Least and greatest solutions of equations over sets of integers

Artur Jeż[1] ⋆ and Alexander Okhotin[2,3] ⋆⋆

[1] Institute of Computer Science, University of Wrocław, Poland
aje@cs.uni.wroc.pl
[2] Department of Mathematics, University of Turku, Finland
alexander.okhotin@utu.fi
[3] Academy of Finland

Abstract. Systems of equations with sets of integers as unknowns are considered, with the operations of union, intersection and addition of sets, $S + T = \{m + n \mid m \in S,\ n \in T\}$. These equations were recently studied by the authors ("On equations over sets of integers", *STACS 2010*), and it was shown that their unique solutions represent exactly the hyperarithmetical sets. In this paper it is demonstrated that greatest solutions of such equations represent exactly the Σ^1_1 sets in the analytical hierarchy, and these sets can already be represented by systems in the *resolved form* $X_i = \varphi_i(X_1, \ldots, X_n)$. Least solutions of such resolved systems represent exactly the recursively enumerable sets.

1 Introduction

Consider equations $\varphi(X_1, \ldots, X_n) = \psi(X_1, \ldots, X_n)$, in which the unknowns X_i are sets of integers, and the expressions φ, ψ may contain addition $S + T = \{m + n \mid m \in S,\ n \in T\}$, Boolean operations and ultimately periodic constants. At a first glance, they might appear as a simple arithmetical object. However, already their simple special case, *expressions* and *circuits* over sets of integers, have a non-trivial computational complexity, studied by McKenzie and Wagner [10] in the case of nonnegative integers and by Travers [18] in the case of all integers.

If only nonnegative integers are allowed in the equations, they become isomorphic to *language equations* [8] over a one-letter alphabet. Language equations over multiple-letter alphabets are known to be computationally complete [15,14]: their unique solutions represent exactly the recursive sets, while their least and greatest solutions represent exactly the recursively enumerable sets and their complements, respectively. This result has been subsequently re-created by the authors [4,5] for the one-letter case, that is, for equations over sets of natural

⋆ Supported by MNiSW grants N206 259035 2008–2010 and N206 492638 2010-2012 and by personal grant FNP START of Polish Foundation for Science.
⋆⋆ Supported by the Academy of Finland under grant 134860.

numbers. As recently shown by Lehtinen and Okhotin [9], this computational universality extends to systems of such a simple form as $\{X+X+C = X+X+D, X+E = F\}$, with a unique unknown X.

The first study of equations over sets of integers, both positive and negative, was recently conducted by the authors [6]. The main result was that a set is representable by a unique solution of such a system if and only if it is *hyper-arithmetical*. Hyper-arithmetical sets are defined as the intersection $\Sigma_1^1 \cap \Pi_1^1$ of the two bottom classes of the analytical hierarchy, and are accordingly a proper superset of the sets representable in first-order Peano arithmetic. The results on unique solutions of such systems are recalled and commented in Section 2. Concerning least and greatest solutions of these equations, one can easily see that they must belong to Π_1^1 and to Σ_1^1, respectively, though no lower bounds are yet known.

This paper begins the study of least and greatest solutions of equations over sets of integers with systems of the following form:

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

This is the same general form as in the most well-known kind of language equations used to define context-free grammars [1]. It is known that such a system has a least solution corresponding to the context-free derivation; it is a folklore knowledge that greatest solutions are context-free as well. Least and greatest solutions are obtained by the fixpoint iteration, in which a solution is always reached after ω iterations. These results extend to a natural generalization of the context-free grammars, the *conjunctive grammars* [11,12].

In this paper, the unknowns in a system (*) are sets of integers, and the operations are union, intersection and addition. Tarski's fixpoint theorem [17] guarantees the existence of a least and a greatest solution, and, as explained in Section 3, an iterative version of Tarski's theorem asserts that a fixpoint is always reachable in ω_1 iterations, that is, iterating over countable ordinals. In Section 4 it is shown that in the case of greatest solutions, all ω_1 iterations are actually used, and that every set in Σ_1^1 can be represented by a greatest solution of such a system. On the other hand, Section 5 demonstrates that least solutions can be always reached in only ω iterations, and the family of sets represented by these solutions is exactly the family of recursively enumerable sets.

