sigma(\widehat{X})$, then

$$X+\{0,4,11\} = \bigcup_{\substack{i\in\{0,4,6,8,9,\\ 10,12,13\}}} \tau_i(\mathbb{N}) \cup \bigcup_{i\in\{1,3,7\}} \tau_i(\mathbb{N}+1) \cup \{11\}.$$

Since $X = \bigcup_{i=0}^{15} \tau_i(\textsc{track}_i(X))$, then

$$\begin{aligned}X+\{0,4,11\} = \Big(\bigcup_i \tau_i(\textsc{track}_i(X))+0\Big) \cup \Big(\bigcup_i \tau_i(\textsc{track}_i(X))+4\Big) \cup \\ \cup \Big(\bigcup_i \tau_i(\textsc{track}_i(X))+11\Big),\end{aligned}$$

and Table 1 presents the form of each particular term in this union. Each i^{th} row represents track number i in X, and each column labelled $+j$ for $j \in \{0,4,11\}$ corresponds to the addition of a number j. The cell (i,j) gives the set $\textsc{track}_i(X)+j$ and the number of the track in which this set appears in the result (this is track $i+j \pmod{16}$). Then each ℓ^{th} track of $X+\{0,4,11\}$ is obtained as a union of all the appropriate sets in the Table 1.

According to the table, the values of the set $\widehat{X}$ are reflected in three tracks of the sum $X+\{0,4,11\}$: in tracks 13, 1 and 8 (in the last two cases, with offset 1). However, at the same time the sum contains full tracks 8 and 13, as well as $\mathbb{N}+1$ in track 1, and the contributions of $\widehat{X}$ to the sum are subsumed by these numbers, as $\tau_{13}(\widehat{X}) \subseteq \tau_{13}(\mathbb{N})$, $\tau_1(\widehat{X}+1) \subseteq \tau_1(\mathbb{N}+1)$ and $\tau_8(\widehat{X}+1) \subseteq \tau_8(\mathbb{N})$. Therefore, the value of the expression does not depend on $\widehat{X}$. Taking the union of all entries of the Table 1 proves that $X+\{0,4,11\}$ equals

$$\bigcup_{\substack{i\in\{0,4,6,8,9,\\ 10,12,13\}}} \tau_i(\mathbb{N}) \cup \bigcup_{i\in\{1,3,7\}} \tau_i(\mathbb{N}+1) \cup \{11\},$$

	$+0$	$+4$	$+11$
$0\colon \{0\}$	$0\colon \{0\}$	$4\colon \{0\}$	$11\colon \{0\}$
$6\colon \mathbb{N}$	$6\colon \mathbb{N}$	$10\colon \mathbb{N}$	$1\colon \mathbb{N}+1$
$8\colon \mathbb{N}$	$8\colon \mathbb{N}$	$12\colon \mathbb{N}$	$3\colon \mathbb{N}+1$
$9\colon \mathbb{N}$	$9\colon \mathbb{N}$	$13\colon \mathbb{N}$	$4\colon \mathbb{N}+1$
$12\colon \mathbb{N}$	$12\colon \mathbb{N}$	$0\colon \mathbb{N}+1$	$7\colon \mathbb{N}+1$
$13\colon \widehat{X}$	$13\colon \widehat{X}$	$1\colon \widehat{X}+1$	$8\colon \widehat{X}+1$

Table 1: Tracks in the sum $\sigma(\widehat{X})+\{0,4,11\}$, only non-empty tracks of $\sigma(\widehat{X})$ are included.

as stated in the lemma. ∎

4. Simulating operations

The goal of this section is to establish the second property of the encoding σ, that is, that a sum of encodings of two sets and a fixed constant set effectively encodes the union of these two sets, while the addition of a different fixed constant set allows encoding the sum of the two original sets. This property is formally stated in the following lemma, along with the actual constant sets:

