On equations over sets of integers

Artur Jez, Alexander Okhotin

Symposium on Theoretical Aspects of Computer Science 2010 (Nancy, France), pp. 477-488
www.stacs-conf.org

ON EQUATIONS OVER SETS OF INTEGERS

ARTUR JEŻ [1] AND ALEXANDER OKHOTIN [2,3]

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

[2] Academy of Finland

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

Abstract. Systems of equations with sets of integers as unknowns are considered. It is shown that the class of sets representable by unique solutions of equations using the operations of union and addition $S+T=\{m+n \mid m \in S,\, n \in T\}$ and with ultimately periodic constants is exactly the class of hyper-arithmetical sets. Equations using addition only can represent every hyper-arithmetical set under a simple encoding. All hyper-arithmetical sets can also be represented by equations over sets of natural numbers equipped with union, addition and subtraction $S \dot{-} T=\{m-n \mid m \in S,\, n \in T,\, m \geqslant n\}$. Testing whether a given system has a solution is Σ^1_1-complete for each model. These results, in particular, settle the expressive power of the most general types of language equations, as well as equations over subsets of free groups.

1. Introduction

Language equations are equations with formal languages as unknowns. The simplest such equations are the context-free grammars [4], as well as their generalization, the conjunctive grammars [15]. Many other types of language equations have been studied in the recent years, see a survey by Kunc [11], and most of them were found to have strong connections to computability. In particular, for equations with concatenation and Boolean operations it was shown by Okhotin [19, 17] that the class of languages representable by their unique (least, greatest) solutions is exactly the class of recursive (r.e., co-r.e.) sets. A computationally universal equation of the simplest form was constructed by Kunc [10], who proved that the greatest solution of the equation $XL = LX$, where $L \subseteq \{a,b\}^*$ is a finite constant language, may be co-r.e.-complete.

A seemingly trivial case of language equations over a *unary alphabet* $\Omega = \{a\}$ has recently been studied. Strings over such an alphabet may be regarded as natural numbers,

1998 ACM Subject Classification: F.4.3 (Formal languages), F.4.1 (Mathematical logic).

Key words and phrases: Language equations, computability, arithmetical hierarchy, hyper-arithmetical hierarchy.

Research supported by the Polish Ministry of Science and Higher Education under grants N N206 259035 2008–2010 and N N206 492638 2010–2012, and by the Academy of Finland under grant 134860.

and languages accordingly become sets of numbers. As established by the authors [8], these equations are as powerful as language equations over a general alphabet: a set of natural numbers is representable by a unique solution of a system with union and elementwise addition if and only if it is recursive. Furthermore, even without the union operation these equations remain almost as powerful [9]: for every recursive set $S \subseteq \mathbb{N}$, its encoding $\sigma(S) \subseteq \mathbb{N}$ satisfying $S = \{n \mid 16n + 13 \in \sigma(S)\}$ can be represented by a unique solution of a system using addition only, as well as ultimately periodic constants. At the same time, as shown by Lehtinen and Okhotin [12], some recursive sets are not representable without an encoding.

Equations over sets of numbers are, on one hand, interesting on their own as a basic mathematical object. On the other hand, these equations form a very special case of language equations with concatenation and Boolean operations, which turned out to be as hard as the general case, and this is essential for understanding language equations. However, it must be noted that these cases do not exhaust all possible language equations. The recursive upper bound on unique solutions [19] is applicable only to equations with *continuous* operations on languages, and using the simplest non-continuous operations, such as homomorphisms or quotient [18], leads out of the class of recursive languages. In particular, a quotient with regular constants was used to represent all sets in the arithmetical hierarchy [18].

The task is to find a natural limit of the expressive power of language equations, which would not assume continuity of operations. As long as operations on languages are expressible in first-order arithmetic (which is true for every common operation), it is not hard to see that unique solutions of equations with these operations always belong to the family of *hyper-arithmetical sets* [14, 20, 21]. This paper shows that this obvious upper bound is in fact reached already in the case of a unary alphabet.

To demonstrate this, two abstract models dealing with sets of numbers shall be introduced. The first model are equations over sets of natural numbers with addition $S + T = \{m + n \mid m \in S,\ n \in T\}$ and subtraction $S \dot{-} T = \{m - n \mid m \in S,\ n \in T,\ m \geqslant n\}$ (corresponding to concatenation and quotient of unary languages), as well as set-theoretic union. The other model has sets of integers, including negative numbers, as unknowns, and the allowed operations are addition and union. The main result of this paper is that unique solutions of systems of either kind can represent every *hyper-arithmetical* set of numbers.

The base of the construction is the authors' earlier result [8] on representing every recursive set by equations over sets of natural numbers with union and addition. In Section 2, this result is adapted to the new models introduced in this paper. The next task is representing every set in the arithmetical hierarchy, which is achieved in Section 3 by simulating existential and universal quantifiers over a recursive set. These arithmetical sets are then used in Section 4 as constants for the construction of equations representing hyper-arithmetical sets. Finally, the constructed equations are encoded in Section 5 using equations over sets of integers with addition only and periodic constant sets.

This result brings to mind a study by Robinson [20], who considered equations, in which the unknowns are functions from $\mathbb{N}$ to $\mathbb{N}$, the only constant is the successor function and the only operation is superposition, and proved that a function is representable by a unique solution of such an equation if and only if it is hyper-arithmetical. Though these equations deal with objects different from sets of numbers, there is one essential thing in common: in both results, unique solutions of equations over second-order arithmetical objects represent hyper-arithmetical sets.