2 Equations over sets of integers

Consider systems of equations of the *resolved form* $X_i = \varphi_i(X_1, \ldots, X_n)$ with $i \in \{1, \ldots, n\}$, where the unknowns X_i are sets of integers, and the expressions φ_i may use the operations of union, intersection and addition of sets, as well as ultimately periodic constants[4]. When such a system has a unique solution, it

[4] A set of integers $S \subseteq \mathbb{Z}$ is *ultimately periodic* if there exist such numbers $n_0 \geqslant 0$ and $p \geqslant 1$, that $n \in S$ if and only if $n + p \in S$ for all n with $|n| \geqslant n_0$.

can be regarded as a definition of the sets in that solution. When a system of this form has multiple solutions, it is known from Tarski's fixpoint theorem [17] that among them there is the *least* and the *greatest solution* with respect to the partial order of componentwise inclusion.

If the unknowns are sets of natural numbers, such equations were first studied by Jeż [2], who established their nontriviality by representing the set $\{4^n \mid n \geqslant 0\}$:

Example 1 (Jeż [2]). The system of equations

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

over sets of natural numbers has a least solution with $X_1 = \{4^n \mid n \geqslant 0\}$, $X_2 = \{2 \cdot 4^n \mid n \geqslant 0\}$, $X_3 = \{3 \cdot 4^n \mid n \geqslant 0\}$ and $X_6 = \{6 \cdot 4^n \mid n \geqslant 0\}$.

To understand this construction, it is useful to consider positional notation of numbers. Let $\Gamma_k = \{\mathtt{0}, \mathtt{1}, \ldots, k-1\}$ be digits in base-k notation. For every $w \in \Gamma_k^*$, let $(w)_k$ be the number defined by this string of digits. For a language $L \subseteq \Gamma_k^*$ of positional notations, define $(L)_k = \{(w)_k \mid w \in L\}$. Now the solution of the above system can be conveniently represented in base-4 notation as $\big((\mathtt{10}^*)_4, (\mathtt{20}^*)_4, (\mathtt{30}^*)_4, (\mathtt{120}^*)_4\big)$. Substituting these four sets into the first equation, one obtains

$$\big((\mathtt{10}^*)_4 + (\mathtt{30}^*)_4\big) \cap \big((\mathtt{20}^*)_4 + (\mathtt{20}^*)_4\big) = \\ = \big((\mathtt{10}^+)_4 \cup (\mathtt{10}^*\mathtt{30}^*)_4 \cup (\mathtt{30}^*\mathtt{10}^*)_4\big) \cap \big((\mathtt{10}^+)_4 \cup (\mathtt{20}^*\mathtt{20}^*)_4\big) = (\mathtt{10}^+)_4,$$

that is, both sums contain some "garbage", yet the garbage in the sums is disjoint, and is accordingly "filtered out" by the intersection. Finally, the union with $\{1\}$ yields the set $\{4^n \mid n \geqslant 0\}$, turning the first equation into an equality. The rest of the equations are verified similarly [2].

The idea of this example was generalised by the authors [3] by representing every set of numbers with their positional notation recognised by a certain kind of cellular automata. These are one-way real-time cellular automata, known under a proper name of *trellis automata* [13].

