Theory Comput Syst (2012) 51:196–228
DOI 10.1007/s00224-011-9352-5

Representing Hyper-arithmetical Sets by Equations over Sets of Integers

Artur Jeż · Alexander Okhotin

Published online: 4 August 2011

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, defined as $S + T = \{m + n \mid m \in S, n \in T\}$, 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 $S \dot{-} T = \{m - n \mid m \in S, n \in T, m \geq n\}$. 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.

Keywords Language equations · Computability · Arithmetical hierarchy · Hyper-arithmetical hierarchy

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 [19]. Many other kinds of language equations have been studied in the recent years, see a survey by Kunc [14], and most of them were found

A. Jeż (✉)
Institute of Computer Science, University of Wrocław, Wroclaw, Poland
e-mail: aje@cs.uni.wroc.pl

A. Okhotin
Department of Mathematics, University of Turku, Turku, Finland
e-mail: alexander.okhotin@utu.fi

to have strong connections to computability. In particular, for equations with concatenation and Boolean operations, it was shown by Okhotin [20, 22] that the families of languages representable by their unique, least and greatest solutions are exactly the recursive, the recursively enumerable (r.e.) and the co-recursively enumerable (co-r.e.) sets, respectively. A computationally universal equation of the simplest form was constructed by Kunc [13], who proved that the greatest solution of the equation $LX = XL$, 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* $\Gamma = \{a\}$ has recently come in the focus of attention [6–11, 15, 16, 23]. Strings over such an alphabet may be regarded as natural numbers, languages accordingly become sets of numbers, and concatenation of such languages turns into elementwise addition of sets. As established by the authors [9], these equations are computationally 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 concatenation (elementwise addition) if and only if it is recursive. Furthermore, even without the union operation, these equations remain almost as powerful [10]: for every recursive set $S \subseteq \mathbb{N}$, its encoding $\sigma(S) \subseteq \mathbb{N}$, satisfying $n \in S \Leftrightarrow 16n + 13 \in \sigma(S)$ (and also containing some elements not equivalent to 13 modulo 16), can be represented by a unique solution of a system using addition only, as well as ultimately periodic constants. As shown by Lehtinen and Okhotin [15], another, more complicated encoding $\pi(S)$ of any recursive set of natural numbers S can be represented by a unique solution of a system of two equations $X + X + C = X + X + D$, $X + E = F$, where $C, D, E, F \subseteq \mathbb{N}$ are ultimately periodic constants. Analogous results hold *least* and *greatest* solutions of all these equations, which represent r.e. and co-r.e. sets, respectively. Besides representing the expressive power of language equations in a system of an ultimately simple form, these equations over sets of numbers provide yet another instance of computational universality in a basic arithmetical object.

However, it must be noted that the kinds of language equations considered in the literature surveyed above do not exhaust all possible language equations. The recursive upper bound on unique solutions [22] is applicable only to equations with *continuous* operations on languages. Most of the basic language-theoretic operations, such as concatenation, Kleene star, all Boolean operations, non-erasing homomorphisms, etc., are indeed continuous, and thus subject to the above methods. On the other hand, it has already been demonstrated that using the simplest non-continuous operations, such as erasing homomorphisms or the quotient [21], in language equations leads out of the class of recursive solutions. In particular, quotient with regular constants was used to represent all sets in the arithmetical hierarchy [21].

How expressive can language equations be, if they are not restricted to continuous 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 prove that unique solutions of equations with these operations always belong to the family of *hyperarithmetical sets*, which are, roughly speaking, the sets representable in first-order Peano arithmetic augmented with quantifier prefixes of unbounded length [18, 24, 25]. This paper shows that this rather 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 [9] on representing every recursive set by equations over sets of natural numbers with union and addition. In Sect. 2, this result is adapted to the new kinds of equations introduced in this paper. The next task is representing every set in the arithmetical hierarchy, which is achieved in Sect. 3 by simulating existential and universal quantification applied to a recursive predicate. The elements of this construction are then used in Sect. 4 for the construction of equations representing hyper-arithmetical sets. In Sect. 5, it is shown how the constructed equations can be further encoded using equations over sets of integers with addition as the only operation and with ultimately periodic constants: this is achieved by a variant of the known construction for equations over sets of natural numbers [10]. The last question considered in the paper is the complexity of testing whether a given system of equations has a solution: in Sect. 6, this problem is proved to be Σ^1_1-complete in the analytical hierarchy (vs. Π^0_1-complete for language equations with continuous operations [9, 22]).

This result brings to mind a study by Robinson [24], 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 exactly the hyper-arithmetical sets.

Some more related work ought to 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 [17], and a variant of these models defined over sets of integers was investigated by Travers [26].

2 Equations and their basic expressive power

The subject of this paper are systems of equations of the form

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

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 $S+T=\{m+n \mid m\in S, n\in T\}$. The constant sets appearing in a system sometimes will be singletons only, sometimes any ultimately periodic constants[1] will be allowed, 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 be considered as well. These systems have subsets of $\mathbb{N}=\{0,1,2,\dots\}$ both as unknowns and as constant languages. Besides addition and Boolean operations, subtraction $S \dot{-} T = \{m-n \mid m\in S, n\in T, m\geqslant n\}$ shall be occasionally used.

Consider systems with a unique solution. Every such system can be regarded as a specification of a set: for instance, the equation $X=(X+\{2\})\cup\{0\}$ specifies the set of non-negative even numbers, which is its unique solution. For every type of systems, a natural question is, what kind of sets can be represented by unique solutions of these systems. For equations over sets of natural numbers with addition and Boolean operations, these are the recursive sets:

Proposition 1 (Jeż, Okhotin [9, Thm. 4]) *The family of sets of natural numbers representable by unique solutions of systems of equations of the form $\varphi_i(X_1,\dots,X_n)=\psi_i(X_1,\dots,X_n)$ with union, addition and singleton constants, is exactly the family of recursive sets. Using other Boolean operations and any recursive constants does not increase their expressive power.*

It is worth mentioning that addition and Boolean operations on sets of natural numbers have an important property of *continuity*: for every function $\varphi\colon (2^{\mathbb{N}})^n \to 2^{\mathbb{N}}$ defined as a superposition of these operations, and for every convergent sequence $\{(S_1^{(i)},\dots,S_n^{(i)})\}_{i=1}^{\infty}$ of n-tuples of sets,[2] $\lim_{i\to\infty}\varphi(S_1^{(i)},\dots,S_n^{(i)})$ exists and coincides with $\varphi(\lim_{i\to\infty}(S_1^{(i)},\dots,S_n^{(i)}))$. This property is crucial for the recursive upper bound in Proposition 1 to hold.

Turning to subtraction of sets of natural numbers, this operation is not continuous, as witnessed by a sequence $S^{(i)}=\{i\}$ with $\lim_{i\to\infty}S^{(i)}=\varnothing$ and a function $\varphi(X)=X\dot{-}X$, for which $\varphi(\lim_{i\to\infty}S^{(i)})=\varphi(\varnothing)=\varnothing$, but $\lim_{i\to\infty}\varphi(S^{(i)})=$

[1] 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 $|n|\geqslant d$.

[2] Such a sequence $\{(S_1^{(i)},\dots,S_n^{(i)})\}_{i=1}^{\infty}$ is called *convergent*, if, for each j-th component, every number $m\in\mathbb{N}$ belongs either to finitely many sets $S_j^{(i)}$ with $i\geqslant 1$, or to all except finitely many of them.

$\lim_{i\to\infty}\{0\} = \{0\}$. Addition of sets of *integers* is also non-continuous. Thus, systems of equations with these operations are not subject to the upper bound methods behind Proposition 1. 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 $X_1, \dots, X_n$, where all operations and constants are expressible in first-order arithmetic, there exists an arithmetical formula $Eq(X_1, \dots, X_n)$, with free second-order variables $X_1, \dots, X_n$ (representing sets of numbers), with no bound second-order variables, and with any bound first-order variables (representing numbers), and this formula $Eq(S_1, \dots, S_n)$ 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

$$(\forall n) n \in X_i \leftrightarrow \big[(\exists n')(\exists n'') n = n' + n'' \wedge n' \in X_j \wedge n'' \in X_k\big].$$

Now consider the following formulae of second-order arithmetic:

$$\varphi(x) = (\exists X_1) \dots (\exists X_n)\big[Eq(X_1, \dots, X_n) \wedge x \in X_1\big],$$
$$\varphi'(x) = (\forall X_1) \dots (\forall X_n)\big[Eq(X_1, \dots, X_n) \to x \in X_1\big].$$

The formula $\varphi(x)$ represents the membership of x in *some* 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 is a Σ^1_1-formula and φ' is a Π^1_1-formula, 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* [18, 25].

Lemma 2 *For every system of equations in variables $X_1, \dots, X_n$, using operations and constants expressible in first-order arithmetic, if it has a unique solution, then all components of this solution 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 an equation (that is, by one of the components of the solution). The result applies 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:

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 the following two systems of equations in variables $X_1, \dots, X_m, Y_1, \dots, Y_{m'}$:*

1. *In the first system, the unknowns are sets of natural numbers, and it uses the operations of addition, subtraction and union and singleton constants,*
2. *The second system has unknowns over sets of integers and uses the operations of addition and union, singleton constants and the constants $\mathbb{N}$ and $-\mathbb{N}$,*

Each system has a unique solution with $X_i = S_i$.

The proof is by transforming the given original system towards the desired forms. 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 the 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 the additional equations $X_i \subseteq \mathbb{N}$ for each variable.

A difference $X \dot{-} R$ for a recursive constant $R \subseteq \mathbb{N}$ is represented as $(X + (-R)) \cap \mathbb{N}$, where the set $-R = \{-n \mid n \in R\}$ is expressed 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 $X_1, \dots, X_n$, using Boolean operations, 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 $X_i = -S_i$.*

Proof Consider that for every expression $\varphi(X_1, \dots, X_n)$ using addition, Boolean operations and constants, $\varphi(-S_1, \dots, -S_n) = -\varphi(S_1, \dots, S_n)$ for any sets S_i: this can be proved by induction on the structure of φ. Therefore, if $(S_1, \dots, S_n)$ is a solution of the original system, then $(-S_1, \dots, -S_n)$ is a solution of the constructed system. The converse claim is symmetric and holds by the same argument. □

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 $X, Y, Y', Z_1, \dots, Z_m$, such that the set of solutions of this system is*

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

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 as another equation $Y' = X \cap \overline{R}$, which may be ignored), and does so without employing the intersection operation.

Proof Consider a system

$$Y \subseteq R, \quad Y' \subseteq \overline{R}, \quad Y \cup Y' = X, \tag{1}$$

where R is the recursive set from the statement and $\overline{R}$ its complement. It is shown, that each solution of this system satisfies $Y = X \cap R$ and $Y' = X \cap \overline{R}$.

The inclusions $Y' \subseteq \overline{R}$ and $Y \subseteq R$ imply that $Y' \cap R = \varnothing$ and $Y \cap R = Y$, and thus

$$R \cap X = R \cap (Y \cup Y') = (R \cap Y) \cup (R \cap Y') = Y \cup \varnothing = Y.$$

Similarly, $\overline{R} \cap X = Y'$.

The system (1) uses constants R and $\overline{R}$. Since both R and $\overline{R}$ are recursive, by Proposition 1, one can effectively construct a system (using union, addition and singleton constants) with a unique solution, such that R and $\overline{R}$ are among its components. Then, this system is appended to (1), and each reference to R and $\overline{R}$ in (1) is replaced by a reference to the corresponding variable. □

This completes the proof of Lemma 3.

So far systems over sets of integers have been employed only for representing sets of natural numbers. A set of integers, both positive and negative, can be specified by first representing its positive and negative subsets individually:

Lemma 4 (Assembling positive and negative subsets) *Let $S \subseteq \mathbb{Z}$ and assume that the sets $S \cap \mathbb{N}$ and $(-S) \cap \mathbb{N}$ are representable by unique solutions of equations over sets of integers using union, addition and ultimately periodic constants. Then, S is representable by equations of the same kind.*

Proof Consider the systems representing $S_+ = S \cap \mathbb{N}$ and $S_- = (-S) \cap \mathbb{N}$. Applying the transformation of Lemma 3.2 to the system for S_- and combining these two systems into one leads to a system of equations in variables $X_+, X_-, X_1, \ldots, X_m$, which has a unique solution with $X_+ = S \cap \mathbb{N}$ and $X_- = S \cap (-\mathbb{N})$. It remains to add one more equation

$$X = X_+ \cup X_-$$

to obtain a unique solution with $X = S$. □

In conjunction with Proposition 1 and Lemma 3, the above Lemma 4 asserts the representability of every recursive set of integers. In the following, these results shall be extended to hyper-arithmetical sets. To that goal, the rest of this paper describes the construction of systems of the form required by Lemma 3.