Some more related work can be mentioned. Halpern [5] studied the decision problem of whether a formula of Presburger arithmetic with set variables is true for all values of these set variables, and showed that it is Π^1_1-complete. The equations studied in this paper can be regarded as a small fragment of Presburger arithmetic with set variables.

Another relevant model are languages over free groups, which have been investigated, in particular, by Anisimov [3] and by d'Alessandro and Sakarovitch [2]. Equations over sets of integers are essentially equations for languages over a monogenic free group.

An important special case of equations over sets of numbers are *expressions* and *circuits* over sets of numbers, which are equations without iterated dependencies. Expressions and circuits over sets of natural numbers were studied by McKenzie and Wagner [13], and a variant of these models defined over sets of integers was investigated by Travers [22].

2. Equations and their basic expressive power

The subject of this paper are systems of equations of the 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 $X_i \subseteq \mathbb{Z}$ are unknown sets of integers, and the expressions φ_i and ψ_i use such operations as union, intersection, complementation, as well as the main arithmetical operation of elementwise addition of sets, defined as $S+T=\{m+n \mid m\in S,\, n\in T\}$. Subtraction $S-T=\{m-n \mid m\in S,\, n\in T\}$ shall be occasionally used. The constant sets contained in a system sometimes will be singletons only, sometimes any ultimately periodic constants will be allowed (a set of integers $S\subseteq\mathbb{Z}$ 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 all n with $|n|\geqslant d$), and in some cases the constants will be drawn from wider classes of sets, such as all recursive sets. Systems over sets of natural numbers shall have subsets of $\mathbb{N}$ both as unknowns and as constant languages; whenever subtraction is used in such equations, it will be used in the form $S \dot{-} T = (S-T)\cap\mathbb{N}$.

Consider systems with a unique solution. Every such system can be regarded as a specification of a set, and for every type of systems there is a natural question of what kind of sets can be represented by unique solutions of these systems. For equations over sets of natural numbers, these are the recursive sets:

Proposition 1 (Jeż, Okhotin [8, Thm. 4])**.** *The family of sets of natural numbers representable by unique solutions of systems of equations of the form $\varphi_i(X_1,\ldots,X_n) = \psi_i(X_1,\ldots,X_n)$ with union, addition and singleton constants, is exactly the family of recursive sets.*

Turning to the more general cases of equations over sets of integers and of equations over sets of natural numbers with subtraction, an upper bound on their expressive power can be obtained by reformulating a given system in the notation of first-order arithmetic.

Lemma 1. *For every system of equations in variables $X_1,\ldots X_n$ using operations expressible in first-order arithmetic there exists an arithmetical formula $Eq(X_1,\ldots,X_n)$, where $X_1,\ldots,X_n$ are free second-order variables, such that $Eq(S_1,\ldots,S_n)$ is true if and only if $X_i=S_i$ is a solution of the system.*

Constructing this formula is only a matter of reformulation. As an example, an equation $X_i = X_j + X_k$ is represented by $(\forall n)\big[n \in X_i \leftrightarrow (\exists n')(\exists n'')n = n' + n'' \wedge n' \in X_j \wedge n'' \in X_k\big]$.

Now consider the following formulae of second-order arithmetic:

$$\varphi(x) = (\exists X_1)\dots(\exists X_n)Eq(X_1,\dots,X_n) \wedge x \in X_1$$
$$\varphi'(x) = (\forall X_1)\dots(\forall X_n)Eq(X_1,\dots,X_n) \rightarrow x \in X_1$$

The formula $\varphi(x)$ represents the membership of x in *any* solution of the system, while $\varphi'(x)$ states that *every* solution of the system contains x. Since, by assumption, the system has a unique solution, these two formulae are equivalent and each of them specifies the first component of this solution. Furthermore, φ and φ' belong to the classes Σ^1_1 and Π^1_1, respectively, and accordingly the solution belongs to the class $\Delta^1_1 = \Sigma^1_1 \cap \Pi^1_1$, known as the class of *hyper-arithmetical sets* [14, 21].

Lemma 2. *For every system of equations in variables $X_1,\dots X_n$ using operations and constants expressible in first-order arithmetic that has a unique solution $X_i = S_i$, the sets S_i are hyper-arithmetical.*

Though this looks like a very rough upper bound, this paper actually establishes the converse, that is, that every hyper-arithmetical set is representable by a unique solution of such equations. The result shall apply to equations of two kinds: over sets of integers with union and addition, and over sets of natural numbers with union, addition and subtraction. In order to establish the properties of both families of equations within a single construction, the next lemma introduces a general form of systems that can be converted to either of the target types of systems:

Lemma 3. *Consider any system of equations $\varphi(X_1,\dots,X_m) = \psi(X_1,\dots,X_m)$ and inequalities $\varphi(X_1,\dots,X_m) \subseteq \psi(X_1,\dots,X_m)$ over sets of natural numbers that uses the following operations: union; addition of a recursive constant; subtraction of a recursive constant; intersection with a recursive constant. Assume that the system has a unique solution $X_i = S_i \subseteq \mathbb{N}$. Then there exist:*

1. *a system of equations over sets of natural numbers in variables $X_1,\dots,X_m,Y_1,\dots,Y_{m'}$ using the operations of addition, subtraction and union and singleton constants, which has a unique solution with $X_i = S_i$;*
2. *a system of equations over sets of integers in variables $X_1,\dots,X_m,Y_1,\dots,Y_{m'}$ using the operations of addition and union, singleton constants and the constants $\mathbb{N}$ and $-\mathbb{N}$, which has a unique solution with $X_i = S_i$.*

Inequalities $\varphi \subseteq \psi$ can be simulated by equations $\varphi \cup \psi = \psi$. For equations over sets of natural numbers, each recursive constant is represented according to Proposition 1, and this is sufficient to implement each addition or subtraction of a recursive constant by a large subsystem using only singleton constants. In order to obtain a system over sets of integers, a straightforward adaptation of Proposition 1 is needed:

Lemma 3.1. *For every recursive set $S \subseteq \mathbb{N}$ there exists a system of equations over sets of integers in variables $X_1,\dots,X_n$ using union, addition, singleton constants and constant $\mathbb{N}$, such that the system has a unique solution with $X_1 = S$.*

This is essentially the system given by Proposition 1, with additional equations $X_i \subseteq \mathbb{N}$.