Proposition 1 (Jeż, Okhotin [3, Thm. 3]). *For every $k \geqslant 2$ and for every trellis automaton M over $\Gamma_k = \{\mathtt{0}, \ldots, k-1\}$, such that $L(M) \cap \mathtt{0}\Gamma_k^* = \varnothing$, there exists and can be effectively constructed a resolved system of equations over sets of natural numbers using the operations of union, intersection and addition and singleton constants, such that its least solution contains a component $(L(M))_k$.*

Trellis automata are notable, in particular, for recognising the language of *computation histories of a Turing machine*, which is generally defined in the form $\mathrm{VALC}(T) = \{C_T(w)\natural w \mid w \in L(T)\}$, where $C_T(w)$ is a sequence of consecutive configurations in the accepting computation of T on w, encoded in a suitable way. This follows from the fact that trellis automata can recognise any finite intersections of linear context-free languages, and VALC(T) is representable as such an intersection. Assume that VALC(T) is defined over an alphabet of k-ary digits Γ_k. Then, any computation represents a number $(C_T(w)\natural w)_k$, and Proposition 1 asserts that the set of such numbers is a solution of some system of equations [3]. A more complicated construction on top of $(\mathrm{VALC}(T))_k$ allows *extracting* $(L)_k$ *out of* VALC(T), leading to a representation of every recursive (r.e., co-r.e.) set by unique (least, greatest, respectively) solution of a system $\varphi_i(X_1, \ldots, X_n) = \psi_i(X_1, \ldots, X_n)$ over sets of natural numbers [4].

When constructing equations over sets of integers, applying Proposition 1 to VALC(T) remains a useful technique. As in the authors' previous work on systems of equations over sets of integers [6], VALC(T) shall be defined over the alphabet of digits in base-7 notation, with each computation encoded by a string $C_T(w) \in \{\mathtt{3}, \mathtt{6}\}^+$, and with

$$\mathrm{VALC}(T) = \{C_T(w)\mathtt{1}w \mid w \in T\}.$$

The exact details of the encoding are not important, as trellis automata are flexible enough to recognise such a variant of VALC(T). Then the corresponding set of numbers

$$\{(C_T(w)\mathtt{1}w)_7 \mid (w)_7 \in L(T)\}$$

is representable by the unique solution of a resolved system of equations over sets of natural numbers with union, intersection and addition [3, Thm. 3]. If every occurrence of every variable X is replaced with $X \cap (\mathbb{N} + 1)$, the system will have the same unique solution if interpreted over sets of integers.

Using equations over sets of integers, the set $(L(T))_7$ can be obtained out of $(\mathrm{VALC}(T))_7$ generally by subtracting the computation history from each number in VALC(T) as follows: $(C_T(w)\mathtt{1}w)_7 - (C_T(w)\mathtt{1}\mathtt{0}^{|w|})_7 = (w)_7$. This has to be done by adding a set of negative numbers to VALC(T), and filtering out numbers of the form $(C_T(w)\mathtt{1}w)_7 - (x)_7$ with $x \neq (C_T(w)\mathtt{1}\mathtt{0}^{|w|})_7$. Since $C_T(w)$ is a string of digits 3 and 6, this subtraction can be regarded as the removal of the prefix $\{\mathtt{3}, \mathtt{6}\}^+$, or as an existential quantification over such prefixes:

Lemma 1 (Representing the existential quantifier [6, Lemma E]). *The value of the expression*

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

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

Then $E(\mathrm{VALC}(T)) = \{(\mathtt{1}w)_7 \,|\, w \in L(T)\}$, and it is left to remove the leading digit 1, which is performed by the expression in the next lemma:

Lemma 2 (Removing leading digit 1 [6]). *The value of the expression*

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

on any $S \subseteq (\mathtt{1}(\Gamma_7^+ \setminus \mathtt{0}\Gamma_7^*))_7$ *is* $\{(w)_7 \mid (\mathtt{1}w)_7 \in S\}$.

The two above lemmata yield a representation of r.e. sets:

Theorem 1. *Every r.e. set $S \subseteq \mathbb{Z}$ is the unique solution of a resolved system of equations over sets of integers using union, intersection and addition, as well as singleton constants and the constants $\mathbb{N}$, $-\mathbb{N}$.*

Proof (sketch). Assume first that $S \subseteq \mathbb{N}$ and let T be a Turing machine accepting S. Then, as long as the constant $\mathrm{VALC}(T)$, and the constants in Lemmata 1 and 2 are given, the expression $\mathit{Remove}_{\mathtt{1}}(E(\mathrm{VALC}(T)))$ yields the set S.