The following technical property of equations over sets of numbers will be useful in proving the correctness of constructions. It was earlier established for sets of natural numbers with the operations of union, intersection and addition, and it is now augmented to accommodate for the subtraction operation:

Proposition 2 ([7, Lem. 4]) *Let $\varphi(X)$ be an expression defined as a composition of the following operations: the variable X; constant sets; union; intersection with a constant set; addition of a constant set; subtraction of a constant set. Then the function $\varphi : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ is distributive over infinite union, that is, $\varphi(X) = \bigcup_{n \in X} \varphi(\{n\})$.*

The existing proof in the cited paper can be straightforwardly extended for the extra operation of subtraction of a constant set.

3 Representing the arithmetical hierarchy

A set of integers is called *arithmetical*, if the membership of a number n in this set is given by a formula $\varphi(n)$ of first-order Peano arithmetic. Each arithmetical set can be represented by a recursive relation with a quantifier prefix, and arithmetical sets form a hierarchy with respect to the number of quantifier alternations in such a formula, known as the *arithmetical hierarchy*. At the bottom of the hierarchy, there are the recursive sets, and every next level is comprised of two classes, Σ_k^0 or Π_k^0, which correspond to the cases of the first quantifier's being existential or universal. For every $k \geqslant 0$, a set is in Σ_k^0, if it can be represented as

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

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_k^0, if it admits a similar representation with the quantifier prefix $\forall x_1 \exists x_2 \dots Q_k x_k$. By the duality of the definition, $\Pi_k^0 = \{S \mid \overline{S} \in \Sigma_k^0\}$. The sets Σ_1^0 and Π_1^0 are the recursively enumerable sets and their complements, respectively. The arithmetical hierarchy is known to be strict: $\Sigma_k^0 \subset \Sigma_{k+1}^0$ and $\Pi_k^0 \subset \Pi_{k+1}^0$ for every $k \geqslant 0$. Furthermore, for every $k \geqslant 0$, the inclusion $\Sigma_k^0 \cup \Pi_k^0 \subset \Sigma_{k+1}^0 \cap \Pi_{k+1}^0$ is proper, that is, 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[3] as follows: a set $S \subseteq \mathbb{N}$ is in Σ_k^0, if it is representable as

$$\begin{aligned} S = \{(w)_7 \mid \exists x_1 \in \{\mathtt{3}, \mathtt{6}\}^* \forall x_2 \in \{\mathtt{3}, \mathtt{6}\}^* \dots Q_k x_k \in \{\mathtt{3}, \mathtt{6}\}^* : \\ (\mathtt{1} x_1 \mathtt{1} x_2 \dots \mathtt{1} x_k \mathtt{1} w)_7 \in R\}, \end{aligned}$$

for some recursive set $R \subseteq \mathbb{N}$, where $(w)_7$ for $w \in \{\mathtt{0}, \mathtt{1}, \dots, \mathtt{6}\}^*$ denotes the natural number with base-7 notation w. It is usually assumed that w has no leading zeroes, that is, $w \in \Gamma_7^* \setminus \mathtt{0}\Gamma_7^*$. In particular, the number 0 is denoted by $w = \varepsilon$. The strings

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

$x_i \in \{\mathtt{3}, \mathtt{6}\}^*$ represent *binary* notation of some numbers, where $\mathtt{3}$ stands for zero and $\mathtt{6}$ stands for one. The notation $(x)_2$ for $x \in \{\mathtt{3}, \mathtt{6}\}^*$ shall be used to denote the number represented by this encoding. The digits $\mathtt{1}$ act as separators. Throughout this paper, the set of base-7 digits $\{\mathtt{0}, \mathtt{1}, \ldots, \mathtt{6}\}$ shall be denoted by Γ_7.

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

$$\begin{aligned} \mathit{Remove}_{\mathtt{1}}(X) &= \{(w)_7 \mid w \in \Gamma_7^* \setminus \mathtt{0}\Gamma_7^*, (\mathtt{1}w)_7 \in X\}, \\ E(X) &= \{(\mathtt{1}w)_7 \mid \exists x \in \{\mathtt{3}, \mathtt{6}\}^* : (x\mathtt{1}w)_7 \in X\}, \\ A(X) &= \{(\mathtt{1}w)_7 \mid \forall x \in \{\mathtt{3}, \mathtt{6}\}^* : (x\mathtt{1}w)_7 \in X\}. \end{aligned}$$

Then,

$$S = \mathit{Remove}_{\mathtt{1}}(Q_k(\ldots \mathit{Remove}_{\mathtt{1}}(A(\mathit{Remove}_{\mathtt{1}}(E(\mathit{Remove}_{\mathtt{1}}(R))))) \ldots)).$$

An expression converting numbers of the form $(\mathtt{1}w)_7$ to $(w)_7$ is constructed using a variant of the previously used method of adding a constant set, and intersecting the sum with another set to filter out unintended sums [6, 7]. Though in this case addition is replaced by subtraction, the general method remains the same:

Lemma 5 (Removing leading digit $\mathtt{1}$) *The value of the expression*

$$\bigcup_{t \in \{0,1\}} \left[\left(X \cap (\mathtt{1}\Gamma_7^t(\Gamma_7^2)^* \setminus \mathtt{1}\mathtt{0}\Gamma_7^*)_7 \right) \dot{-} (\mathtt{1}\mathtt{0}^*)_7 \right] \cap (\Gamma_7^t(\Gamma_7^2)^* \setminus \mathtt{0}\Gamma_7^*)_7$$

on any $S \subseteq \mathbb{N}$ *is* $\mathit{Remove}_{\mathtt{1}}(S) = \{(w)_7 \mid (\mathtt{1}w)_7 \in S\}$.

Proof Denote the given expression by $\varphi(X)$. According to Proposition 2, it is distributive over infinite union, so it is sufficient to evaluate it on a single number n, and then obtain $\varphi(S)$ as $\bigcup_{n \in S} \varphi(\{n\})$.

The expression is designed to process a number $n = (\mathtt{1}w)_7$ with $w \in \Gamma_7^* \setminus \mathtt{0}\Gamma_7^*$ by subtracting the particular number $(\mathtt{1}\mathtt{0}^{|w|})_7$, which removes the leading digit as intended:

$$\begin{array}{r} \mathtt{1}\ w_1\ w_2 \ldots w_{|w|} \\ -\ \underline{\mathtt{1}\ \ \mathtt{0}\ \ \mathtt{0}\ \ldots\ \ \mathtt{0}} \\ w_1\ w_2 \ldots w_{|w|} \end{array}$$

However, the subtraction of the entire set $(\mathtt{1}\mathtt{0}^*)_7$ yields as many as $|w|$ other differences, in which $\mathtt{1}$ is subtracted from other digits, and all these differences need to be filtered out by the final intersection. Since the second leading digit of n is non-zero by assumption, all these erroneous differences have the same number of base-7 digits as n, while the correct difference has one less digit. For this reason, the cases of an even and an odd number of digits in n are treated separately, and the final intersection verifies that the number of digits modulo two has changed, which happens only in the correct differences.

The number $(\mathtt{1})_7 = 1$ is processed correctly, because the only possible subtraction is $(\mathtt{1})_7 - (\mathtt{1})_7 = (\varepsilon)_7$, and hence $\varphi(\{1\}) = \{0\} = \{(\varepsilon)_7\}$, as in the definition of $\mathit{Remove}_{\mathtt{1}}$.

Assume that $n = (\mathtt{1}iw)_7$ for some $i \in \Gamma_7 \setminus \{\mathtt{0}\}$ and $w \in \Gamma_7^*$. Then the only nonempty term in $\varphi(\{n\})$ is the one corresponding to $t = |iw| \pmod 2$, and accordingly

$$\varphi(\{n\}) = \left[\left(\{n\} \cap (\mathtt{1}\Gamma_7^t(\Gamma_7^2)^* \setminus \mathtt{10}\Gamma_7^*)_7\right) \dot{-} (\mathtt{10}^*)_7\right] \cap (\Gamma_7^t(\Gamma_7^2)^* \setminus \mathtt{0}\Gamma_7^*)_7.$$

Consider any number $m = (\mathtt{10}^\ell)_7$ subtracted from n. If $\ell = |iw|$, the difference $n - m = (iw)_7$ is in $(\Gamma_7^t(\Gamma_7^2)^* \setminus \mathtt{0}\Gamma_7^*)_7$, and hence in $\varphi(\{n\})$. If $\ell \leqslant |w|$, then, taking into account that $i > 0$, the difference is

$$(\mathtt{1}iw)_7 - (\mathtt{10}^\ell)_7 \geqslant (\mathtt{1}iw)_7 - (\mathtt{10}^{|w|})_7 = (\mathtt{1}(i-1)w)_7,$$

and therefore has the same number of base-7 digits as n. Accordingly, it is filtered out by the intersection with $(\Gamma_7^t(\Gamma_7^2)^* \setminus \mathtt{0}\Gamma_7^*)_7$. If $\ell > |w| + 1$, then the resulting number is negative and it is filtered out as well. This shows that $\varphi(\{n\})$ produces only the numbers in $\mathit{Remove}_{\mathtt{1}}(\{n\})$. □

With Lemma 5 established and the expression given therein proved to implement the function $\mathit{Remove}_{\mathtt{1}}(X)$, the notation $\mathit{Remove}_{\mathtt{1}}(X)$ shall be used in equations to refer to this expression.

Next, consider the function $E(X)$ representing the existential quantifier ranging over strings in $\{\mathtt{3}, \mathtt{6}\}^*$. This function is not continuous, and accordingly, it cannot be expressed using addition and Boolean operations only. It can be implemented by an expression involving subtraction as follows:

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

$$\left[X \cap (\mathtt{1}\Gamma_7^*)_7\right] \cup \left[\left((X \cap (\{\mathtt{3}, \mathtt{6}\}\Gamma_7^*)_7) \dot{-} (\{\mathtt{3}, \mathtt{6}\}^+\mathtt{0}^*)_7\right) \cap (\mathtt{1}\Gamma_7^*)_7\right]$$

on any $S \subseteq (\{\mathtt{3}, \mathtt{6}\}^*\mathtt{1}\Gamma_7^*)_7$ *is* $E(S) = \{(\mathtt{1}w)_7 \mid \exists x \in \{\mathtt{3}, \mathtt{6}\}^* : (x\mathtt{1}w)_7 \in S\}$.

Note, that $E(X)$ can already produce any recursively enumerable set from a recursive argument. Thus, a single application of the non-continuous subtraction operation can already surpass the upper bound of Proposition 1.

Proof Denote the whole expression by $[X \cap (\mathtt{1}\Gamma_7^*)_7] \cup \varphi(X)$, where $\varphi(X) = [(X \cap (\{\mathtt{3}, \mathtt{6}\}\Gamma_7^*)_7) \dot{-} (\{\mathtt{3}, \mathtt{6}\}^+\mathtt{0}^*)_7] \cap (\mathtt{1}\Gamma_7^*)_7$. The first subexpression $X \cap (\mathtt{1}\Gamma_7^*)_7$ takes care of the case of $x = \varepsilon$, while the second subexpression $\varphi(X)$ represents the function $\{(\mathtt{1}w)_7 \mid \exists x \in \{\mathtt{3}, \mathtt{6}\}^+ : (x\mathtt{1}w)_7 \in S\}$, where the quantification is over nonempty strings.

The expression $\varphi(X)$ is constructed by generally the same method of subtraction followed by intersection as in Lemma 5. Since $\varphi(X)$ is, by Proposition 2, distributive over infinite union, it is enough to consider the value of φ on a single number $n = (x1w)_7 \in S$ with $x \in \{3, 6\}^+$, and show that $\varphi(\{(x1w)_7\}) = \{(1w)_7\}$.