Now a difference $X \dot{-} R$ for a recursive constant $R \subseteq \mathbb{N}$ shall be represented as $(X + (-R)) \cap \mathbb{N}$, where the set $-R = \{-n \mid n \in R\}$ is specified by taking a system for R and applying the following transformation:

Lemma 3.2 (Representing sets of opposite numbers)**.** *Consider a system of equations over sets of integers, in variables $X_1, \ldots, X_n$, using union and addition, and any constant sets, which has a unique solution $X_i = S_i$. Then the same system, with each constant $C \subseteq \mathbb{Z}$ replaced by the set of the opposite numbers $-C$, has the unique solution $X_i = -S_i$.*

The last step in the proof of Lemma 3 is eliminating intersection with recursive constants. This is done as follows:

Lemma 3.3 (Intersection with constants)**.** *Let $R \subseteq \mathbb{N}$ be a recursive set. Then there exists a system of equations over sets of natural numbers using union, addition and singleton constants, which has variables $X, Y, Y', Z_1, \ldots, Z_m$, such that the set of solutions of this system is*

$$\big\{\ (X = S,\ Y = S \cap R,\ Y' = S \cap \overline{R},\ Z_i = S_i) \ \big|\ S \subseteq \mathbb{N}\ \big\},$$

where $S_1, \ldots, S_m$ are some fixed sets.

In plain words, the constructed system works as if an equation $Y = X \cap R$ (and also another equation $Y' = X \cap \overline{R}$, which may be ignored). This completes the transformations needed for Lemma 3.

The last basic element of the construction is representing a set of integers (both positive and negative) by first representing its positive and negative subsets individually:

Lemma 4 (Assembling positive and negative subsets)**.** *Let sets $S \cap \mathbb{N}$ and $(-S) \cap \mathbb{N}$ be representable by unique solutions of equations over sets of integers using union, addition, and ultimately periodic constants. Then S is representable by equations over integers using only union, addition and ultimately periodic constants.*

3. Representing the arithmetical hierarchy

Each arithmetical set can be represented by a recursive relation with a quantifier prefix, and arithmetical sets form the *arithmetical hierarchy* based on the number of quantifier alternations in such a formula. The bottom of the hierarchy are the recursive sets, and every next level is comprised of two classes, Σ^0_k or Π^0_k, which correspond to the cases of the first quantifier's being existential or universal. For every $k \geqslant 1$, a set is in Σ^0_k if it can be represented as

$$\{w \mid \exists x_1 \forall x_2 \ldots Q_k x_k\ R(w, x_1, \ldots, x_k)\}$$

for some recursive relation R, where $Q_k = \forall$ if k is even and $Q_k = \exists$ if k is odd. A set is in Π^0_k if it admits a similar representation with the quantifier prefix $\forall x_1 \exists x_2 \ldots Q_k x_k$. It is easy to see that $\Pi^0_k = \{L \mid \overline{L} \in \Sigma^0_k\}$. The sets Σ^0_1 and Π^0_1 are the recursively enumerable sets and their complements, respectively. The arithmetical hierarchy is known to be strict: $\Sigma^0_k \subset \Sigma^0_{k+1}$ and $\Pi^0_k \subset \Pi^0_{k+1}$ for every $k \geqslant 0$. Furthermore, for every $k \geqslant 1$ the inclusion $\Sigma^0_k \cup \Pi^0_k \subset \Sigma^0_{k+1} \cap \Pi^0_{k+1}$ is proper, i.e., there is a gap between the k-th and $(k+1)$-th level.

For this paper, the definition of arithmetical sets shall be arithmetized in base-7 notation[1] as follows: a set $S \subseteq \mathbb{N}$ is in Σ^0_k if it is representable as

$$S = \{\ (w)_7 \mid \exists x_1 \in \{\mathtt{3},\mathtt{6}\}^* \forall x_2 \in \{\mathtt{3},\mathtt{6}\}^* \ldots Q_k x_k \in \{\mathtt{3},\mathtt{6}\}^* (\mathtt{1} x_1 \mathtt{1} y_1 \mathtt{1} \ldots x_k \mathtt{1} y_k \mathtt{1} w)_7 \in R\},$$

for some recursive set $R \subseteq \mathbb{N}$, where $(w)_7$ for $w \in \{\mathtt{0},\mathtt{1},\ldots,\mathtt{6}\}^*$ denotes the natural number with base-7 notation w. The strings $x_i \in \{\mathtt{3},\mathtt{6}\}^*$ represent *binary* notation of some numbers,

[1]Base 7 is the smallest base, for which the details of the constructions could be conveniently implemented.

where 3 stands for zero and 6 stands for one. The notation $(x)_2$ for $x \in \{3,6\}^*$ shall be used to denote the number represented by this encoding. The digits 1 act as separators. Throughout this paper, the set of base-7 digits $\{0,1,\ldots,6\}$ shall be denoted by Ω_7.