The constant $\mathrm{VALC}(T) \subseteq \mathbb{N}$, as well as the constant sets of natural numbers in the Lemmata, are representable by equations over sets of natural numbers by Proposition 1. This construction is replicated for equations over sets of integers, by applying an intersection with a constant $\mathbb{N}$. The constant sets of negative integers in Lemmata 1 and 2 are represented as if the sets of the opposite numbers, negating all constants in the system.

This construction can be applied to any r.e. set of negative integers by representing the set of opposite numbers as above, and then by replacing every constant C by $-C$. Finally, any r.e. set of integers $S \subseteq \mathbb{Z}$ is represented as a union of its positive and negative subsets. ◻

The natural counterpart of the "existential quantifier" $E(X)$ is the function $A(X)$, defined as $A(S) = \{(\mathtt{1}w)_7 \mid \forall x \in \{\mathtt{3},\mathtt{6}\}^* \ (x\mathtt{1}w)_7 \in S\}$. Equations of the general form $\varphi_i(X_1, \ldots, X_n) = \psi(X_1, \ldots, X_n)$ representing $A(X)$ were constructed by the authors [6]. Then, applying $A(X)$ and $E(X)$ to a recursive set finitely many times allowed constructing every set from the arithmetical hierarchy, and doing this iteratively led to the representation of every hyper-arithmetical set as a unique solution of such a system [6]. Intuitively, that system implemented an equation $X = A(E(X)) \cup C$, for a recursive constant $C \subseteq ((\mathtt{1}\{3,6\}^+)^*\mathtt{10}\Gamma_7^*)_7$, in which the digit blocks $\{3,6\}^+$ correspond to the quantified variables, $\mathtt{1}$ is a separator, while $\mathtt{10}$ marks the end of the quantifier prefix. Processing the latter requires an extra equation:

Lemma 3 (Removing leading digits $\mathtt{10}$ [6]). *The value of the expression*

$$\mathit{Remove}_{\mathtt{10}}(Z) = (Z \cap \{(\mathtt{10})_7\} - \{(\mathtt{10})_7\}) \cup \bigcup_{i \in \Gamma_7 \setminus \{0\}} \bigcup_{t \in \{0,1,2\}} (Z \cap (\mathtt{10}i\Gamma_7^t(\Gamma_7^3)^*)_7) - (\mathtt{10}^*)_7 \cap (i\Gamma_7^t(\Gamma_7^3)^*)_7$$

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

Proposition 2 (Jeż, Okhotin [6, Thm. 2]). *For every hyper-arithmetical set $S \subseteq \mathbb{Z}$ there is a system of equations over subsets of $\mathbb{Z}$ using union, addition, singleton constants and the constants $\mathbb{N}$ and $-\mathbb{N}$, with a unique solution $(S, \ldots)$.*

This representation result has a matching upper bound: whenever such a system has a unique solution, it is a hyper-arithmetical set [6]. The proof of this upper bound can actually be split into two statements: first, least solutions are demonstrated to be in the class Π_1^1, and second, greatest solutions always belong to Σ^1_1. As unique solutions are both least and greatest at the same time, they are in the class $\Pi^1_1 \cap \Sigma^1_1 = \Delta^1_1$, that is, are hyper-arithmetical. These bounds are based upon the following translation of equations into an arithmetical formula:

Proposition 3 (Jeż, Okhotin [6]). *For every system of equations in variables $X_1, \ldots, X_n$ using operations expressible in first-order arithmetic there exists an arithmetical formula $Eq(X_1, \ldots, X_n)$, where $X_1, \ldots, X_n$ are free second-order variables, such that $Eq(S_1, \ldots, S_n)$ is true if and only if $X_i = S_i$ is a solution of the system.*