The general plan is to subtract the number $(x0^{|1w|})_7$ from n, which directly gives the required result:

$$\begin{array}{r} x_1\ x_2 \ldots x_{|x|}\ 1\ w_1\ w_2 \ldots w_{|w|} \\ -\ \underline{x_1\ x_2 \ldots x_{|x|}\ 0\ \ 0\ \ 0\ \ldots\ \ 0} \\ 1\ w_1\ w_2 \ldots w_{|w|} \end{array}$$

The subtraction is followed by a check that the leading digit of the result is 1, represented by an intersection with $(1\Gamma_7^*)_7$. The question is, whether any unintended numbers obtained by such a subtraction could pass through the subsequent intersection.

In general, the expression $\{n\} \dot{-} (\{3, 6\}^+ 0^*)_7$ allows subtracting any number of the form $(z0^\ell)_7$ with $\ell \geqslant 0$ and $z \in \{3, 6\}^+$. It is claimed that as long as the difference $(x1w)_7 - (z0^\ell)_7$ is in $(1\Gamma_7^*)_7$, the subtraction has been done according to the plan (in other words, any unintended subtraction is filtered out by the intersection).

Claim Let $x, z \in \{3, 6\}^+$, $w, w' \in \Gamma_7^*$ and $\ell \geqslant 0$ satisfy $(x1w)_7 - (z0^\ell)_7 = (1w')_7$. Then $x = z$ and $w = w'$.

Proof It is first shown that these two numbers have the same number of digits, that is, $|x1w| = |z0^\ell|$.

- If $|x1w| < |z0^\ell|$, then $(x1w)_7 < (z0^\ell)_7$, and so the difference $(x1w)_7 - (z0^\ell)_7$ is negative.
- If $|x1w| > |z0^\ell|$, the difference is positive, but its leading digit cannot be 1. To see this, consider that the number $(x1w)_7$ is greater than $3 \cdot 7^{|x1w|-1}$ and less than $7 \cdot 7^{|x1w|-1}$, because $x \in \{3, 6\}^+$. On the other hand, $(z0^\ell)_7 < 7 \cdot 7^{|z0^\ell|-1} \leqslant 7^{|x1w|-1}$. Thus,

  $$2 \cdot 7^{|x1w|-1} < (x1w)_7 - (z0^\ell)_7 < 7 \cdot 7^{|x1w|-1},$$

  and so the leading digit of $(x1w)_7 - (z0^\ell)_7$ is between 2 and 6, contradiction.

The only remaining possibility is that $|x1w| = |z0^\ell|$.

It remains to show, that if $|x1w| = |z0^\ell|$ and $(x1w)_7 - (z0^\ell)_7 = (1w')_7$, then $x = z$ and $w = w'$. To this end, it is shown by induction on the length of x, that $x = z$; the other claim, that $w = w'$, naturally follows. For the simplicity of the argument, both strings of digits x and z are allowed to be empty: that is, assume $x, z \in \{3, 6\}^*$ instead of $x, z \in \{3, 6\}^+$.

For the induction basis, consider $x = \varepsilon$. If $z \neq \varepsilon$, then the leading digit of $(z0^\ell)_7$ is 3 or 6. Then $(z0^\ell)_7 > (1w)_7$, and therefore the subtraction $(x1w)_7 - (z0^\ell)_7$ results in a negative number, which is a contradiction. Therefore, $z = \varepsilon$, as claimed.

For the induction step, consider $x = x_1x'$, where $x_1 \in \{3, 6\}$ is the first digit of x. Note, that $z \neq \varepsilon$, as in such a case $(x1w)_7 - (z0^\ell)_7 = (x1w)_7 - 0 = (x1w)_7$, and this number has the leading digit $x_1 \neq 1$, contradiction. Hence, $z = z_1z'$, where $z_1 \in \{3, 6\}$ is the first digit of z. It is shown that $z_1 = x_1$.

– If $z_1 > x_1$, then the difference $(x1w)_7 - (z0^\ell)_7$ is negative.
– If $z_1 < x_1$, then $x_1 = 6$ and $z_1 = 3$. Accordingly, $(x1w)_7 > 6 \cdot 7^{|x1w|-1}$ and $(z0^\ell)_7 < 4 \cdot 7^{|z0^\ell|-1} = 4 \cdot 7^{|x1w|-1}$, and their difference lies within the following bounds:

$$2 \cdot 7^{|x1w|-1} < (x1w)_7 - (z0^\ell)_7 < 7 \cdot 7^{|x1w|-1}.$$

Consequently, the leading digit of this difference is between 2 and 6, contradiction.

Therefore, $x_1 = z_1$, and $(x_1x'1w)_7 - (z_1z'0^\ell)_7 = (x'1w)_7 - (z'0^\ell)_7$. Since $|x'| < |x|$, the induction assumption asserts that $x' = z'$.

So it was shown that $|x1w| = |z0^\ell|$ and $x = z$. Thus, $(x1w)_7 - (z0^\ell)_7 = (1w)_7$, which concludes the proof. □

Getting back to the proof of Lemma E, the above claim implies that, for $x \in \{3, 6\}^+$,

$$\varphi(\{(x1w)_7\}) = \{(1w)_7\}.$$

The value of the entire expression for $S \subseteq \{3, 6\}^*1\Gamma_7^*$ is

$$\big[S \cap (1\Gamma_7^*)_7\big] \cup \varphi(S) = \{(1w)_7 \mid \exists x \in \{3, 6\}^* : (x1w)_7 \in S\} = E(S),$$

as claimed in the lemma. □

With the existential quantifier implemented, the next task is to represent a universal quantifier. Though it would be convenient to devise an expression implementing $A(X)$, this provably cannot be done, as long as the operations are limited to addition, subtraction, union and intersection. It was recently shown by the authors [12, Thm. 3] that though a superposition of these operations need not be continuous, it always has a weaker property of $\cup$-continuity.[4] However, $A(X)$ is not $\cup$-continuous, which is witnessed by an ascending sequence $S^{(i)} = \{(x1)_7 \mid x \in \{3, 6\}^*,\ (x)_7 \leqslant i\}$ with $A(\lim_{i\to\infty} S^{(i)}) = A\big(\{(x1)_7 \mid x \in \{3, 6\}^*\}\big) = \{0\}$, but $A(S^{(i)}) = \varnothing$ and thus $\lim_{i\to\infty} A(S^{(i)}) = \varnothing$. For this reason, the universal quantifier has to be implemented implicitly, as a solution of an equation.

The equation representing the function $A(X)$ shall use the another function representing the set of pre-images of $E(X)$:

$$E^{-1}(X) = \{(x1w)_7 \mid x \in \{3, 6\}^* : (1w)_7 \in X\}.$$

It will be shown later that E^{-1} is a quasi-inverse of $A(X)$, in the sense that $A(E^{-1}(S)) = S$ for all $S \subseteq (1\Gamma_7^*)_7$ and $E^{-1}(A(T)) \subseteq T$ for every set $T \subseteq (\{3, 6\}^*1\Gamma_7^*)_7$. Unlike $A(X)$, the function $E^{-1}(X)$ can be represented by an expression over sets of natural numbers.

[4] A function φ is $\cup$-continuous if $\lim_{i\to\infty}(\varphi(S^{(i)})) = \varphi(\lim_{i\to\infty}(S^{(i)}))$ for every *ascending* sequence $S^{(0)} \subseteq S^{(0)} \subseteq \cdots \subseteq S^{(i)} \subseteq \cdots$.

Lemma E^{-1} (Inverse of the existential quantifier) *The value of the expression*

$$(X \cap (1\Gamma_7^*)_7) \cup \big[\big((X \cap (1\Gamma_7^*)_7) + (\{3,6\}^+0^*)_7\big) \cap (\{3,6\}^+1\Gamma_7^*)_7\big]$$

on any $S \subseteq \mathbb{N}$ *is* $E^{-1}(S) = \{(x1w)_7 \mid x \in \{3,6\}^*, (1w)_7 \in S\}$.

Proof As in Lemma E, the expression is represented as $[X \cap (1\Gamma_7^*)_7] \cup \varphi(X)$, where $\varphi(X) = ((X \cap (1\Gamma_7^*)_7) + (\{3,6\}^+0^*)_7) \cap (\{3,6\}^+1\Gamma_7^*)_7$. An empty string $x = \varepsilon$ is appended in the first subexpression, and $\varphi(X)$ appends nonempty strings. It is claimed that $\varphi(X) = \{(x1w)_7 \mid x \in \{3,6\}^+, (1w)_7 \in S\}$.

The structure of the expression $\varphi(X)$ representing the function $E^{-1}(X)$ mirrors the expression for the function $E(X)$ constructed in Lemma E. As in the proof of Lemma E, the function φ is distributive over infinite union, and it is sufficient to evaluate it on a singleton $\{(1w)_7\}$. This expression operates by adding an arbitrary number of the form $(x0^\ell)_7$, with $x \in \{3,6\}^+$, and its intended meaning is to add $(x0^{|1w|})_7$ as follows:

$$\begin{array}{r r} & 1\; w_1\; w_2 \dots w_{|w|} \\ + & x_1\; x_2 \dots x_{|x|}\; 0\; 0\; 0 \dots 0 \\ \hline & x_1\; x_2 \dots x_{|x|}\; 1\; w_1\; w_2 \dots w_{|w|} \end{array}$$

However, any numbers $(x0^\ell)_7$ with $\ell \neq |1w|$ can be added as well. To see that all such sums are filtered out by the subsequent intersection, note, that the equality $(1w)_7 + (x0^\ell)_7 = (y1w')_7$ can be equivalently reformulated as $(y1w')_7 - (x0^\ell)_7 = (1w)_7$. Then the assumptions of the Claim in the proof of Lemma E are satisfied, and therefore $x = y$ and $w = w'$.

This shows that for an arbitrary $n = (1w)_7$,

$$\varphi(\{n\}) = (\{3,6\}^+1w)_7,$$

and accordingly the entire expression has the value $(\{3,6\}^*1w)_7 = E^{-1}(\{(1w)_7\})$, as claimed. □

Now, for an arbitrary set $S \subseteq (\{3,6\}^*1\Gamma_7^*)_7$, the set $A(S)$ shall be expressed by a system of equations with an unknown Y, which has a unique solution $Y = A(S)$. The condition $Y \subseteq A(S)$ is specified by the inequality $E^{-1}(Y) \subseteq S$. In order to represent the converse inclusion $A(S) \subseteq Y$, the construction requires the complement of S up to a certain set, such as $\widetilde{S} = (\{3,6\}^*1\Gamma_7^*)_7 \setminus S$ (there is a more general definition below). Then this inclusion is equivalent to $S \subseteq E^{-1}(Y \cup E(\widetilde{S}))$. The equivalence between these conditions is verified in the following lemma.

Lemma A (Representing the universal quantifier) *Let* $S, \widetilde{S} \subseteq (\{3,6\}^*1\Gamma_7^*)_7$ *be two disjoint sets, and let their union* $S \cup \widetilde{S}$ *be of the form* $\{(x1z)_7 \mid x \in \{3,6\}^*, w \in L_0\}$, *for some language* $L_0 \subseteq \Gamma_7^*$. *Then the following system of equations over sets of integers*

$$Y \subseteq (1\Gamma_7^*)_7 \tag{2a}$$

$$E^{-1}(Y) \subseteq S \subseteq E^{-1}(Y \cup E(\widetilde{S})), \tag{2b}$$

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

Proof The first claim is that $E^{-1}(Y) \subseteq S$ if and only if $Y \subseteq A(S)$.

$\Rightarrow$ Let $E^{-1}(Y) \subseteq S$. Applying A to both sides of the inequality gives $A(E^{-1}(Y)) \subseteq A(S)$. Note that

$$A(E^{-1}(T)) = A\big(\big\{(x\mathtt{1}w)_7 \mid x \in \{\mathtt{3},\mathtt{6}\}^*, (\mathtt{1}w)_7 \in T\big\}\big) = T \quad \text{(for all } T \subseteq (\mathtt{1}\Gamma_7^*)_7).$$

Therefore, $Y = A(E^{-1}(Y)) \subseteq A(S)$, which proves the first statement.