Lemma 4.1. *For all sets $X, Y, Z \subseteq \mathbb{N}$,*

$$\sigma(Y)+\sigma(Z)+\{0,1\} = \sigma(X)+\sigma(\{0\})+\{0,1\} \text{ if and only if } Y+Z=X$$

and

$$\sigma(Y)+\sigma(Z)+\{0,2\} = \sigma(X)+\sigma(X)+\{0,2\} \text{ if and only if } Y\cup Z=X.$$

Proof. The goal is to show that for all $Y, Z \subseteq \mathbb{N}$, the sum

$$\sigma(Y)+\sigma(Z)+\{0,1\}$$

encodes the set $Y+Z+1$ on one of its tracks, while the contents of all other tracks do not depend on Y or on Z. Similarly, the sum

$$\sigma(Y)+\sigma(Z)+\{0,2\}$$

has a track that encodes $Y\cup Z$, while the rest of its tracks also do not depend on Y and Z.

The common part of both of the above sums is $\sigma(Y)+\sigma(Z)$, so let us calculate it first. Since

$$\sigma(Y) = \{0\}\cup\tau_6(\mathbb{N})\cup\tau_8(\mathbb{N})\cup\tau_9(\mathbb{N})\cup\tau_{12}(\mathbb{N})\cup\tau_{13}(Y) \quad \text{and}$$
$$\sigma(Z) = \{0\}\cup\tau_6(\mathbb{N})\cup\tau_8(\mathbb{N})\cup\tau_9(\mathbb{N})\cup\tau_{12}(\mathbb{N})\cup\tau_{13}(Z),$$

by the distributivity of union, the sum $\sigma(Y)+\sigma(Z)$ is a union of 36 terms, each being a sum of two individual tracks. Every such sum is contained in a single track as well, and Table 2 gives a case inspection of the form of all these terms. Each of its six rows corresponds to one of the nonempty tracks of $\sigma(Y)$, while its six columns refer to the nonempty tracks in $\sigma(Z)$. Then the cell gives the sum of these tracks, in the form of the track number and track contents: that is, for row representing $\textsc{track}_i(\sigma(Y))$ and for column representing $\textsc{track}_j(\sigma(Z))$, the cell (i,j) represents the set $\textsc{track}_i(\sigma(Y))+\textsc{track}_j(\sigma(Z))$, which is

	0: $\{0\}$	6: $\mathbb{N}$	8: $\mathbb{N}$	9: $\mathbb{N}$	12: $\mathbb{N}$	13: Z
0: $\{0\}$	0: $\{0\}$	6: $\mathbb{N}$	8: $\mathbb{N}$	9: $\mathbb{N}$	12: $\mathbb{N}$	13: Z
6: $\mathbb{N}$	6: $\mathbb{N}$	12: $\mathbb{N}$	14: $\mathbb{N}$	15: $\mathbb{N}$	2: $\mathbb{N}+1$	3: ?
8: $\mathbb{N}$	8: $\mathbb{N}$	14: $\mathbb{N}$	0: $\mathbb{N}+1$	1: $\mathbb{N}+1$	4: $\mathbb{N}+1$	5: ?
9: $\mathbb{N}$	9: $\mathbb{N}$	15: $\mathbb{N}$	1: $\mathbb{N}+1$	2: $\mathbb{N}+1$	5: $\mathbb{N}+1$	6: ?
12: $\mathbb{N}$	12: $\mathbb{N}$	2: $\mathbb{N}+1$	4: $\mathbb{N}+1$	5: $\mathbb{N}+1$	8: $\mathbb{N}+1$	9: ?
13: Y	13: Y	3: ?	5: ?	6: ?	9: ?	10: $(Y+Z)+1$

Table 2: Tracks in the sum $\sigma(Y) + \sigma(Z)$. Question marks denote subsets of $\mathbb{N} + 1$ that depend on Y or Z and whose actual values are unimportant.