Constructing this formula is only a matter of reformulation. As an example, an equation $X_i = X_j + X_k$ is represented by

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

Applying existential quantification to the set variables produces a Σ^1_1-formula $\varphi(x) = (\exists X_1)\ldots(\exists X_n)Eq(X_1, \ldots, X_n) \wedge (x \in X_1)$ representing the greatest solution, while universal quantification leads to a Π^1_1-formula $\varphi'(x) = (\forall X_1)\ldots(\forall X_n)Eq(X_1, \ldots, X_n) \to (x \in X_1)$ for the least solution:

Proposition 4. *For every system of equations in variables $X_1, \ldots, X_n$ using operations expressible in first-order arithmetic that has a least (greatest) solution $X_i = S_i$, the sets S_i are in the class Π^1_1 (in Σ^1_1, respectively).*

3 Resolved systems and their properties

A system of equations is called *explicit* or *resolved* if it is of the form

$$X_i = \varphi_i(X_1, \ldots, X_n) \quad (1 \leqslant i \leqslant n). \tag{1}$$

When the unknowns are formal languages, such equations are used to define the context-free grammars and their generalization, the conjunctive grammars [11].

It is convenient to regard (1) as a single equation $X = \varphi(X)$, where X is an unknown n-tuple of sets, while $\varphi = (\varphi_1, \ldots, \varphi_n)$ is an operator on the set of such n-tuples. A solution of such equation is known as a *fixpoint* of the operator φ. As long as φ is *monotone* under some partial ordering, that is, if

$$A \preccurlyeq A' \Longrightarrow \varphi(A) \preccurlyeq \varphi(A'),$$

a least and a greatest fixpoint exists by Tarski's [17] theorem.

In case of vectors of sets of integers, the partial ordering is defined by $(S_1, \ldots, S_n) \sqsubseteq (T_1, \ldots, T_n)$ if $S_i \subseteq T_i$ for each i. The operations of union, intersection and addition are all monotone with respect to this ordering.

Another general property of operators is continuity. A sequence of sets $\{A_n\}_{n\geqslant 0}$ is *convergent* if for every element $x \in \bigcup_n A_n$ the set $\{n | x \in A_n\}$ is either finite or co-finite; in such a case $\lim_{n\to\infty} A_n = \{x \mid x \text{ is in infinitely many } A_n\text{'s}\}$. Now φ is *continuous*, if for every convergent sequence $\{A_n\}_{n=1}^{\infty}$,

$$\lim_{n\to\infty} \varphi(A_n) = \varphi(\lim_{n\to\infty} A_n).$$

A composition of monotone (continuous) operators is monotone (continuous). Provided that a system (1) has monotone and continuous right-hand sides, its least solution is reached by ω iterations of φ, beginning with a vector of empty sets: $\bigsqcup_{k=1}^{\infty} \varphi^k(\varnothing, \ldots, \varnothing)$. If the iteration begins with the top element $(\overline{\varnothing}, \ldots, \overline{\varnothing})$, then the greatest solution is similarly reached after ω steps of a similar iteration, with intersection instead of union. This is the case with language equations using concatenation, union and intersection [1,12], or similar equations over sets of natural numbers [3].

However, when equations over sets of *integers* are considered (that is, if negative numbers are allowed), the addition of such sets is no longer continuous: consider $\varphi(X) = X + X$ and a sequence $X_n = \{-n, n\}$. Then $\lim_{n\to\infty} X_n = \varnothing$ and $\varphi(\lim_{n\to\infty} X_n) = \varnothing$. On the other hand, $0 \in X_n + X_n$ for each n, and accordingly $0 \in \lim_{n\to\infty} \varphi(X_n)$. This makes the above ω-step fixpoint iteration inapplicable to such systems, as the vector obtained after ω steps need not be a solution.