In general, the construction of a system of equations representing the set S begins with representing R, and proceeds with evaluating the quantifiers, eliminating the prefixes $1x_1$, $1x_2$, and so on until $1x_k$. In the end, all numbers $(1w)_7$ with $(w)_7 \in S$ will be produced. These manipulations can be expressed in terms of the following three functions:

$$\begin{aligned} \mathit{Remove}_1(X) &= \{(w)_7 \mid (1w)_7 \in X\}, \\ E(X) &= \{(1w)_7 \mid \exists x \in \{3,6\}^* : \ (x1w)_7 \in X\}, \\ A(X) &= \{(1w)_7 \mid \forall x \in \{3,6\}^* : \ (x1w)_7 \in X\}. \end{aligned}$$

The expression converting numbers of the form $(1w)_7$ to $(w)_7$ is constructed as follows:

Lemma 5 (Removing leading digit 1)**.** *The value of the expression*

$$(X - \{1\} \cap \{0\}) \cup \bigcup_{i \in \Omega_7 \setminus \{0\}} \bigcup_{t \in \{0,1\}} \left[(X \cap (1i\Omega_7^t(\Omega_7^2)^*)_7) \dot{-} (10^*)_7\right] \cap (i\Omega_7^t(\Omega_7^2)^*)_7 \tag{3.1}$$

on any $S \subseteq (1(\Omega_7^* \setminus 0\Omega_7^*))_7$ *is* $\{(w)_7 \mid (1w)_7 \in S\}$*. The value on* $S \subseteq (10\Omega_7^*)_7$ *equals* $\varnothing$*.*

With Lemma 5 established and the expression (3.1) proved to implement the function $\mathit{Remove}_1(X)$, the notation $\mathit{Remove}_1(X)$ is used in equations to refer to this subexpression.

Next, consider the function $E(X)$ representing the existential quantifier ranging over strings in $\{3,6\}^*$. This function can be implemented by a single expression as follows:

Lemma E (Representing the existential quantifier)**.** *The value of the expression*

$$(X \cap (1\Omega_7^*)_7) \cup \left(\left[(X \cap (\{3,6\}^+1\Omega_7^*)_7) \dot{-} (\{3,6\}^+0^*)_7\right] \cap (1\Omega_7^*)_7\right)$$

on any $S \subseteq (\{3,6\}^*1\Omega_7^*)_7$ *is* $E(S) = \{(1w)_7 \mid \exists w' \in \{3,6\}^*(w'1w)_7 \in S\}$*.*

Note that $E(X)$ can already produce any recursively enumerable set from a recursive argument, and therefore it is essential to use subtraction in the expression.

With the existential quantifier implemented, the next task is to represent a universal quantifier. Ideally, one would be looking for an expression implementing $A(X)$, but, unfortunately, no such expression was found, and the actual construction given below implements the universal quantifier using multiple equations. The first step is devising an equation representing the function $f(X) = \{(x1w)_7 \mid x \in \{3,6\}^*, \ (1w)_7 \in X\}$, which appends every string of digits in $\{3,6\}^*$ to numbers in its argument set.

Lemma 6. *For every constant set* $X \subseteq (1\Omega_7^*)_7$*, the equation*

$$Y = X \cup \mathit{Append}_{3,6}(Y), \quad \textit{where}$$

$$\begin{aligned} \mathit{Append}_{3,6}(Y) = &\bigcup_{i,j \in \{3,6\}} \left[\left(\left[(Y \cap (j\Omega_7^*)_7) + (20^*)_7\right] \cap (2j\Omega_7^*)_7\right) + ((i-2)0^*)_7\right] \cap (ij\Omega_7^*)_7 \\ &\cup \bigcup_{i \in \{3,6\}} \left[(Y \cap (1\Omega_7^*)_7) + (i0^*)_7\right] \cap (i1\Omega_7^*)_7 \end{aligned}$$

has the unique solution $Y = \{(x1w)_7 \mid x \in \{3,6\}^*, (1w)_7 \in X\}$*.*

Lemma A (Representing the universal quantifier)**.** *Let $S, \widetilde{S} \subseteq (\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}\Omega_7^*)_7$ be any sets, such that $\widetilde{S} \cap S = \varnothing$ and for $x', x \in \{\mathtt{3},\mathtt{6}\}^*$ $(x\mathtt{1}w)_7 \in S$ and $(x'\mathtt{1}w)_7 \notin S$ implies $(x'\mathtt{1}w)_7 \in \widetilde{S}$. Then the following system of equations over sets of integers in variables Y, $\widetilde{Y}$ and Z*

$$\begin{aligned} Y &= Z \cup \mathit{Append}_{\mathtt{3},\mathtt{6}}(Y) \\ \widetilde{Y} &= E(\widetilde{S}) \cup \mathit{Append}_{\mathtt{3},\mathtt{6}}(\widetilde{Y}) \\ Z &\subseteq (\mathtt{1}\Omega_7^+)_7 \\ Y \subseteq S &\subseteq Y \cup \widetilde{Y}, \end{aligned}$$

has the unique solution $Z = A(S) = \{(\mathtt{1}w)_7 \,|\, \forall x \in \{\mathtt{3},\mathtt{6}\}^* : (x\mathtt{1}w)_7 \in S\}$, $Y = \{(y\mathtt{1}w)_7 \,|\, y \in \{\mathtt{3},\mathtt{6}\}^*, \forall x \in \{\mathtt{3},\mathtt{6}\}^* : (x\mathtt{1}w)_7 \in S\}$, $\widetilde{Y} = \{(y\mathtt{1}w)_7 \,|\, y \in \{\mathtt{3},\mathtt{6}\}^*, \exists x \in \{\mathtt{3},\mathtt{6}\}^* : (x\mathtt{1}w)_7 \in \widetilde{S}\}$.