$\Leftarrow$ Assume $Y \subseteq A(S)$ and consider any number $(x\mathtt{1}w)_7 \in E^{-1}(Y)$. Then $(\mathtt{1}w)_7 \in Y$, and hence $(\mathtt{1}w)_7 \in A(S)$ by the assumption. From this it follows that $(x\mathtt{1}w)_7 \in S$.

The second claim needed to establish the lemma is that $S \subseteq E^{-1}(Y \cup E(\widetilde{S}))$ if and only if $A(S) \subseteq Y$.

$\Rightarrow$ If $S \subseteq E^{-1}(Y \cup E(\widetilde{S}))$, then $A(S) \subseteq A(E^{-1}(Y \cup E(\widetilde{S}))) = Y \cup E(\widetilde{S})$. Consider that the sets $A(S)$ and $E(\widetilde{S})$ are disjoint: indeed, if $(\mathtt{1}w)_7 \in E(\widetilde{S})$, then $(x\mathtt{1}w)_7 \in \widetilde{S}$ for some x, and hence $(x\mathtt{1}w)_7 \notin S$, which rules out the membership of $(\mathtt{1}w)_7$ in $A(S)$. Therefore, $A(S) \subseteq Y \cup E(\widetilde{S})$ implies $A(S) \subseteq Y$.

$\Leftarrow$ Assume that $A(S) \subseteq Y$ and consider any number $(x\mathtt{1}w)_7 \in S$. Consider the following two possibilities:

- If there exists $y \in \{\mathtt{3},\mathtt{6}\}^*$ with $(y\mathtt{1}w)_7 \in \widetilde{S}$, then $(\mathtt{1}w)_7 \in E(\widetilde{S})$.
- Otherwise, $(y\mathtt{1}w)_7 \notin \widetilde{S}$ for all $y \in \{\mathtt{3},\mathtt{6}\}^*$. Then $(y\mathtt{1}w)_7 \in S$ for all such y, and hence $(\mathtt{1}w)_7 \in A(S)$. By the assumption, this implies $(\mathtt{1}w)_7 \in Y$.

In both cases, $(\mathtt{1}w)_7 \in Y \cup E(\widetilde{S})$, and therefore $(x\mathtt{1}w)_7 \in E^{-1}(Y \cup E(\widetilde{S}))$. □

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. Finally, this system is transformed according to Lemma 3 to both target forms:

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, singleton constants and the constants* $\mathbb{N}$ *and* $-\mathbb{N}$ (*addition, subtraction, union and singleton constants, respectively*).

Proof The statement will be first established in the case of S being a set of non-negative integers, and the system using union, intersection with recursive constants, addition of recursive constants and subtraction of recursive constants (as required by Lemma 3). If $S \subseteq \mathbb{N}$ is an arithmetical set, then it belongs to some level of the arithmetical hierarchy, that is, $S \in \Sigma_k^0$ or $S \in \Pi_k^0$ for some $k \geqslant 0$. A system of equations over sets of integers representing S is constructed inductively on k.

The base case is S being recursive. Then it is representable by Proposition 1.

Let $S \in \Sigma_k^0$ for some $k \geqslant 1$. Then S can be represented in the form

$$S = \{(w)_7 \mid \exists x \in \{\mathtt{3},\mathtt{6}\}^* : (x\mathtt{1}w)_7 \in T\} = Remove_1(E(T)),$$

for some $T \in \Pi^0_{k-1}$. By the induction hypothesis, there is a system of equations in variables $X, X_1, \ldots, X_m$, which has a unique solution with $Y = T$. Adding an extra equation

$$Y = Remove_1(E(X)),$$

constructed according to Lemma 5 and Lemma E, yields a unique solution with $Y = S$.

Assume $S \in \Pi^0_k$ with $k \geqslant 1$. Such a set is representable as

$$S = \{(w)_7 \mid \forall x \in \{3, 6\}^* : (x1w)_7 \in T\} = Remove_1(A(T)),$$

where $T \in \Sigma^0_{k-1}$. The set

$$T' = \{(x1w)_7 \mid x \in \{3, 6\}^*, (x1w)_7 \notin T\} = (\{3, 6\}^*1(\Gamma_7^* \setminus 0\Gamma_7^*))_7 \setminus T$$

is accordingly in Π^0_{k-1}. By the induction hypothesis, both T and T' are representable by a system of equations in variables $X, X', X_1, \ldots, X_m$, whose unique solution has $X = T$ and $X' = T'$. The condition $Y = A(T)$, where Y is a new variable, is represented by the following system of equations, constructed as in Lemma A:

$$Y \subseteq (1\Gamma_7^*)_7$$
$$E^{-1}(Y) \subseteq X \subseteq E^{-1}(Y \cup E(X')).$$

According to the lemma, the resulting system has a unique solution with $Y = \{(1w)_7 \mid \forall x \in \{3, 6\}^* : (x1w)_7 \in T\}$. Adding an extra equation

$$Y' = Remove_1(Y)$$

leads to a representation of S, which proves this last case of the induction step.

An equivalent system of equations over sets of natural numbers, using union, addition, subtraction and singleton constants, can be constructed according to Lemma 3.

Consider an arithmetical set $S \subseteq \mathbb{Z}$. Then the sets $S_+ = S \cap \mathbb{N}$ and $S_- = (-S) \cap (\mathbb{N} \setminus \{0\})$ are both arithmetical, and since $S_+, S_- \subseteq \mathbb{N}$, each of them is representable by a unique solution of some system of equations by the above argument. By Lemma 3 and Lemma 4, this system is converted to the target form. □

Since every arithmetical set is representable by a unique solution, Lemma 3.3 can now be strengthened to the following result to be used later on:

Corollary 1 (Intersection with arithmetical constants) *Let $R \subseteq \mathbb{N}$ be an arithmetical set. Then there is a system of equations over sets of natural numbers using union, addition and singleton constants, in variables $X, Y, Y', Z_1, \ldots, Z_m$, such that the set of solutions of this system is*

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

for some fixed sets $S_1, \ldots, S_m \subseteq \mathbb{N}$.

With this statement established, Lemma 3 can be accordingly improved to handle systems with arithmetical constants. Such systems shall now be used to represent an even greater family of sets.

4 Representing Hyper-arithmetical Sets

Each arithmetical set is defined by applying a fixed quantifier prefix to a base recursive set. In particular, it is not possible to evaluate quantifier prefixes of varying (unbounded) length when testing the membership of different numbers. The more general definition of *hyper-arithmetical sets* allows expressing the limit over all finite quantifier prefixes, and thus continues the arithmetical hierarchy to transfinite levels. It turns out that the definition of hyper-arithmetical sets can be represented in equations over sets of numbers by further extending the methods established in the previous section.

4.1 Definition of Hyper-arithmetical Sets

Following Moschovakis [18, Sec. 8E] and Aczel [1, Thm. 2.2.3], hyper-arithmetical sets shall be defined in set-theoretical terms, as an *effective σ-ring*. Let $f_1, f_2, \ldots$ be an effective enumeration of all partial recursive functions from $\mathbb{N}$ to $\mathbb{N}$. A family of sets $\mathscr{B} = \{B_i \mid i \in I\} \cup \{C_i \mid i \in I\}$, where $I \subseteq \mathbb{N}$ is an index set, is called an effective σ-ring, if there exist two injective recursive functions $\tau_1, \tau_2 \colon \mathbb{N} \to \mathbb{N}$ with disjoint images, such that

1. $\mathscr{B}$ contains the sets $B_{\tau_1(e)} = \mathbb{N} \setminus \{e\}$ and $C_{\tau_1(e)} = \{e\}$ for all $e \in \mathbb{N}$, and
2. for all numbers $e \in \mathbb{N}$, if f_e is a total function and the image of f_e is contained in I, then $\mathscr{B}$ contains

$$B_{\tau_2(e)} = \bigcup_{n \in \mathbb{N}} C_{f_e(n)}, \qquad C_{\tau_2(e)} = \bigcap_{n \in \mathbb{N}} B_{f_e(n)}.$$

The sets B_i and C_i are intended to be complements of each other, which shall be demonstrated later.

Informally, an effective σ-ring contains all singletons and co-singletons and is closed under *effective σ-union* and *effective σ-intersection*. Hyper-arithmetical sets are, by definition, *the smallest effective σ-ring*.[5] The existence of the smallest effective σ-ring is demonstrated constructively, by defining the smallest set of indices $I \subseteq \mathbb{N}$ as a union of a transfinite sequence of sets I_λ, indexed by countable ordinals λ. The below definition at the same time establishes that every effective σ-ring must contain the indices in each I_λ.

[5] And thus may be regarded as the recursion-theoretic counterpart of σ-rings considered in descriptive set theory. A σ-ring is any family of sets closed under countable union and countable intersection. The smallest σ-ring containing all open sets is known as the Borel sets, and hyper-arithmetical sets are their analogue in the recursion theory.

The base set of indices $I_0 = \{\tau_1(e) \mid e \in \mathbb{N}\}$ represents singleton sets $B_{\tau_1(e)}$ and their complements $C_{\tau_1(e)}$. The set of indices for every countable ordinal is defined inductively as follows. For a successor ordinal $\lambda + 1$, let

$$I_{\lambda+1} = \{\tau_2(e) \mid e \in \mathbb{N} \text{ and } \forall n f_e(n) \in I_\lambda\} \cup I_\lambda,$$

and for limit ordinal λ, define

$$I_\lambda = \bigcup_{\alpha<\lambda} I_\alpha.$$

The idea behind this definition is that when an index $i \in I_\lambda$ is obtained at some step, the sets B_i and C_i can be simultaneously defined by referring only to the previously defined sets B_j and C_j.

The convergence of the sequence I_λ after a transfinite yet countable number of steps is established as follows.

Proposition 3 ([18, Thm. 1A.1]) *There exists a countable ordinal λ, for which $I_\lambda = I_{\lambda+1}$.*

Proof Suppose the contrary. Then, for each countable ordinal λ, the set $I_{\lambda+1} \setminus I_\lambda$ is nonempty. Hence, I contains at least one element per countable ordinal, and therefore $|I| \geqslant |\{\lambda \mid \lambda \text{ countable ordinal}\}| > \aleph_0$. This is a contradiction, as $I \subseteq \mathbb{N}$ is a countable set. □

Proposition 4 *If $I_\lambda = I_{\lambda+1}$ for some ordinal λ, then $\mathcal{B}_\lambda = \{B_i \mid i \in I_\lambda\} \cup \{C_i \mid i \in I_\lambda\}$ is the smallest effective σ-ring.*

Proof The condition $I_\lambda = I_{\lambda+1}$ means that the corresponding collection of sets $\mathcal{B}_\lambda$ is closed under effective σ-union and effective σ-intersection. At the same time, by the construction, every set of indices satisfying the definition of an effective σ-ring must contain I_λ, which makes I_λ the smallest such set. Furthermore, every effective σ-ring must contain all sets in $\mathcal{B}_\lambda$, and therefore it is *the smallest* effective σ-ring. □

Now I can be defined as I_λ, as in Proposition 4, which completes the definition of hyper-arithmetical sets. Notably, the class of sets thus defined does not depend upon the choice of the functions τ_1 and τ_2 [18, Sec. 8E], and it forms the bottom of the analytical hierarchy:

The Suslin-Kleene Theorem ([18, Thm. 8E.1], [1, Thm. 2.2.3]) *Hyper-arithmetical sets are exactly the sets in $\Sigma^1_1 \cap \Pi^1_1$.*

4.2 Trees, Well-Founded Orders and Induction

Each hyper-arithmetical set is defined as a formula over the previously defined sets. Its dependencies upon the other sets form a tree with internal nodes of a countable degree representing infinite union or intersection. With every index $i \in I$, one can associate a *tree of i* labeled with indices from I: its root is labeled with i, and each vertex $\tau_2(e')$ in the tree has children labeled with $\{f_{e'}(n) \mid n \in \mathbb{N}\}$. Vertices of the form $\tau_1(e')$ have no children; these are the only *leaves* in the tree. While formally the vertices are labeled with indices, it is convenient to think that each node i denotes the corresponding set B_i (or C_i), with the levels of Bs and Cs alternating as per the definition of these sets. This convention is depicted in Fig. 1.