	$\sigma(Y)$	$\sigma(Z)$	$\sigma(Y)+\sigma(Z)$	$\sigma(Y)+\sigma(Z)+\{0,1\}$	$\sigma(Y)+\sigma(Z)+\{0,2\}$
0	$\{0\}$	$\{0\}$	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$
1	$\varnothing$	$\varnothing$	$\mathbb{N}+1$	$\mathbb{N}$	$\mathbb{N}+1$
2	$\varnothing$	$\varnothing$	$\mathbb{N}+1$	$\mathbb{N}+1$	$\mathbb{N}$
3	$\varnothing$	$\varnothing$	?	$\mathbb{N}+1$	$\mathbb{N}+1$
4	$\varnothing$	$\varnothing$	$\mathbb{N}+1$	$\mathbb{N}+1$	$\mathbb{N}+1$
5	$\varnothing$	$\varnothing$	$\mathbb{N}+1$	$\mathbb{N}+1$	$\mathbb{N}+1$
6	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$
7	$\varnothing$	$\varnothing$	$\varnothing$	$\mathbb{N}$	$\mathbb{N}+1$
8	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$
9	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$
10	$\varnothing$	$\varnothing$	$Y+Z+1$	$\mathbb{N}$	$\mathbb{N}$
11	$\varnothing$	$\varnothing$	$\varnothing$	$Y+Z+1$	$\mathbb{N}$
12	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$
13	Y	Z	$Y \cup Z$	$\mathbb{N}$	$Y \cup Z$
14	$\varnothing$	$\varnothing$	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$
15	$\varnothing$	$\varnothing$	$\mathbb{N}$	$\mathbb{N}$	$\mathbb{N}$

Table 3: Tracks in the sums of $\sigma(Y) + \sigma(Z)$ with constants.

bound to be on track $i + j \pmod{16}$. For example, the sum of track 8 of $\sigma(Y)$ and track 9 of $\sigma(Z)$ falls onto track $1 = 8 + 9 \pmod{16}$ and equals

$$\tau_8(\mathbb{N}) + \tau_9(\mathbb{N}) = \{8 + 9 + 16(m + n) \mid m, n \geqslant 0\} = \{1 + 16n \mid n \geqslant 1\} = \tau_1(\mathbb{N} + 1),$$

while adding track 13 of $\sigma(Y)$ to track 13 of $\sigma(Z)$ results in

$$\tau_{13}(Y) + \tau_{13}(Z) = \{26 + 16(m + n) \mid m \in Y,\ n \in Z\} = \tau_{10}(Y + Z + 1),$$

which is reflected in the table. Each question mark denotes a track with unspecified contents. Though these contents can be calculated, it is actually irrelevant, because it does not influence the value of the subsequent sums $\sigma(Y) + \sigma(Z) + \{0, 1\}$ and $\sigma(Y) + \sigma(Z) + \{0, 2\}$. What is important is that none of these tracks contain 0.

Now the value of each i^{th} track of $\sigma(Y) + \sigma(Z)$ is obtained as the union of all sums in Table 2 that belong to the i^{th} track. The final values of these tracks are presented in the corresponding column of Table 3.

Now the contents of the tracks in $\sigma(Y) + \sigma(Z) + \{0,1\}$ can be completely described. The calculations are given in Table 3, and the result is that for all Y and Z,

$$\begin{aligned}
&\textsc{track}_{11}(\sigma(Y) + \sigma(Z) + \{0,1\}) = Y + Z + 1,\\
&\textsc{track}_i(\sigma(Y) + \sigma(Z) + \{0,1\}) = \mathbb{N} + 1 && \text{for } i \in \{2,3,4,5\},\\
&\textsc{track}_i(\sigma(Y) + \sigma(Z) + \{0,1\}) = \mathbb{N} && \text{for all other } i.
\end{aligned}$$

It easily follows that

$$X = Y + Z$$

if and only if

$$\sigma(X) + \sigma(\{0\}) + \{0,1\} = \sigma(Y) + \sigma(Z) + \{0,1\},$$

as, clearly, $X = X + \{0\}$.