Once the above quantifiers process a number $(\mathtt{1}x_k\mathtt{1}x_{k-1}\ldots\mathtt{1}x_1\mathtt{1}w)_7$, reducing it to $(\mathtt{1}w)_7$, the actual number $(w)_7$ is obtained from this encoding by Lemma 5.

Theorem 1. *Every arithmetical set $S \subseteq \mathbb{Z}$ ($S \subseteq \mathbb{N}$) is representable as a component of a unique solution of a system of equations over sets of integers (sets of natural numbers, respectively) with φ_j, ψ_j using the operations of addition and union and ultimately periodic constants (addition, subtraction, union and singleton constants, respectively).*

4. Representing hyper-arithmetical sets

Following Moschovakis [14, Sec. 8E] and Aczel [1, Thm. 2.2.3], *hyper-arithmetical* sets $B_1, B_2, \ldots$ shall be defined as the *smallest effective σ-ring*, which is the recursion-theoretic counterpart to Borel sets (the smallest family of sets containing all open sets and closed under countable union and countable intersection).

Let $f_1, f_2, \ldots$ be an enumeration of all partial recursive functions and let τ_1, τ_2 be two recursive functions. Then, for all $k \in \mathbb{N}$,

$$B_{\tau_1(k)} = \mathbb{N} \setminus \{k\}, \quad C_{\tau_1(k)} = \{k\}$$

Moreover, for all numbers $k \in \mathbb{N}$, if f_k is a total function, then

$$B_{\tau_2(k)} = \bigcup_{n\in\mathbb{N}} C_{f_k(n)}, \quad C_{\tau_2(k)} = \bigcap_{n\in\mathbb{N}} B_{f_k(n)},$$

where the former operation is known as *effective σ-union*, while the latter is *effective σ-intersection*. Note that the only distinction between B_e and C_e is that the former is defined as a union and the latter as an intersection. As the definitions are dual, $B_e = \overline{C_e}$.

The family of sets $\mathcal{B} = \{B_e, C_e \mid e \in I\}$, where $I \subseteq \mathbb{N}$ is an index set, is called an effective σ-ring, if it contains $\{B_{\tau_1(e)}, C_{\tau_1(e)} \mid e \in \mathbb{N}\}$ and is closed under effective σ-union and effective σ-intersection. Then the hyper-arithmetical sets are defined as the smallest effective σ-ring, which can be formally defined as the least fixed point of a certain operator on the set $\mathcal{A} = 2^{\mathbb{N}\times 2^{\mathbb{N}}\times 2^{\mathbb{N}}}$, where a triple (e, B_e, C_e) indicates that the sets B_e and C_e have been defined for the index e in the above inductive definition, and an operator $\Phi : \mathcal{A} \to \mathcal{A}$ represents one step of this inductive definition. Furthermore, this least fixed point can be obtained constructively by a transfinite induction on countable ordinals, which is essential for any proofs about hyper-arithmetical sets. It is known [14, Sec. 8E] [1, Thm. 2.2.3] that for some (easy) choices of τ_1 and τ_2 the smallest effective σ-ring coincides with Δ^1_1 sets.

Fix those two functions and the corresponding $\mathcal{B}$. Note that the definition is valid not for every choice of τ_1 and τ_2: in particular, they must be one-to-one and have disjoint images.

With every set $B_e \in \mathcal{B}$ one can associate a *tree of* B_e, labelled with sets from $\mathcal{B}$: its root is labelled with B_e, and each vertex $B_{\tau_2(e')}$ ($C_{\tau_2(e')}$, respectively) in the tree has children labelled with $\{C_{f_{e'}(n)} \mid n \in \mathbb{N}\}$ ($\{B_{f_{e'}(n)} \mid n \in \mathbb{N}\}$, respectively). Vertices of the form $B_{\tau_1(e')}$ or $C_{\tau_1(e')}$ have no children; these are the only leaves in the tree.

A partial order $\prec$ is *well-founded*, if it has no infinite descending chain. Extending this notion to oriented trees, a tree is well-founded if it contains no infinite downward path.

Lemma 7. *For each pair of sets $B_e, C_e \in \mathcal{B}$ the trees of B_e, C_e are well-founded.*

The well-foundedness of a set allows using the *well-founded induction principle*: given a property ϕ and a well founded order $\prec$ on a set A, $\phi(n)$ is true for all $n \in A$ if

$$(\forall m \prec n\ \phi(m)) \Rightarrow \phi(n).$$

This principle shall be used in the proof of the main construction, which is described in the rest of this section. Note, that the basis of the induction are $\prec$-minimal elements n of A, as for them $\phi(n)$ has to be shown directly.

Fix B_{i_0} as the target set in the root. Consider a path of length k in this tree, going from B_{i_0} to C_{i_1}, B_{i_2}, …, B_{i_k} (or C_{i_k}, depending on the parity of k). Then, for each j-th set in this path, $i_j = f_{\tau_2^{-1}(i_{j-1})}(n_j)$ for some number n_j, and the path is uniquely defined by the sequence of numbers $n_1, \ldots, n_k$. Consider the binary encoding of each of these numbers written using digits `3` and `6` (representing zero and one, respectively), and let *Resolve* be a partial function that maps finite sequences of such "binary" strings representing numbers $n_1, \ldots, n_k$ to the number i_k of the set B_{i_k} or C_{i_k} in the end of this path. The value of this function can be formally defined by induction:

$$\mathit{Resolve}(\langle\rangle) = i_0, \qquad \mathit{Resolve}(x_1, \ldots, x_k) = f_{\tau_2^{-1}(\mathit{Resolve}(x_1,\ldots,x_{k-1}))}((x_k)_2),$$

Note that *Resolve* may be undefined if some τ_2-preimage is undefined.