Proofs involving hyper-arithmetical sets naturally tend to require induction on the structure of such trees. The following property of these trees is essential for carrying out the induction:

Lemma 6 *For every index $i \in I$, the tree of i has no infinite downward path.*

Proof Suppose, for the sake of contradiction, that there exist such indices. Consider the ordinals λ, for which I_λ contains at least one such index i. As any set of ordinals has a minimal element, there exists the smallest such λ. Fix any $i \in I_\lambda$, for which the tree of i has an infinite downward path.

The ordinal λ cannot be a limit ordinal, as then

$$I_\lambda = \bigcup_{\alpha<\lambda} I_\alpha,$$

and therefore $i \in I_\alpha$ for some $\alpha < \lambda$, which contradicts the minimality of λ.

Suppose that $\lambda \neq 0$, by definition, $I_0 = \{\tau_1(e) \mid e \in \mathbb{N}\}$, and thus $i \in I_0$ implies that $i = \tau_1(e)$ for some $e \in \mathbb{N}$. On the other hand, since i has a child, $i = \tau_2(e')$ for some $e' \in \mathbb{N}$. This contradicts the assumption that τ_1 and τ_2 have disjoint images.

Therefore, λ is a successor ordinal, and thus $\lambda = \alpha + 1$ for some ordinal α. Then, by the definition of $I_{\alpha+1}$, all i's children are in I_α; in particular, $i_1 \in I_\alpha$. As i_1 has an infinite downward path and $\alpha < \lambda$, this contradicts the minimality of λ.

Hence, no tree in I has an infinite downward path. □

This tree of dependencies naturally induces an order $\prec$ on the set I: the indices $i = \tau_1(n)$ are the minimal elements of this order and for each $i = \tau_2(e) \in I$, the indices $f_e(n)$ are the direct predecessors of i. The absence of infinite downward paths in the tree implies that the order $\prec$ is *well-founded*, that is, has no infinite descending chain.

Well-founded orders generalize the usual order on natural numbers; in particular, one can use the *well-founded induction principle*: given a predicate P and a well founded order $\prec$ on a set A, if $P(x)$ is true for all $\prec$-minimal elements of A, and if

$$(\forall y \prec x \ P(y)) \Longrightarrow P(x),$$

then $P(x)$ holds for all $x \in A$. This principle shall be used in the proof of the main construction, which is described in the rest of this section.

As a simple example of this proof technique, well-founded induction is employed to show that $B_i = \overline{C_i}$ for each $i \in I$.

Lemma 7 *For each $B_i, C_i \in \mathcal{B}$, it holds that $B_i = \overline{C_i}$.*

Proof As promised, the claim is shown by a well-founded induction on $i \in I$. Each $\prec$-minimal element of I is of the form $\tau_1(e)$ for some e. By definition, $B_{\tau_1(e)} = \mathbb{N} \setminus \{e\}$ and $C_{\tau_1(e)} = \{e\}$, and so the claim holds for $\prec$-minimal elements of I.

If i is not a minimal element of I, it can be represented as $\tau_2(e)$. Then, by definition,

$$B_{\tau_2(e)} = \bigcup_n C_{f_e(n)},$$

where $f_e(n) \prec \tau_2(e)$ for each n. Thus, the induction assumption guarantees that $C_{f_e(n)} = \overline{B_{f_e(n)}}$, for each n. Therefore,

$$B_{\tau_2(e)} = \bigcup_n C_{f_e(n)} = \bigcup_n \overline{B_{f_e(n)}} = \overline{\bigcap_n B_{f_e(n)}} = \overline{C_{\tau_2(e)}},$$

which proves the induction step. □

It is worth mentioning that this property is guaranteed to hold only for the *smallest* effective σ-ring. If the tree of B_i is not well-founded, the values of B_i and C_i may diverge.

4.3 Equations Representing Hyper-arithmetical Sets

Consider an arbitrary hyper-arithmetical set, and let i_0 be its index. The definition of this set B_{i_0} is illustrated by the tree in Fig. 1. The goal is to encode the set B_{i_0} together with all the sets B_j and C_j it depends upon, in a single set. The dependencies between these sets are then expressed uniformly, by a self-reference to this encoding.

Each of the sets in the tree in Fig. 1 is identified by an *address* in this tree, which is a finite sequence of natural numbers identifying a path of length k leading to the set in question. Consider such a path, $B_{i_0}, C_{i_1}, B_{i_2}, \ldots, C_{i_k}$ (or B_{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 $n_1, \ldots, n_k$.

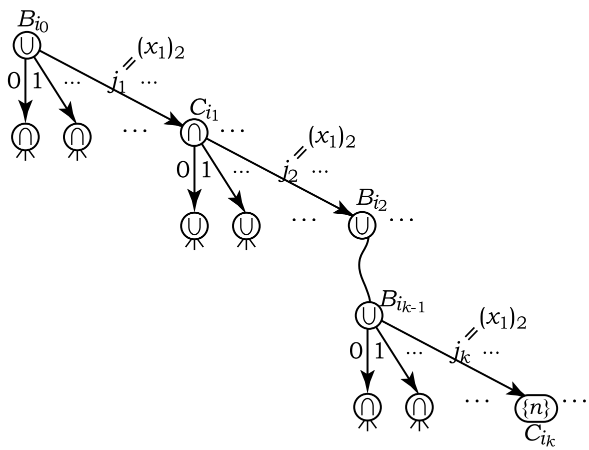

Fig. 1 The tree of dependencies of B_{i_0}, with $Node(x_1, x_2, \ldots, x_k) = i_k$

Consider the binary encoding of each of these numbers written using digits $\mathtt{3}$ and $\mathtt{6}$ (representing zero and one, respectively), and let *Node* be a partial function that maps finite sequences of such "binary" strings representing numbers $n_1, \dots, n_k$ to the index i_k in the end of this path. The value of this function is formally defined by induction as follows:

$$Node(\langle\rangle) = i_0,$$
$$Node(x_1, \dots, x_k) = f_{\tau_2^{-1}(Node(x_1,\dots,x_{k-1}))}((x_k)_2).$$

The index $Node(x_1, \dots, x_k)$ is well-defined, as long as all τ_2-preimages along the path are defined. Since $i_0 \in I$ is a well-defined index of an element of the smallest effective σ-ring, these τ_2-preimages will be defined for all sets B_{i_0} logically depends upon. The $Node(x_1, \dots, x_k)$ is undefined only for all such addresses $x_1, \dots, x_k$, that a shorter address $x_1, \dots, x_\ell$ with $\ell < k$ points to a leaf of the tree: that is, when attempting to resolve an address beyond the end of a path. Consequently, if $Node(x_1, \dots, x_{k-1}, x_k)$ is defined for some x_k, then it is defined for every $x_k \in \{\mathtt{3}, \mathtt{6}\}^*$.

Let $(w)_7$ with $w \in \Gamma_7^* \setminus \mathtt{0}\Gamma_7^*$ be an arbitrary number. Its membership in the sets located under a valid address $(x_1, \dots, x_k)$ in the tree—that is, with well-defined $Node(x_1, \dots, x_k)$—shall be encoded as the number $(\mathtt{1}x_k\mathtt{1}x_{k-1}\dots\mathtt{1}x_1\mathtt{10}w)_7$, where the digits $\mathtt{10}$ unambiguously separate the address from the encoded number. Denote the set of all valid encodings of this kind by

$$Paths = \{(\mathtt{1}x_k\mathtt{1}x_{k-1}\dots\mathtt{1}x_1\mathtt{10}w)_7 \,|\, k \geqslant 0, x_i \in \{\mathtt{3},\mathtt{6}\}^*, Node(x_1, \dots, x_k) \text{ is defined}\}.$$

Since $(w)_7$ belongs either to $B_{Node(x_1,\dots,x_k)}$ or to $C_{Node(x_1,\dots,x_k)}$, its membership status is reflected by arranging the above encodings between the following two sets:

$$T_0 = \{(\mathtt{1}x_k\mathtt{1}x_{k-1}\dots\mathtt{1}x_1\mathtt{10}w)_7 \mid k \geqslant 0, x_i \in \{\mathtt{3},\mathtt{6}\}^*, (w)_7 \in B_{Node(x_1,\dots,x_k)}\},$$
$$T_1 = \{(\mathtt{1}x_k\mathtt{1}x_{k-1}\dots\mathtt{1}x_1\mathtt{10}w)_7 \mid k \geqslant 0, x_i \in \{\mathtt{3},\mathtt{6}\}^*, (w)_7 \in C_{Node(x_1,\dots,x_k)}\}.$$

In particular, a number $(\mathtt{10}w)_7$ with an empty x_i-prefix is in T_0 if and only if $(w)_7 \in B_{Node(\langle\rangle)} = B_{i_0}$.

Note the following basic property of these sets:

Lemma 8 *The sets T_0 and T_1 are disjoint, and their union is Paths.*

The proof is immediate, from the fact that $B_i \cap C_i = \varnothing$ and $B_i \cup C_i = \mathbb{N}$ for all well-defined $i \in I$.

The goal is to construct such a system of equations, that the sets T_0 and T_1 will be among the components of its unique solution. These two sets encode all the sets in $\mathcal{B}$ needed to compute B_{i_0}. A system of equations involving these two variables will represent the (potentially) infinitely many dependencies between the required sets in $\mathcal{B}$ using finitely many equations. The general idea is to implement an equation of the form $X_0 = A(Remove_1(E(Remove_1(X_0)))) \cup const$, in which the functions $E(X)$ and $A(X)$ defined in Sect. 3 represent effective σ-union and σ-intersection, respectively. However, since the function $A(X)$ cannot be implemented as an expression, this intuitive idea of an equation shall be executed using the approach of Lemma A, by simulating universal quantification implicitly using a pair of inequalities.

The equations will use the following constant sets representing the membership of numbers in the leaves of the tree of B_{i_0}:

$$\begin{aligned} R_0 = \{(\mathtt{1}x_k\mathtt{1}x_{k-1}\dots\mathtt{1}x_1\mathtt{10}w)_7 \mid \\ k \geqslant 0, x_i \in \{\mathtt{3},\mathtt{6}\}^*, \exists e \in \mathbb{N} : \mathit{Node}(x_1,\dots,x_k) = \tau_1(e), (w)_7 \in B_{\tau_1(e)}\}, \\ R_1 = \{(\mathtt{1}x_k\mathtt{1}x_{k-1}\dots\mathtt{1}x_1\mathtt{10}w)_7 \mid \\ k \geqslant 0,\ x_i \in \{\mathtt{3},\mathtt{6}\}^*, \exists e \in \mathbb{N} : \mathit{Node}(x_1,\dots,x_k) = \tau_1(e), (w)_7 \in C_{\tau_1(e)}\}. \end{aligned}$$

These sets $R_i \subseteq T_i$ form the basis of the inductive definition encoded in the equations.

Lemma 9 *The function Node is partial recursive. The sets Paths, R_0 and R_1 are recursively enumerable.*

Proof To see that *Node* is a partial recursive function, consider the following semi-algorithm for computing its value. If the argument is an empty sequence, the algorithm returns i_0. If it is $(x_1,\dots,x_{k+1})$ for $k \geqslant 0$, the algorithm recursively invokes itself to calculate $\mathit{Node}(x_1,\dots,x_k) = i_k$, and then considers all numbers $e \in \mathbb{N}$ until it finds one with $\tau_2(e) = i_k$. If this ever happens, the number $f_e((x_{k+1})_2)$ is computed and returned. In case any of the numbers do not exist, the algorithm does not terminate.

The set *Paths* is recursively equivalent to the set of arguments, on which *Node* is defined, and thus the above procedure can be used to test the membership in *Paths*.

For R_0, the semi-decision procedure is as follows: given a number with a base-7 notation $\mathtt{1}x_k\mathtt{1}x_{k-1}\dots\mathtt{1}x_1\mathtt{10}w$, first calculate $i_k = \mathit{Node}(x_1,x_2,\dots,x_k)$, then search for $e \in \mathbb{N}$ with $\tau_1(e) = i_k$, and finally reject if $(w)_7 = e$, otherwise accept. A semi-algorithm for R_1 is similar, with acceptance and rejection switched. If *Node* is not defined, or if i_k is not $\tau_1(e)$ for any e, these semi-algorithms do not terminate. □

Thus, all these sets are arithmetical, and therefore can be represented by systems of equations with unique solutions. This allows using them in a new system of equations as constants.