For the set $\sigma(Y) + \sigma(Z) + \{0,2\}$, in the same way, for all Y and Z,

$$\begin{aligned}
&\textsc{track}_{13}(\sigma(Y) + \sigma(Z) + \{0,2\}) = Y \cup Z,\\
&\textsc{track}_j(\sigma(Y) + \sigma(Z) + \{0,2\}) = \mathbb{N} + 1 && \text{for } j \in \{1,3,4,5,7\},\\
&\textsc{track}_j(\sigma(Y) + \sigma(Z) + \{0,2\}) = \mathbb{N} && \text{for all other } j,
\end{aligned}$$

and therefore for all X, Y, Z,

$$X = Y \cup Z$$

if and only if

$$\sigma(X) + \sigma(X) + \{0,2\} = \sigma(Y) + \sigma(Z) + \{0,2\},$$

since $X = X \cup X$.

Both claims of the lemma follow. ∎

5. Simulating a system of equations

Using the encoding defined above, it is now possible to represent a system with union and addition by a system with addition only. Since Lemma 4.1 on the simulation of individual operations is applicable only to equations of a simple form, the first task is to convert a given system to such a form:

Lemma 5.1. *For every system of equations over sets of numbers in variables $(X_1, \ldots, X_n)$ using union, addition and constants from a set $\mathcal{C}$ there exists a system in variables $(X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+m})$ with all equations of the form $X_i = X_j + X_k$, $X_i = X_j \cup X_k$ or $X_i = C$ with $C \in \mathcal{C}$, such that the set of solutions of this system is*

$$\big\{(S_1, \ldots, S_n, \ldots, f_i(S_1, \ldots, S_n), \ldots) \;\big|\; (S_1, \ldots, S_n) \text{ is a solution of the original system}\big\},$$

for some monotone functions $f_1, \ldots, f_m$.

The construction is by a straightforward decomposition of equations, with new variables representing subexpressions of the sides of the original equations. Once the equations are thus transformed, the system can be encoded as follows.

Lemma 5.2. *For every system of equations over sets of numbers in variables $(\dots, X, \dots)$ and with all equations of the form $X = Y + Z$, $X = Y \cup Z$ or $X = C$, there exists a system in variables $(\dots, X', \dots)$, using only addition and constants $\{0,1\}$, $\{0,2\}$, $\{0,4,11\}$, $\sigma(\{0\})$, $\sigma(C)$ with C used in the original system and the ultimately periodic constant from Lemma 3.3, such that $(\dots, S'_X, \dots)$ is a solution of the latter system if and only if $S'_X = \sigma(S_X)$ for each variable X, for some solution $(\dots, S_X, \dots)$ of the former system.*

Proof. The proof is by a direct transformation of this system according to Lemmata 3.3 and 4.1. First, the new system contains the following equation for each variable X':

$$X' + \{0,4,11\} = \bigcup_{\substack{i \in \{0,4,6,8,9,\\ 10,12,13\}}} \tau_i(\mathbb{N}) \cup \bigcup_{i \in \{1,3,7\}} \tau_i(\mathbb{N}+1) \cup \{11\}. \tag{5.1}$$

Next, for each equation $X = Y + Z$ in the original system, there is a corresponding equation

$$X' + \sigma(\{0\}) + \{0,1\} = Y' + Z' + \{0,1\} \tag{5.2}$$

in the new system. Similarly, for each equation of the form $X = Y \cup Z$, the new system contains an equation

$$X' + X' + \{0,2\} = Y' + Z' + \{0,2\}. \tag{5.3}$$

Finally, every equation $X = C$ in the original system is represented in the new system by the following equation:

$$X' = \sigma(C). \tag{5.4}$$

By Lemma 3.3, (5.1) ensures that each solution $(\dots, S'_X, \dots)$ of the constructed system satisfies $S'_X = \sigma(S_X)$ for some sets S_X. It is claimed that $(\dots, S_X, \dots)$ satisfies each equation of the original system if and only if $(\dots, \sigma(S_X), \dots)$ satisfies the corresponding equation (5.2–5.4) of the constructed system. Consider each pair of corresponding equations:

- Consider an equation $X = Y \cup Z$ from the original system. Then there is a corresponding equation (5.2), and, by Lemma 4.1, $(\dots, S_X, \dots)$ satisfies the original equation if and only if $(\dots, \sigma(S_X), \dots)$ satisfies (5.2).
- Similarly, by Lemma 4.1, an equation of the form $X = Y + Z$ is satisfied by $(\dots, S_X, \dots)$ if and only if $(\dots, \sigma(S_X), \dots)$ satisfies the corresponding equation (5.3).
- For each equation of the form $X = C$ it is claimed that a set S_X satisfies it if and only if $\sigma(S_X)$ satisfies the corresponding equation (5.4). Indeed, $\sigma(S_X) = \sigma(C)$ if and only if $\textsc{track}_{13}(\sigma(S_X)) = \textsc{track}_{13}(\sigma(C))$, and since $\textsc{track}_{13}(\sigma(S_X)) = S_X$ and $\textsc{track}_{13}(\sigma(C)) = C$, this is equivalent to $S_X = C$.

This shows that $(\dots, S_X, \dots)$ satisfies the original system if and only if $(\dots, \sigma(S_X), \dots)$ satisfies the constructed system, which proves the correctness of the construction. ∎

Note that σ is a bijection between the sets of solutions of the two systems. Then, in particular, if the original system has a unique solution, then the constructed system has a unique solution as well, which encodes the solution of the original system.

Furthermore, it is important that the encoding σ respects inclusion, that is, if $X \subseteq Y$, then $\sigma(X) \subseteq \sigma(Y)$. Consider the partial order on solutions of a system, defined as $(S_1, \dots, S_n) \preccurlyeq (S'_1, \dots, S'_n)$ if $S_i \subseteq S'_i$ for all i. Now if one solution of the original system is less than another, then the corresponding solutions of the constructed system maintain this relation. Therefore, if the original system has a least (greatest) solution with respect to this partial order, then so does the new one, and its least (greatest) solution is the image of the least (greatest) solution of the original system.

These observations allow applying Lemmata 5.1 and 5.2 to encode each system in Theorem 2.4 within a system using addition only.

Theorem 5.3. *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 the operation of addition and ultimately periodic constants, which has a unique (least, greatest, respectively) solution with $X_1 = T$, where $S = \{n \mid 16n + 13 \in T\}$.

Note that S is computationally reducible to T via the "$16n + 13$" transduction, hence the following statements:

Corollary 5.4. *For every recursive set $S \subseteq \mathbb{N}$ there exists a system of equations over sets of natural numbers using addition and ultimately periodic constants that has a unique solution, which is computationally as hard as S.*

Corollary 5.5. *There exists a system of equations over sets of natural numbers using addition and ultimately periodic constants which has a least (greatest) solution with its first component being r.e.-complete (co-r.e.-complete, respectively).*

Finally, the decision problems for these systems of equations turn out to be as hard as in the case of union and addition:

Theorem 5.6. *The problem of testing whether a system of equations over sets of natural numbers using addition and ultimately periodic constants has a solution is Π_1-complete. The problem of whether it has a unique, least or greatest solution is Π_2-complete. The problem of whether it has finitely many solutions is Σ_3-complete.*

The above results equally apply to *language equations* over a one-letter alphabet with concatenation as the only allowed operation and with regular constants.