The goal is to construct a system of equations, such that the following two sets are among the components of its unique solution:

$$\mathit{Goal}_0 = \{(\mathtt{1}x_k\mathtt{1}x_{k-1}\ldots\mathtt{1}x_1\mathtt{10}w)_7 \mid k \geqslant 0, x_i \in \{\mathtt{3},\mathtt{6}\}^*, (w)_7 \in B_{\mathit{Resolve}(x_1,\ldots,x_k)}\},$$
$$\mathit{Goal}_1 = \{(\mathtt{1}x_k\mathtt{1}x_{k-1}\ldots\mathtt{1}x_1\mathtt{10}w)_7 \mid k \geqslant 0, x_i \in \{\mathtt{3},\mathtt{6}\}^*, (w)_7 \in C_{\mathit{Resolve}(x_1,\ldots,x_k)}\}.$$

These sets encode the sets $B_0, B_1, \ldots$ needed to compute B_{i_0}. In this way the (possibly infinite) amount of equations defining sets in hyper-arithmetical hierarchy is encoded in a finite amount of equations using only small number of variables. The set B_i in the node with path to the root encoded by $x_k, x_{k-1}, \ldots, x_1 \in \{\mathtt{3},\mathtt{6}\}^*$ is represented by $\{(\mathtt{1}x_k\mathtt{1}\ldots\mathtt{1}x_k\mathtt{10}w)_7 \mid (w)_7 \in B_i\} \subseteq \mathit{Goal}_0$.

The following set defines the admissible encodings, that is, numbers encoding paths in the tree of B_{i_0}:

$$\mathit{Admissible} = \{(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{10}w)_7 | k \geqslant 0, x_i \in \{\mathtt{3},\mathtt{6}\}^*, \mathit{Resolve}(x_1,\ldots,x_k) \text{ is defined}\}$$

The next two sets represent the leaves of the tree of B_{i_0}, and the numbers in those leaves:

$$R_0 = \{(\mathtt{1}x_k\mathtt{1}x_{k-1}\ldots\mathtt{1}x_1\mathtt{10}w)_7 \mid$$
$$k \geqslant 0,\ x_i \in \{\mathtt{3},\mathtt{6}\}^*, \exists e \in \mathbb{N} : \mathit{Resolve}(x_1,\ldots,x_k) = \tau_1(e),\ (w)_7 \in B_{\tau_1(e)}\},$$

$$R_1 = \{(1x_k1x_{k-1}\dots1x_110w)_7 \mid k \geqslant 0,\ x_i \in \{3,6\}^*, \exists e \in \mathbb{N} : \mathit{Resolve}(x_1,\dots,x_k) = \tau_1(e),\ (w)_7 \in C_{\tau_1(e)}\}.$$

Lemma 8. *The sets Goal_i, Admissible, R_i are r.e. sets, Resolve is an r.e. predicate.*

Consider the following system of equations:

$$X_0 = E(\mathit{Remove}_1(X_1)) \cup R_0 \tag{4.1}$$
$$X_1 = Z \cup R_1 \tag{4.2}$$
$$\widetilde{Y} = E(\mathit{Remove}_1(X_1)) \cup \mathit{Append}_{3,6}(\widetilde{Y}) \tag{4.3}$$
$$Y = Z \cup \mathit{Append}_{3,6}(Y) \tag{4.4}$$
$$Y \subseteq \mathit{Remove}_1(X_0 \cap \mathit{Admissible}) \subseteq Y \cup \widetilde{Y} \tag{4.5}$$
$$Z \subseteq (1\Omega_7^+)_7 \tag{4.6}$$
$$X_0, X_1 \subseteq \mathit{Admissible} \tag{4.7}$$
$$X_0 \cap R_1 = X_1 \cap R_0 = \varnothing \tag{4.8}$$

Its intended unique solution has $X_0 = \mathit{Goal}_0$ and $X_1 = \mathit{Goal}_1$, and accordingly encodes the set B_{i_0}, as well as all sets of $\mathcal{B}$ on which B_{i_0} logically depends. The system implements the functions $E(X)$ and $A(X)$ to represent effective σ-union and σ-intersection, respectively. For that purpose, the expression for $E(X)$ introduced in Lemma E, as well as the system of equations implementing $A(X)$ defined in Lemma A, are applied iteratively to the same variables X_0 and X_1. Intuitively, the above system may be regarded as an implementation of an equation $X_0 = A(E(X_0)) \cup \mathit{const}$.

The proof uses the principle of induction on well-founded structures. The membership of numbers of the form $(1x_k1x_{k-1}\dots1x_110w)_7$ in the variables X_0 and X_1, where $k \geqslant 0$, $x_i \in \{3,6\}^*$ and $w \in \Omega_7^* \setminus 0\Omega_7^*$, is first proved for larger k's and then inductively extended down to $k = 0$, which allows extracting B_{i_0} out of the solution. The well-foundedness of the tree of B_{i_0} means that although B_{i_0} depends upon infinitely many sets, each dependency is over a finite path ending with a constant, that is, the self-dependence of numbers in X_0, X_1 on the numbers in X_0, X_1 reaches a constant R_0, R_1 in finitely many steps (yet the number of steps is unbounded).