Consider the following system of equations, which uses subexpressions defined in Lemmata 5, E and E^{-1}:

$$X_0 = E(\mathit{Remove}_{\mathtt{1}}(X_1)) \cup R_0 \quad \text{(3a)}$$

$$X_1 = Y \cup R_1 \quad \text{(3b)}$$

$$Y \subseteq (\mathtt{1}\Gamma_7^*)_7 \quad \text{(3c)}$$

$$E^{-1}(Y) \subseteq \mathit{Remove}_{\mathtt{1}}(X_0) \subseteq E^{-1}(Y \cup E(\mathit{Remove}_{\mathtt{1}}(X_1))) \quad \text{(3d)}$$

$$X_0 \cup X_1 = \mathit{Paths} \quad \text{(3e)}$$

$$X_0 \cap R_1 = X_1 \cap R_0 = \varnothing \quad \text{(3f)}$$

Its intended unique solution is $X_0 = T_0$, $X_1 = T_1$ and $Y = A(Remove_{\mathtt{1}}(T_0))$. 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.

4.4 Proof of Correctness

The goal is now to show that the constructed system indeed has a unique solution of the stated form. The proof is by first verifying that it is a solution, and then by showing that every solution must be of this form. Both parts of the argument are based upon the following characterization of the self-dependencies of T_0 and T_1 specified in a single node of the tree.

Lemma 10 *Let $x_1, \ldots, x_k \in \{\mathtt{3}, \mathtt{6}\}^*$ be an address of a node in the tree, and assume that $X_1 \cap (\mathtt{1}\{\mathtt{3}, \mathtt{6}\}^*\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}\Gamma_7^*)_7 = T_1 \cap (\mathtt{1}\{\mathtt{3}, \mathtt{6}\}^*\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}\Gamma_7^*)_7$. Then $(\mathtt{1}x_k\mathtt{1} \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \in E(Remove_{\mathtt{1}}(X_1))$ if and only if $Node(x_1, \ldots, x_k) = \tau_2(e)$ for some $e \in \mathbb{N}$ and $(w)_7 \in B_{Node(x_1,\ldots,x_k)}$.*

Proof By the definition of E,

$$(\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \in E(Remove_{\mathtt{1}}(X_1)) \tag{4a}$$

holds if and only if

$$\exists x_{k+1} \in \{\mathtt{3}, \mathtt{6}\}^* : (x_{k+1}\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \in Remove_{\mathtt{1}}(X_1), \tag{4b}$$

which is in turn equivalent to

$$\exists x_{k+1} \in \{\mathtt{3}, \mathtt{6}\}^* : (\mathtt{1}x_{k+1}\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \in X_1. \tag{4c}$$

By the assumption, $(\mathtt{1}x_{k+1}\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \in X_1$ holds if and only if $(\mathtt{1}x_{k+1}\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \in T_1$, and by the definition of T_1, the latter is equivalent to $(w)_7 \in C_{Node(x_1,\ldots,x_k,x_{k+1})}$, which additionally implies that $Node(x_1, \ldots, x_k) = \tau_2(e)$ for some $e \in \mathbb{N}$. Thus, (4c) holds if and only if

$$\exists x_{k+1} \in \{\mathtt{3}, \mathtt{6}\}^* : (w)_7 \in C_{f_e((x_{k+1})_2)}, \tag{4d}$$

or, equivalently,

$$(w)_7 \in \bigcup_{x_{k+1} \in \{\mathtt{3},\mathtt{6}\}^*} C_{f_e((x_{k+1})_2)} = \bigcup_{n \in \mathbb{N}} C_{f_e(n)},$$

with the last equality following from the fact that $(x_{k+1})_2$ for all $x_{k+1} \in \{\mathtt{3}, \mathtt{6}\}^*$ enumerates all natural numbers. The latter set is $B_{\tau_2(e)}$ by definition, which is equal to $B_{Node(x_1,\ldots,x_k)}$ by the definition of $\tau_2(e)$. □

The next lemma symmetrically asserts the representation of C_i by $A(Remove_{\mathtt{1}}(X_0))$.

Lemma 11 *Let* $x_1, \ldots, x_k \in \{\mathtt{3}, \mathtt{6}\}^*$ *and let* $X_0 \cap (\mathtt{1}\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}\Gamma_7^*)_7 = T_0 \cap (\mathtt{1}\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}\Gamma_7^*)_7$. *Then* $(\mathtt{1}x_k\mathtt{1} \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \in A(Remove_{\mathtt{1}}(X_0))$ *if and only if* $Node(x_1, \ldots, x_k) = \tau_2(e)$ *with* $e \in \mathbb{N}$ *and* $(w)_7 \in C_{Node(x_1,\ldots,x_k)}$.

The proof is the same as for Lemma 10, with $A(Remove_{\mathtt{1}}(X_0))$ instead of $E(Remove_{\mathtt{1}}(X_1))$, with "$\forall x_{k+1}$" instead of "$\exists x_{k+1}$", with $(\mathtt{1}x_{k+1}\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7$ in X_0 instead of X_1, and with $(w)_7$ in $\bigcap_{x_{k+1}} B_{f_e((x_{k+1})_2)} = C_{\tau_2(e)}$ instead of $\bigcup_{x_{k+1}} C_{f_e((x_{k+1})_2)} = B_{\tau_2(e)}$.

The above lemmata are used to show that the intended solution is indeed a solution.

Lemma 12 *The following assignment of sets to variables forms a solution of the system of equations* (3a)–(3f):

$$X_0 = T_0 = \{(\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \mid k \geqslant 0, x_i \in \{\mathtt{3},\mathtt{6}\}^*, (w)_7 \in B_{Node(x_1,\ldots,x_k)}\},$$
$$X_1 = T_1 = \{(\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \mid k \geqslant 0, x_i \in \{\mathtt{3},\mathtt{6}\}^*, (w)_7 \in C_{Node(x_1,\ldots,x_k)}\},$$
$$Y = A(Remove_{\mathtt{1}}(T_0)).$$

Proof To see that the first equation (3a) holds true under this substitution, that is, $T_0 = E(Remove_{\mathtt{1}}(T_1)) \cup R_0$, consider that $T_0 \subseteq Paths$ and $E(Remove_{\mathtt{1}}(T_1))$, $R_0 \subseteq Paths$, and hence it is sufficient to check that a number $(\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \in Paths$ is in $E(Remove_{\mathtt{1}}(T_1)) \cup R_0$ if and only if it belongs to T_0.

By Lemma 10, $(\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \in E(Remove_{\mathtt{1}}(T_1))$ holds if and only if $Node(x_1, \ldots, x_k) = \tau_2(e)$ for some $e \in \mathbb{N}$ and $(w)_7 \in B_{Node(x_1,\ldots,x_k)}$. At the same time, by the definition of R_0, $(\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \in R_0$ if and only if $Node(x_1, \ldots, x_k) = \tau_1(e)$ for $e \in \mathbb{N}$ and $(w)_7 \in B_{Node(x_1,\ldots,x_k)}$. Combining these two cases together, $(\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \in E(Remove_{\mathtt{1}}(T_1)) \cup R_0$ if and only if $Node(x_1, \ldots, x_k)$ is defined and $(w)_7 \in B_{Node(x_1,\ldots,x_k)}$, which is exactly the condition of the membership of n in T_0.

The second equation (3b) is verified similarly. Both sides of the equality $T_1 = A(Remove_{\mathtt{1}}(T_0)) \cup R_1$ are subsets of *Paths*, and hence it is sufficient to check that every number of the form $(\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7$ with $k \geqslant 0$ and $x_i \in \{\mathtt{3},\mathtt{6}\}^*$ is in $A(Remove_{\mathtt{1}}(T_0)) \cup R_1$ if and only if it is in T_1. The proof is the same as for (3a), this time using Lemma 11.

The next pair of equations (3c)–(3d) is checked according to Lemma A, with $S = Remove_{\mathtt{1}}(T_0)$ and $\widetilde{S} = Remove_{\mathtt{1}}(T_1)$. Firstly, one should validate its assumptions. The sets $Remove_{\mathtt{1}}(T_0)$ and $Remove_{\mathtt{1}}(T_1)$ are disjoint, because so are T_0 and T_1, due to Lemma 8. The union of these sets is

$$\begin{aligned} &Remove_{\mathtt{1}}(T_0) \cup Remove_{\mathtt{1}}(T_1) \\ &\quad = Remove_{\mathtt{1}}(T_0 \cup T_1) = Remove_{\mathtt{1}}(Paths) \\ &\quad = \{(x_k\mathtt{1}x_{k-1} \ldots \mathtt{1}x_1\mathtt{1}\mathtt{0}w)_7 \mid k \geqslant 1, x_i \in \{\mathtt{3},\mathtt{6}\}^*, Node(x_1, \ldots, x_k) \text{ is defined}\} \\ &\quad = \{(x_k\mathtt{1}w)_7 \mid x_k \in \{3,6\}^*, w \in L_0\}, \end{aligned}$$

where the language

$$L_0 = \{(x_{k-1}\mathtt{1}\dots\mathtt{1}x_1\mathtt{10}w)_7 \mid k \geqslant 0, x_i \in \{\mathtt{3},\mathtt{6}\}^*,$$
$$\mathit{Node}(x_1,\dots,x_{k-1},x) \text{ is defined for some } x \in \{\mathtt{3},\mathtt{6}\}^*\}$$

represents the set of well-defined addresses of internal nodes in the tree. Thus, both assumptions of Lemma A are validated, and it asserts that (3c)–(3d) hold true, that is, that $A(\mathit{Remove}_{\mathtt{1}}(T_0)) \subseteq (\mathtt{1}\Gamma_7^*)_7$ and

$$E^{-1}\big(A(\mathit{Remove}_{\mathtt{1}}(T_0))\big) \subseteq \mathit{Remove}_{\mathtt{1}}(T_0)$$
$$\subseteq E^{-1}\big(A(\mathit{Remove}_{\mathtt{1}}(T_0)) \cup E(\mathit{Remove}_{\mathtt{1}}(T_1))\big).$$

Equation (3e) turns into an equality $T_0 \cup T_1 = \mathit{Paths}$, which is true by Lemma 8.

To see that the last equation (3f) is satisfied, consider that $T_0 \cap R_1 \subseteq T_0 \cap T_1 = \varnothing$ by Lemma 8, and similarly $T_1 \cap R_0 \subseteq T_1 \cap T_0 = \varnothing$. □

The second and the more difficult task is to demonstrate that every solution of the system must coincide with the given solution. The argument uses the well-founded induction on the structure of the tree. The membership of numbers of the form $(\mathtt{1}x_k\mathtt{1}x_{k-1}\dots\mathtt{1}x_1\mathtt{10}w)_7$, with $k \geqslant 0$, $x_i \in \{\mathtt{3},\mathtt{6}\}^*$ and $w \in \Gamma_7^* \setminus \mathtt{0}\Gamma_7^*$, in the variables X_0 and X_1 is first determined for larger k's and then inductively extended down to $k = 0$. Lemmata 10 and 11 are specifically designed to handle the induction step in this argument.

Lemma 13 *If $\mathit{Node}(x_1,\dots,x_k) = i$ is defined, then, for every solution of the system, and for every number $(\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7$ with $w \in \Gamma_7^* \setminus \mathtt{0}\Gamma_7^*$,*

1. *$(\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7$ is in X_0 if and only if $(w)_7$ is in B_i;*
2. *$(\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7$ is in X_1 if and only if $(w)_7$ is in C_i.*

Proof The proof proceeds by a well-founded induction on the index $i \in I$, with respect to the ordering $\prec$ on I. Each descending sequence of indices corresponds to a path in the tree of B_{i_0}, and all such paths are finite by Lemma 6, which justifies the use of the well-founded induction principle.

Induction basis

Consider an index minimal according to $\prec$, which is of the form $i = \mathit{Node}(x_1,\dots,x_k) = \tau_1(e)$ with $e \in \mathbb{N}$. The first claimed equivalence for X_0 and B_i is established as follows.