When all is known about a system (1) is that its right-hand sides are monotone, Tarski's [17] fixpoint theorem asserts that it has a least and a greatest solution. This result can be shown using a transfinite induction as follows. Denote by ω_1 the first uncountable ordinal. For each ordinal $\alpha \leqslant \omega_1$, define the vector of sets after α iterations of φ:

$$S^{(0)} = (\varnothing, \ldots, \varnothing) \tag{2a}$$

$$S^{(\alpha+1)} = \varphi(S^{(\alpha)}) \tag{2b}$$

$$S^{(\alpha)} = \bigsqcup_{\gamma<\alpha} S^{(\gamma)} \quad \text{when } \alpha \text{ is a limit ordinal} \tag{2c}$$

Lemma 4. *$S^{(\omega_1)}$ is the least fixpoint of the system* (1).

The proof proceeds along the following steps. First it is shown that $S^{(\alpha)}$ is a weakly increasing sequence, that is, $S^{(\alpha)} \sqsubseteq S^{(\gamma)}$ for all ordinals $\alpha < \gamma$. Then it is proved that there are countably many ordinals α with $S^{(\alpha)} \sqsubset S^{(\alpha+1)}$, and accordingly the sequence converges to a fixpoint in fewer than ω_1 steps. This fixpoint is then proved to be the least. All arguments are by a transfinite induction on the ordinals.

Similarly, for the greatest solutions define

$$T^{(0)} = (\mathbb{Z}, \ldots, \mathbb{Z}) \tag{3a}$$

$$T^{(\alpha+1)} = \varphi(T^{(\alpha)}) \tag{3b}$$

$$T^{(\alpha)} = \bigsqcap_{\gamma<\alpha} T^{(\gamma)} \quad \text{when } \alpha \text{ is a limit ordinal} \tag{3c}$$

Lemma 5. *$T^{(\omega_1)}$ is the greatest fixpoint of the system* (1).

4 Greatest solutions

The greatest solution of any system of equations over sets of integers is in Σ_1^1 in the analytical hierarchy. It shall now be proved that, conversely, every set $S \subseteq \mathbb{N}$ in Σ_1^1 is representable. The construction combines the definition of a certain Σ_1^1-complete set $\mathcal{T} \subseteq \mathbb{N}$ with a reduction function from S to $\mathcal{T}$. Representing any Σ_1^1-subset of $\mathbb{Z}$ is achieved by a simple additional step.

The announced Σ_1^1-hard set contains the yes-instances of the following problem (its complement is Π_1^1-complete [16, Thm. 16-XX]): "Given a Turing machine M working on natural numbers, determine whether there exists an infinite sequence of strings $\{x_i\}_{i=1}^{\infty}$ with $x_i \in \{\mathtt{3}, \mathtt{6}\}^+$, such that, for all $k \geqslant 0$, the number $(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{1})_7$ is in $L(M)$". Base-7 notations of these numbers encode finite sequences of natural numbers, and are formatted for processing by Lemmata 1 and 2. Now for any Σ_1^1 set S there exists a total recursive reduction function f_S, such that

$$n \in S \iff \exists\{x_i\}_{i=1}^{\infty} \forall k \geqslant 0 \ (\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{1})_7 \in L(M_{f_S(n)}),$$

where $M_0, M_1, \ldots, M_i, \ldots$ is any effective enumeration of Turing machines.

Fix S and its reduction f_S witnessing $S \leqslant_{rec} \mathcal{T}$. Define the set

$$C = \big\{(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{10}s)_7 \;\big|\; s \in \Gamma_7^* \setminus \mathtt{0}\Gamma_7^*, \\ \forall k' \leqslant k \ (\mathtt{1}x_{k'}\mathtt{1}x_{k'-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{1})_7 \in L(M_{f_S((s)_7)})\big\},$$

which is r.e.: given a number $(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{10}s)_7$, a Turing machine calculates its base-7 notation, extracts $(s)_7$, constructs $M_{f_S((s)_7)}$ and simulates it on each input $(\mathtt{1})_7, \ldots, (\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{1})_7$. If they are all accepted, this number belongs to C. By Theorem 1, C can be represented as a unique (and, in particular, the greatest) solution of a resolved system of equations.