Lemma 9. *The unique solution of the system* (4.1)–(4.8) *is*

$$X_0 = \mathit{Goal}_0 = \{(1x_k\dots1x_110w)_7 \mid k \geqslant 0, x_i \in \{3,6\}^*, (w)_7 \in B_{\mathit{Resolve}(x_1,\dots,x_k)}\}$$
$$X_1 = \mathit{Goal}_1 = \{(1x_k\dots1x_110w)_7 \mid k \geqslant 0, x_i \in \{3,6\}^*, (w)_7 \in C_{\mathit{Resolve}(x_1,\dots,x_k)}\}$$
$$Y = \{(x_{k+1}1x_k\dots1x_110w)_7 \mid k \geqslant 0, x_i \in \{3,6\}^*, \forall x_{k+1} : (w)_7 \in B_{\mathit{Resolve}(x_1,\dots,x_{k+1})}\}$$
$$\widetilde{Y} = \{(x_{k+1}1x_k\dots1x_110w)_7 \mid k \geqslant 0, x_i \in \{3,6\}^*, \exists x_{k+1} : (w)_7 \in C_{\mathit{Resolve}(x_1,\dots,x_{k+1})}\}$$
$$Z = \mathit{Goal}_1 \setminus R_1 = \{(1x_k\dots1x_110w)_7 \mid k \geqslant 0, e \in \mathbb{N}, x_i \in \{3,6\}^*, \mathit{Resolve}(x_1,\dots,x_k) = \tau_2(e), (w)_7 \in C_{\tau_2(e)}\}$$

Then, in order to obtain the set B_{i_0}, it remains to intersect $X_0 = \mathit{Goal}_0$ with the recursive constant set $(10\Omega_7^*)_7$, and then remove the leading digits 10 by a construction analogous to the one in Lemma 5.

Theorem 2. *For every hyper-arithmetical set $B \subseteq \mathbb{Z}$ ($B \subseteq \mathbb{N}$) there is a system of equations over subsets of $\mathbb{Z}$ (over subsets of $\mathbb{N}$, respectively) using union, addition and ultimately*

periodic constants (union, addition, subtraction and singleton constants, respectively), such that $(B, \dots)$ is its unique solution.

5. Equations with addition only

Equations over sets of natural numbers with addition as the only operation can represent an *encoding* of every recursive set, with each number $n \in \mathbb{N}$ represented by the number $16n + 13$ in the encoding [9]. In order to define this encoding, for each $i \in \{0, 1, \dots, 15\}$ and for every set $S \subseteq \mathbb{Z}$, denote:

$$\tau_i(S) = \{16n + i \mid n \in S\}.$$

The encoding of a set of natural numbers $\widehat{S} \subseteq \mathbb{N}$ is defined as

$$S = \sigma_0(\widehat{S}) = \{0\} \cup \tau_6(\mathbb{N}) \cup \tau_8(\mathbb{N}) \cup \tau_9(\mathbb{N}) \cup \tau_{12}(\mathbb{N}) \cup \tau_{13}(\widehat{S}),$$

Proposition 2 ([9, Thm. 5.3])**.** *For every recursive set S there exists a system of equations over sets of natural numbers in variables $X, Y_1, \dots, Y_m$ using the operation of addition and ultimately periodic constants, which has a unique solution with $X = \sigma_0(S)$.*

This result is proved by first representing the set S by a system with addition and union, and then by representing addition and union of sets using addition of their σ_0-encodings.

The purpose of this section is to obtain a similar result for equations over sets of integers: namely, that they can represent the same kind of encoding of every hyper-arithmetical set. For every set $\widehat{S} \subseteq \mathbb{Z}$, define its *encoding* as the set

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

The subset $S \cap \{16n + i \mid n \in \mathbb{Z}\}$ is called the *i-th track of S*.

The first result on this encoding is that the condition of a set X being an encoding of any set can be specified by an equation of the form $X + C = D$.

Lemma 10 (cf. [9, Lemma 3.3])**.** *A set $X \subseteq \mathbb{Z}$ satisfies an equation*

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

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

Now, assuming that the given system of equations with union and addition is decomposed to have all equations of the form $X = Y + Z$, $X = Y \cup Z$ or $X = const$, these equations can be simulated in a new system as follows:

Lemma 11 (cf. [9, Lemma 4.1])**.** *For all sets $X, Y, Z \subseteq \mathbb{Z}$,*

$$\sigma(Y) + \sigma(Z) + \{0, 1\} = \sigma(X) + \sigma(\{0\}) + \{0, 1\} \quad \text{if and only if} \quad Y + Z = X$$
$$\sigma(Y) + \sigma(Z) + \{0, 2\} = \sigma(X) + \sigma(X) + \{0, 2\} \quad \text{if and only if} \quad Y \cup Z = X.$$

Using these two lemmata, one can simulate any system with addition and union by a system with addition only. Taking systems representing different hyper-arithmetical sets, the following result on the expressive power of systems with addition can be established:

	Sets representable by unique solutions	Complexity of decision problems	
		solution existence	solution uniqueness
over $2^{\mathbb{N}}$, with $\{+,\cup\}$	Δ_1^0 (recursive) [8]	Π_1^0-complete [8]	Π_2^0-complete [8]
over $2^{\mathbb{N}}$, with $\{+\}$	encodings of Δ_1^0 [9]	Π_1^0-complete [9]	Π_2^0-complete [9]
over $2^{\mathbb{N}}$, with $\{+,\dot{-},\cup\}$	Δ_1^1 (hyper-arithmetical)	Σ_1^1-complete	$\Pi_1^1 \leqslant \cdot \leqslant \Delta_2^1$
over $2^{\mathbb{Z}}$, with $\{+,\cup\}$	Δ_1^1	Σ_1^1-complete	$\Pi_1^1 \leqslant \cdot \leqslant \Delta_2^1$
over $2^{\mathbb{Z}}$, with $\{+\}$	encodings of Δ_1^1	Σ_1^1-complete	$\Pi_1^1 \leqslant \cdot \leqslant \Delta_2^1$

Table 1: Summary of the results.