⊜ If $(w)_7 \in B_i$, then $(\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7 \in R_0$, and, by (3a), $(\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7 \in X_0$.

⊖ Conversely, if $(\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7 \in X_0$, then, by (3f), $(\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7 \notin R_1$, and accordingly $(w)_7 \notin C_i$, or, equivalently, $(w)_7 \in B_i$.

This proves the equivalence for X_0 in the base case of the induction. The other equivalence for X_1 is established by exactly the same argument.

Induction step

For the induction step, fix $x_1, \dots, x_k$, with $i = Node(x_1, \dots, x_k) = \tau_2(e)$ for some $e \in \mathbb{N}$. Assume that the claim of the lemma holds for all $i' = Node(x_1, \dots, x_k, x_{k+1})$. The task is to show that the same statement holds for i. Note that, under the induction assumption, Lemmata 10 and 11 are applicable.

Induction step: X_0 and infinite union

By (3a), $(\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}w)_7 \in X_0$ holds if and only if this number belongs to $E(Remove_1(X_1))$ or to R_0. The former is equivalent to $(w)_7 \in B_i$ by Lemma 10, while the latter is impossible because $Node(x_1, \dots, x_k) = \tau_2(e)$ for some e.

Induction step: X_1 and infinite intersection

Consider the following consequence of (3d), obtained by intersecting it with the set $(\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7$:

$$\begin{aligned}
&E^{-1}(Y) \cap (\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7 \\
&\quad\subseteq Remove_1(X_0) \cap (\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7 \\
&\quad\subseteq E^{-1}\big(Y \cup E(Remove_1(X_1)\big) \cap (\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7.
\end{aligned}$$

This equation shall now be equivalently transformed to match the form required by Lemma A. Consider, that for every set $S \subseteq \mathbb{N}$,

$$\begin{aligned}
&E^{-1}(S) \cap (\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7 = E^{-1}\big(S \cap (\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7\big), \\
&E(S) \cap (\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7 = E\big(S \cap (\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7\big) \quad \text{and} \\
&Remove_1(S) \cap (\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7 \\
&\quad = Remove_1\big(S \cap (\mathtt{1}\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7\big).
\end{aligned}$$

Using these identities, the above consequence of (3d) can be restated as

$$\begin{aligned}
&E^{-1}\big(Y \cap (\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7\big) \\
&\quad\subseteq Remove_1\big(X_0 \cap (\mathtt{1}\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7\big) \\
&\quad\subseteq E^{-1}\Big[\big(Y \cap (\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7\big) \\
&\qquad \cup E\big(Remove_1(X_1 \cap (\mathtt{1}\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7)\big)\Big],
\end{aligned}$$

which, introducing new variables

$$\begin{aligned}
&X_0' = Remove_1\big(X_0 \cap (\mathtt{1}\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7\big), \\
&X_1' = Remove_1\big(X_1 \cap (\mathtt{1}\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7\big) \quad \text{and} \\
&Y' = Y \cap (\mathtt{1}x_k \dots \mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7,
\end{aligned}$$

is written as follows:

$$E^{-1}(Y') \subseteq X'_0 \subseteq E^{-1}(Y' \cup E(X'_1)). \tag{5a}$$

Equation (3c) has a similar consequence:

$$Y' \subseteq (\mathtt{1}\Gamma_7^*)_7. \tag{5b}$$

The values of X'_0 and X'_1 are in fact uniquely determined. Consider, that $(x\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7 \in X'_0$ if and only if $(x\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7 \in \mathit{Remove}_\mathtt{1}(X_0)$, which is in turn equivalent to $(\mathtt{1}x\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7 \in X_0$. The latter, by the induction hypothesis, holds if and only if $(w)_7 \in B_{\mathit{Node}(x_1,\dots,x_k,x)}$. Hence,

$$X'_0 = \{(x\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7 \mid x \in \{\mathtt{3},\mathtt{6}\}^*, (w)_7 \in B_{\mathit{Node}(x_1,\dots,x_k,x)}\}, \quad \text{and}$$
$$X'_1 = \{(x\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7 \mid x \in \{\mathtt{3},\mathtt{6}\}^*, (w)_7 \in C_{\mathit{Node}(x_1,\dots,x_k,x)}\}$$

by a similar argument. These two sets are thereby disjoint, as so are $B_{\mathit{Node}(x_1,\dots,x_k,x)}$ and $C_{\mathit{Node}(x_1,\dots,x_k,x)}$ for any x. Furthermore, each number $(x\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7$ is either in X'_0 or in X'_1, depending on whether $(w)_7$ is in $B_{\mathit{Node}(x_1,\dots,x_k,x)}$ or in $C_{\mathit{Node}(x_1,\dots,x_k,x)}$, and thus the union of these two variables is exactly

$$X'_0 \cup X'_1 = (\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}(\Gamma_7^* \setminus \mathtt{0}\Gamma_7^*))_7.$$

This allows applying Lemma A to (5), and according to this lemma, the value of Y' is completely determined as $Y' = A(X'_0)$. In the original variables,

$$Y \cap (\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7, = A\big(\mathit{Remove}_\mathtt{1}\big(X_0 \cap (\mathtt{1}\{\mathtt{3},\mathtt{6}\}^*\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}\Gamma_7^*)_7\big)\big). \tag{6}$$

Using the latter equality, the induction step for X_1 is proved as follows. By (3b), a number $(\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7$ with $w \in \Gamma_7^* \setminus \mathtt{0}\Gamma_7^*$ is in X_1 if and only if it is in Y or in R_1. Since $\mathit{Node}(x_1,\dots,x_k) = \tau_2(e)$ for some e, the latter is impossible, and hence the statement is equivalent to $(\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7 \in Y$. By (6), this holds if and only if the number $(\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7$ belongs to $A(\mathit{Remove}_\mathtt{1}(X_0))$. The latter, by Lemma 11, is equivalent to $(w)_7 \in C_{\mathit{Node}(x_1,\dots,x_k)}$, as claimed. □

Now, Lemmata 12 and 13 together assert, that there is exactly one solution of the intended form:

Lemma 14 *The system (3a)–(3f) has a unique solution* $X_0 = T_0$, $X_1 = T_1$, $Y = A(\mathit{Remove}_\mathtt{1}(T_0))$.

Proof This assignment is a solution by Lemma 12.

Let X_0, X_1, Y be an arbitrary solution of the system. Equation (3e) ensures that every number in X_0 or in X_1 is of the form $(\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7$, with $\mathit{Node}(x_1,\dots,x_k)$ defined and with $w \in \Gamma_7^* \setminus \mathtt{0}\Gamma_7^*$. Then, by Lemma 13, $(\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7 \in X_0$ if and only if $(w)_7 \in B_{\mathit{Node}(x_1,\dots,x_k)}$, which in turn is equivalent to $(\mathtt{1}x_k\dots\mathtt{1}x_1\mathtt{10}w)_7 \in T_0$. The case of X_1 is proved by the same argument.

Thus, it is proved that $X_0 = T_0$ and $X_1 = T_1$. Finally, applying Lemma A with $S = X_0$ and $\widehat{S} = X_1$ to (3c)–(3d) proves that Y is fixed at $A(\mathit{Remove}_\mathtt{1}(T_0))$. □

4.5 Representing the Actual Set B_{i_0}

Besides the desired sets B_{i_0} and C_{i_0}, the sets T_0 and T_1, defined by the above system of equations, encode all sets on which B_{i_0} and C_{i_0} depend. Intersecting T_0 with the constant set $(\mathtt{10}\Gamma_7^*)_7$ produces the set $\{\mathtt{10}w \mid (w)_7 \in B_{i_0}\}$, and in order to obtain B_{i_0} as it is, one has to remove the leading digits $\mathtt{10}$ by the following function:

$$Remove_{\mathtt{10}}(X) = \{(w)_7 \mid w \in \Gamma_7^* \setminus \mathtt{0}\Gamma_7^*, (\mathtt{10}w)_7 \in X\}.$$

This function is implemented by a construction analogous to the one in Lemma 5.

Lemma 15 *The value of the expression*

$$\bigcup_{t\in\{0,1,2\}} \big[(X \cap (\mathtt{10}\Gamma_7^t(\Gamma_7^3)^* \setminus \mathtt{100}\Gamma_7^*)_7) \dot{-} (\mathtt{10}^*)_7\big] \cap (\Gamma_7^t(\Gamma_7^3)^* \setminus \mathtt{100}\Gamma_7^*)_7$$

on any $S \subseteq (\mathtt{10}(\Gamma_7^* \setminus \mathtt{0}\Gamma_7^*))_7$ *is* $Remove_{\mathtt{10}}(S) = \{(w)_7 \mid (\mathtt{10}w)_7 \in S\}$.

The expression $Remove_{\mathtt{10}}$ works generally similarly to $Remove_{\mathtt{1}}$, and it is intended to operate as follows:

$$\begin{array}{r} \mathtt{1}\ \mathtt{0}\ w_1\ w_2 \ldots w_0 \\ -\ \underline{\mathtt{1}\ \mathtt{0}\ \mathtt{0}\ \ \mathtt{0}\ \ldots\ \mathtt{0}} \\ w_1\ w_2 \ldots w_0 \end{array}$$

In this way, the correct subtraction reduces the number of digits by two, while an incorrect subtraction of $(\mathtt{10}^i)_7$ with $i < |\mathtt{0}w|$ may reduce the number of digits by one or leave it unchanged. Accordingly, the expression considers the cases of different number of digits modulo 3, rather than modulo 2, as in Lemma 5. In all other respects, the proof of Lemma 15 is the same as the proof of Lemma 5.

Theorem 2 *For every hyper-arithmetical set* $B \subseteq \mathbb{Z}$ ($B \subseteq \mathbb{N}$), *there is a system of equations over sets of integers* (*over sets of natural numbers*, *respectively*) *using union*, *addition*, *singleton constants and the constants* $\mathbb{N}$ *and* $-\mathbb{N}$ (*union*, *addition*, *subtraction and singleton constants*, *respectively*), *which has a unique solution with* B *as one of the components*.

Proof Assume first that $B \subseteq \mathbb{N}$. Let $B = B_{i_0}$ according to the enumeration of hyper-arithmetical sets, and construct the corresponding system of (3a)–(3f).

By Lemma 14, this system has a unique solution with the X_0-component

$$T_0 = \{(\mathtt{1}x_k\mathtt{1}x_{k-1}\ldots\mathtt{1}x_1\mathtt{10}w)_7 \mid k \geqslant 0, x_i \in \{\mathtt{3}, \mathtt{6}\}^*, (w)_7 \in B_{Node(x_1,\ldots,x_k)}\}.$$

Construct an additional equation

$$X = Remove_{\mathtt{10}}(X_0 \cap (\mathtt{10}\Gamma_7^*)_7).$$

Then its unique solution is

$$X = Remove_{\mathtt{10}}(\{(\mathtt{10}w)_7 \mid (w)_7 \in B_{Node(\langle\rangle)}\}) = B_{i_0}.$$

Thus the set B_{i_0} has been represented by a system of equations in the intermediate form required by Lemma 3, enhanced by Corollary 1 to allow recursively enumerable constants. According to the lemma, the set B_{i_0} can be represented by a system of equations over sets of natural numbers, using union, addition and subtraction, with singleton constants.

The set C_{i_0} is represented by applying similar operations to the set T_1.

For a hyper-arithmetical set of arbitrary integers, its positive and negative parts are first represented as shown above, and then Lemma 4 yields the system representing the whole set. □

This main result of the paper deserves being re-stated for language equations with the quotient operation, $K \cdot L^{-1} = \{u \mid \exists v \in L : uv \in K\}$.

Corollary 2 *For every hyper-arithmetical unary language $L \subseteq a^*$ there is a system of language equations using union, concatenation, quotient and constant $\{a\}$, such that $(L, \ldots)$ is its unique solution.*

5 Equations with Addition Only

It is known that 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 [10]. In order to define this encoding, for each $i \in \{0, 1, \ldots, 15\}$ 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

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

and the following result is known:

Proposition 5 ([10, Thm. 5.3]) *For every recursive set S there exists a system of equations over sets of natural numbers in variables $X, Y_1, \ldots, Y_m$ 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, as in Proposition 1, and then by representing addition and union of sets using addition of their σ_0-encodings.