6. Systems with finite constants

The constructions above essentially use three infinite ultimately periodic constants: one of them is the right-hand side of the equation from Lemma 3.3, and the other two are the sets $\sigma(\{0\})$ and $\sigma(\{1\})$ used in Lemma 5.2 to represent constants $\{0\}$ and $\{1\}$. It will now be shown that the use of such constants is necessary, and systems using only addition and *finite* constants cannot specify any non-trivial infinite sets.

This is done by demonstrating that every solution $(\ldots, S, \ldots)$ of such a system can be *pruned* in the sense that each of its infinite components can be replaced by an empty set and the resulting vector remains a solution.

Lemma 6.1. *If a system of equations in variables $(\ldots, X_j, \ldots, Y_i, \ldots)$ using addition and only finite constants has a solution $(\ldots, F_j, \ldots, S_i, \ldots)$, where each F_j is finite and each S_i infinite, then $(\ldots, F_j, \ldots, \varnothing, \ldots)$ is a solution of this system.*

In a similar way, infinite components of a solution can be augmented to co-finite sets. For every nonempty set $S \subseteq \mathbb{N}$, consider its *upward closure* $S + \mathbb{N}$, which is always co-finite.

Lemma 6.2. *If a system of equations in variables $(\ldots, X_j, \ldots, Y_i, \ldots)$ using addition and only finite constants has a solution $(\ldots, F_j, \ldots, S_i, \ldots)$, where each F_j is finite and each S_i infinite, then $(\ldots, F_j, \ldots, S_i + \mathbb{N}, \ldots)$ is a solution as well.*

Theorem 6.3. *If a system of equations using addition and finite constants has a least (greatest, unique) solution $(\ldots, S_i, \ldots)$, then each S_i is finite (finite or co-finite, finite, respectively).*

Since equations with finite constants have so trivial solutions, it is natural to expect their decision problems to be much easier than in Theorem 5.6. Establishing the exact complexity of these problems is left for future work.

7. Conclusion

The study of language equations has progressed by showing the computational universality of simpler and simpler models [10, 6, 5]. The equations proved universal in this paper are the simplest considered so far: the constructions use systems of equations $X = YZ$ and $X = C$ over an alphabet $\Sigma = \{a\}$, with ultimately periodic constants $C \subseteq a^*$. Little room is left for further improvement, as infinite constants were proved to be essential.

The results have been obtained in terms of equations over sets of numbers using the operation of addition, which is the main subject of additive combinatorics [14]. Hopefully, this work will lead to some further connections between computability and number theory.

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", *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, "Complexity of solutions of equations over sets of natural numbers", *STACS 2008*.
[5] A. Jeż, A. Okhotin, "On the computational completeness of equations over sets of natural numbers", *ICALP 2008*, part II, LNCS 5126, 63–74.
[6] M. Kunc, "The power of commuting with finite sets of words", *Theory of Computing Systems*, 40:4 (2007), 521–551.
[7] E. L. Leiss, "Unrestricted complementation in language equations over a one-letter alphabet", *Theoretical Computer Science*, 132 (1994), 71–93.
[8] P. McKenzie, K. Wagner, "The complexity of membership problems for circuits over sets of natural numbers", *Computational Complexity*, 16:3 (2007), 211–244.
[9] A. Okhotin, "Conjunctive grammars", *Journal of Automata, Languages and Combinatorics*, 6:4 (2001), 519–535.
[10] A. Okhotin, "Decision problems for language equations with Boolean operations", *ICALP 2003*.
[11] A. Okhotin, "Unresolved systems of language equations: expressive power and decision problems", *Theoretical Computer Science*, 349:3 (2005), 283–308.
[12] A. Okhotin, O. Yakimova, "On language equations with complementation", *DLT 2006*, 420–432.
[13] L. J. Stockmeyer, A. R. Meyer, "Word problems requiring exponential time", *STOC 1973*, 1–9.
[14] T. Tao, V. Vu, *Additive Combinatorics*, Cambridge University Press, 2006.