Theorem 3. *For every hyper-arithmetical set $S \subseteq \mathbb{Z}$ there exists a system of equations over sets of integers using the operation of addition and ultimately periodic constants, which has a unique solution with $X_1 = T$, where $S = \{n \mid 16n \in T\}$.*

6. Decision problems

Having a solution (solution existence) and having exactly one solution (solution uniqueness) are basic properties of a system of equations. For language equations with continuous operations, *solution existence* is Π_1^0-complete [19], and it remains Π_1^0-complete already in the case of a unary alphabet, concatenation as the only operation and regular constants [9], that is, for equations over sets of natural numbers with addition only. For the same formalisms, *solution uniqueness* is Π_2^0-complete.

Consider equations over sets of integers. Since their expressive power extends beyond the arithmetical hierarchy, the decision problems should accordingly be harder. In fact, the solution existence is Σ_1^1-complete, which will now be proved using a reduction from the following problem:

Proposition 3 (Rogers [21, Thm. 16-XX])**.** *Consider trees with nodes labelled by finite sequences of natural numbers, such that a node $(x_1, \ldots, x_{k-1}, x_k)$ is a son of $(x_1, \ldots, x_{k-1})$, and the empty sequence ε is the root. Then the following problem is Π_1^1-complete: "Given a description of a Turing machine recognizing the set of nodes of a certain tree, determine whether this tree has no infinite paths".*

In other words, a given Turing machine recognizes sequences of natural numbers, and the task is to determine whether there is *no* infinite sequence of natural numbers, such that all of its prefixes are accepted by the machine. The Σ_1^1-complete complement of the problem is testing whether such an infinite sequence exists, and it can be reformulated as follows:

Corollary 1. *The following problem is Σ_1^1-complete: "Given a Turing machine M working on natural numbers, determine whether there exists an infinite sequence of strings $\{x_i\}_{i=1}^{\infty}$ with $x_i \in \{\mathtt{3},\mathtt{6}\}^*$, such that M accepts $(\mathtt{1}x_k\mathtt{1}x_{k-1}\ldots\mathtt{1}x_1\mathtt{1})_7$ for all $k \geqslant 0$".*

This problem can be reduced to testing existence of a solution of equations over sets of numbers.

Theorem 4. *The problem of whether a given system of equations over sets of integers with addition and ultimately periodic constants has a solution is Σ_1^1-complete.*

Now consider the solution uniqueness property. The following upper bound on its complexity naturally follows by definition:

Theorem 5. *The problem of whether a given system of equations over sets of integers using addition and ultimately periodic constants has a unique solution can be represented as a conjunction of a Σ^1_1-formula and a Π^1_1-formula, and is accordingly in Δ^1_2. At the same time, the problem is Π^1_1-hard.*

The exact hardness of testing solution uniqueness is still open. The properties of different families of equations over sets of numbers are summarized in Table 1.

References

[1] P. Aczel, "An introduction to inductive definitions", in: J. Barwise (Ed.), *Handbook of Mathematical Logic*, 739–783, North-Holland, 1977.

[2] F. d'Alessandro, J. Sakarovitch, "The finite power property in free groups", *Theoretical Computer Science*, 293:1 (2003), 55–82.

[3] A. V. Anisimov, "Languages over free groups", *Mathematical Foundations of Computer Science*, (MFCS 1975, Mariánské Lázně, September 1–5, 1975), LNCS 32, 167–171.

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

[5] J. Y. Halpern, "Presburger arithmetic with unary predicates is Π^1_1 complete", *Journal of Symbolic Logic*, 56:2 (1991), 637–642.

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

[7] A. Jeż, A. Okhotin, "Conjunctive grammars over a unary alphabet: undecidability and unbounded growth", *Theory of Computing Systems*, 46:1 (2010), 27–58.

[8] A. Jeż, A. Okhotin, "On the computational completeness of equations over sets of natural numbers" *ICALP 2008* (Reykjavik, Iceland, July 7–11, 2008), LNCS 5126, 63–74.

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

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

[11] 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.

[12] 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, 360–371.

[13] P. McKenzie, K. Wagner, "The complexity of membership problems for circuits over sets of natural numbers", *Computational Complexity*, 16:3 (2007), 211–244.

[14] Y. Moschovakis, *Elementary Induction on Abstract Structures*, North-Holland, 1974.

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

[16] A. Okhotin, "Conjunctive grammars and systems of language equations", *Programming and Computer Software*, 28:5 (2002), 243–249.

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

[18] A. Okhotin, "Computational universality in one-variable language equations", *Fundamenta Informaticae*, 74:4 (2006), 563–578.

[19] A. Okhotin, "Decision problems for language equations", *Journal of Computer and System Sciences*, 76 (2010), to appear; earlier version at ICALP 2003.

[20] J. Robinson, "An introduction to hyperarithmetical functions", *Journal of Symbolic Logic*, 32:3 (1967), 325–342.

[21] H. Rogers, Jr., *Theory of Recursive Functions and Effective Computability*, McGraw-Hill, 1967.

[22] S. D. Travers, "The complexity of membership problems for circuits over sets of integers" *Theoretical Computer Science*, 369:1–3 (2006), 211–229.