Exactly the same methods turn out to be applicable to equations over sets of integers, which can represent a closely similar 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 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 16 (cf. [10, Lemma 3.3]) *A set $S \subseteq \mathbb{Z}$ satisfies an equation*

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

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

Next, 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 17 (cf. [10, 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:

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 $S = \{n \mid 16n \in T\}$.*

All proofs can be carried out using the same methods as in the case of equations over sets of natural numbers [10]. They are thereby omitted, and shall be properly presented in the full version of the cited paper.

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, testing *solution existence* is a Π^0_1-complete decision problem [22], and it remains Π^0_1-complete already in the case of a unary alphabet, concatenation as the only operation and regular constants [10], that is, for equations over sets of natural numbers with addition only. For the same formalisms, *solution uniqueness* is Π^0_2-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 6 (Rogers [25, Thm. 16-XX]) *Consider trees with nodes labeled by finite sequences of natural numbers, such that a node $(x_1, \dots, x_{k-1}, x_k)$ 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 would be 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:

Proposition 6.1 *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* $x_i \in \{3, 6\}^*$, *such that* M *accepts* $(1x_k1x_{k-1}\ldots 1x_11)_7$ *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.*

Proof For any fixed system of equations, the statement that it has a solution naturally belongs to Σ_1^1: taking the arithmetical formula $Eq(X_1, \ldots, X_n)$ from Lemma 1, it suffices to write a second-order statement

$$(\exists X_1)\ldots(\exists X_n)\ Eq(X_1, \ldots, X_n).$$

Furthermore, note that a given system can be effectively transformed to such a formula.

Consider, that the condition of a given closed Σ_1^1-formula's being true can be specified by a certain *universal* Σ_1^1*-formula* $\varphi(x)$, with $\varphi(n)$ true if and only if n is a number representing a true closed Σ_1^1-formula [25, Cor. 16-XX(a)]. This leads to a Σ_1^1-formula representing the existence of solution of a system given as an argument.

In order to prove that testing solution existence is Σ_1^1-hard, it is sufficient to reduce the problem from Proposition 6.1 to it. Let M be the given Turing machine. Since $L(M) \in \Sigma_1^0$, there is a system of equations over sets of integers in variables $Y, Y_1, \ldots, Y_m$, which has a unique solution with $Y = L(M)$, and this system can be effectively constructed from the description of M. Introducing an extra variable X, consider the following additional equations, where the expressions E and $Remove_1$ are taken from Lemma E and Lemma 5:

$$X \subseteq Y$$

$$\{1\} \subseteq X$$

$$X = E(Remove_1(X)))$$

The variable X represents a subset of Y containing the set of finite prefixes of one or more infinite sequences. The claim is that this system has a solution if and only if there exists an infinite sequence $x_1, x_2, \ldots, x_k, \ldots$, such that each number $(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{1})_7$, for all $k \geqslant 0$, is accepted by M.

⇒ Assume, that the system has a solution. Then, an infinite sequence $x_1, \ldots, x_k, \ldots$, with $(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{1})_7 \in X$ for each $k \geqslant 0$, is constructed inductively as follows. The base case is that the unique element with $k = 0$ is defined, and it is ensured by the equation $\{1\} \subseteq X$. Assume, that the elements are defined up to $k \geqslant 0$. Then, $(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{1})_7 \in X = E(\mathit{Remove}_1(X))$. As $E(\mathit{Remove}_1(X)) = \{\mathtt{1}w \mid \exists x(\mathtt{1}x\mathtt{1}w)_7 \in X\}$, there exists x with $(\mathtt{1}x\mathtt{1}x_k\mathtt{1}\ldots, \mathtt{1}x_1)_7 \in X$. Let $x_{k+1} = x$. Since $X \subseteq Y = L(M)$, the number $(\mathtt{1}x\mathtt{1}x_k\mathtt{1}\ldots, \mathtt{1}x_1)_7$ is accepted by M.

⇐ Conversely, assume that there is an infinite sequence $x_1, x_2, \ldots, x_k, \ldots$, such that each $(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{1})_7$, for all $k \geqslant 0$, is accepted by M. Then let $X = \{(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{1})_7 \mid k \geqslant 0\}$ be the set of finite prefixes of this particular sequence. This X, together with $Y = L(M)$, forms a solution of the constructed system. Indeed,

$$\begin{aligned}
E(\mathit{Remove}_1(X))) &= E\big(\mathit{Remove}_1(\{(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{1})_7 \mid k \geqslant 0\})\big) \\
&= E\big(\{(x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{1})_7 \mid k \geqslant 0\}\big) \\
&= \{(\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{1})_7 \mid k \geqslant 1\} \\
&= \{(\mathtt{1}x_k\mathtt{1}\ldots\mathtt{1}x_1\mathtt{1})_7 \mid k \geqslant 0\} \\
&= X,
\end{aligned}$$

and the rest of the equations clearly hold, as $X \subseteq Y$ and, by the construction, $1 \in X$. Thus the system has a solution. □

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

Proposition 7 *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.*

Proof The property of having at most one solution can be expressed by the following Π^1_1-formula:

$$\begin{aligned}
&(\forall X_1)\ldots(\forall X_n)(\forall X'_1)\ldots(\forall X'_n)\big[Eq(X_1, \ldots, X_n) \wedge Eq(X'_1, \ldots, X'_n)\big] \\
&\quad \to (\forall n)(\forall i)\big(n \in X_i \leftrightarrow n \in X'_i\big)
\end{aligned}$$

Then the condition of having a unique solution is a conjunction of the latter formula with the Σ^1_1-formula expressing solution existence. The resulting conjunction can be reformulated both as a Σ^1_2-formula and as a Π^1_2-formula. □

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.

Table 1 Summary of the results

	Sets representable by unique solutions	Complexity of decision problems	
		solution existence	solution uniqueness
over $2^{\mathbb{N}}$, with $\{+, \cup\}$	Δ^0_1 (recursive) [9]	Π^0_1-complete [9]	Π^0_2-complete [9]
over $2^{\mathbb{N}}$, with $\{+\}$	encodings of Δ^0_1 [10]	Π^0_1-complete [10]	Π^0_2-complete [10]
over $2^{\mathbb{N}}$, with $\{+, \dot{-}, \cup\}$	Δ^1_1 (hyper-arithmetical)	Σ^1_1-complete	in Δ^1_2
over $2^{\mathbb{Z}}$, with $\{+, \cup\}$	Δ^1_1	Σ^1_1-complete	in Δ^1_2
over $2^{\mathbb{Z}}$, with $\{+\}$	encodings of Δ^1_1	Σ^1_1-complete	in Δ^1_2

7 Conclusion and Open Problems

The paper has determined the natural limit of the expressive power of language equations involving erasing operations. Just like the recursive sets are the natural upper bound for equations with continuous operations [22], and this upper bound is reached by ultimately simple specimens of such equations [9, 10, 16], the hyper-arithmetical sets, which might have looked as a very rough upper bound, have been found representable by equations with the simplest sets of erasing operations. In addition, these simple equations can be regarded as a basic arithmetical object representing an important variant of formal arithmetic.

One natural direction of the future work is to consider representability of sets by least and greatest solutions of equations over sets of integers, and this has already been attempted by the authors [12].

There is an important question left unanswered in this paper: What is the exact complexity of the solution uniqueness problem for equations over sets of integers? In particular, is it Σ^1_1-hard or Π^1_1-hard?

Acknowledgements We are grateful to the anonymous referees for pointing out several issues with the presentation and for suggesting useful improvements. A. Jeż supported by MNiSW grant N206 492638 2010–2012 and by a grant funded by Human Capital Programme. A. Okhotin supported by the Academy of Finland under grant 134860.

References

1. Aczel, P.: An introduction to inductive definitions. In: Barwise, J. (ed.) Handbook of Mathematical Logic, pp. 739–783. North-Holland, Amsterdam (1977)
2. d'Alessandro, F., Sakarovitch, J.: The finite power property in free groups. Theor. Comput. Sci. **293**(1), 55–82 (2003)
3. Anisimov, A.V.: Languages over free groups. In: Mathematical Foundations of Computer Science (MFCS 1975), Mariánské Lázně, September 1–5, 1975. LNCS, vol. 32, pp. 167–171 (1975)
4. Ginsburg, S., Rice, H.G.: Two families of languages related to ALGOL. J. ACM **9**, 350–371 (1962)

5. Halpern, J.Y.: Presburger arithmetic with unary predicates is Π_1^1 complete. J. Symb. Log. **56**(2), 637–642 (1991)
6. Jeż, A.: Conjunctive grammars can generate non-regular unary languages. Int. J. Found. Comput. Sci. **19**(3), 597–615 (2008)
7. Jeż, A., Okhotin, A.: Conjunctive grammars over a unary alphabet: undecidability and unbounded growth. Theory Comput. Syst. **46**(1), 27–58 (2010)
8. Jeż, A., Okhotin, A.: Complexity of equations over sets of natural numbers. Theory Comput. Syst. **48**(2), 319–342 (2011)
9. Jeż, A., Okhotin, A.: On the computational completeness of equations over sets of natural numbers. In: 35th International Colloquium on Automata, Languages and Programming (ICALP 2008), Reykjavik, Iceland, July 7–11, 2008. LNCS, vol. 5126, pp. 63–74 (2008)
10. Jeż, A., Okhotin, A.: Equations over sets of natural numbers with addition only. In: STACS 2009, Freiburg, Germany, 26–28 February 2009, pp. 577–588 (2009)
11. Jeż, A., Okhotin, A.: One-nonterminal conjunctive grammars over a unary alphabet. Theory Comput. Syst. **49**(2), 319–342 (2011)
12. Jeż, A., Okhotin, A.: Least and greatest solutions of equations over sets of integers. In: Mathematical Foundations of Computer Science (MFCS 2010), Brno, Czech Republic, 23–27 August 2010. LNCS, vol. 6281, pp. 441–452 (2010)
13. Kunc, M.: The power of commuting with finite sets of words. Theory Comput. Syst. **40**(4), 521–551 (2007)
14. Kunc, M.: What do we know about language equations? In: Developments in Language Theory (DLT 2007), Turku, Finland, July 3–6, 2007. LNCS, vol. 4588, pp. 23–27 (2007)
15. Lehtinen, T., Okhotin, A.: On equations over sets of numbers and their limitations. Int. J. Found. Comput. Sci. **22**(2), 377–393 (2011)
16. Lehtinen, T., Okhotin, A.: On language equations $XXK = XXL$ and $XM = N$ over a unary alphabet. In: Developments in Language Theory (DLT 2010), London, Ontario, Canada, 17–20 August 2010. LNCS, vol. 6224, pp. 291–302 (2010)
17. McKenzie, P., Wagner, K.: The complexity of membership problems for circuits over sets of natural numbers. Comput. Complex. **16**(3), 211–244 (2007)
18. Moschovakis, Y.: Elementary Induction on Abstract Structures. North-Holland, Amsterdam (1974)
19. Okhotin, A.: Conjunctive grammars. J. Autom. Lang. Comb. **6**(4), 519–535 (2001)
20. Okhotin, A.: Unresolved systems of language equations: expressive power and decision problems. Theor. Comput. Sci. **349**(3), 283–308 (2005)
21. Okhotin, A.: Computational universality in one-variable language equations. Fundam. Inform. **74**(4), 563–578 (2006)
22. Okhotin, A.: Decision problems for language equations. J. Comput. Syst. Sci. **76**(3–4), 251–266 (2010)
23. Okhotin, A., Rondogiannis, P.: On the expressive power of univariate equations over sets of natural numbers. In: IFIP International Conference on Theoretical Computer Science (TCS 2008), Milan, Italy, 8–10 September 2008. IFIP, vol. 273, pp. 215–227 (2008)
24. Robinson, J.: An introduction to hyperarithmetical functions. J. Symb. Log. **32**(3), 325–342 (1967)
25. Rogers, H., Jr.: Theory of Recursive Functions and Effective Computability. McGraw-Hill, New York (1967)
26. Travers, S.D.: The complexity of membership problems for circuits over sets of integers. Theor. Comput. Sci. **369**(1–3), 211–229 (2006)