For any fixed number $(s)_7 \in \mathbb{N}$, the set C induces a set of finite sequences

$$\big\{(n_1, \ldots, n_{k-1}, n_k) \;\big|\; (\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{10}s)_7 \in C, \text{ where each } x_i \text{ represents} \\ \text{the binary notation of } n_i, \text{ using } \mathtt{3} \text{ for zero and } \mathtt{6} \text{ for one}\big\}.$$

This set of sequences is closed under taking prefixes, and thus may be regarded as a *tree*. Each sequence is a *node* of the tree. A node $(n_1, n_2, \ldots, n_{k-1}, n_k)$ is a *child* of the node $(n_1, n_2, \ldots, n_{k-1})$, which is its *parent*. The empty sequence is the unique node without a parent, that is, the *root* of the tree; a node is a *leaf* if it has no children. A tree has an *infinite path* if there exists such a sequence $(n_1, n_2, \ldots, n_k, \ldots)$ that all of its finite prefixes belong to the tree. This tree terminology shall be adopted for a fixed $(s)_7$ when referring to C: for example, $(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{10}s)_7 \in C$ is the parent of $(\mathtt{1}x_{k+1}\mathtt{1}x_k\mathtt{1}\ldots\mathtt{1}x_1\mathtt{10}s)_7 \in C$, etc.

In this terminology, an element $(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{10}s)_7 \in C$ is said to *have an infinite path* if the tree corresponding to s has an infinite path beginning with the node corresponding to this element; or, equivalently, if

$$\exists\{x_{k+i}\}_{i=0}^{\infty} \forall \ell \geqslant 0 \ (\mathtt{1}x_{k+\ell}\mathtt{1}x_{k+\ell-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{10}s)_7 \in C$$

In particular, a number $(s)_7$ is in S if and only if the element $(\mathtt{10}s)_7$ has an infinite path. The goal is to construct an equation with the greatest solution comprised exactly of numbers with an infinite path. Since the greatest solution is a limit of a descending chain of sets, see Lemma 5 and (3), the equation shall iteratively shorten finite paths, so that the numbers without an infinite path are eventually eliminated.

For every node with finitely many descendants there is a well-defined *height* of its subtree. This concept is generalised to trees with infinite paths and infinite degrees of nodes as follows. The *rank* of an element of C, see Rogers [16, §16], is an ordinal defined by

$$r(x) = \begin{cases} 1, & \text{if } x \text{ is a leaf,} \\ \sup\{r(y)+1 \mid y \text{ is a child of } x\}, & \text{otherwise.} \end{cases} \tag{4a}$$

For some elements of C the recursion does not terminate, and the definition is extended by

$$r(x) = \omega_1, \quad \text{when } r(x) \text{ is not defined by (4a).} \tag{4b}$$

Lemma 6. *The rank of an element* $(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{10}s)_7 \in C$ *is not defined by* (4a) *if and only if it has an infinite path.*

As argued by Rogers [16, Thm. 16-XVIII(a)], all ordinals assigned by (4a) are countable. By definition, $\omega_1 > \alpha$ for every countable ordinal α, that is, for every rank defined in (4a). Now it can be said that the elements without an infinite path are those with a countable rank. There exists a natural approach of removing these elements by an iterative removal of the leaves. While it is easily seen that this works for elements with a finite rank, it is not so obvious, what happens for elements ranked with an infinite ordinal. Nevertheless, it turns out that this approach works in the general case of countable ordinals.

Consider an equation

$$X = C \cap E(Remove_{\mathtt{1}}(X)),$$

Denote its right-hand side by $\varphi(X) = C \cap E(Remove_{\mathtt{1}}(X))$, and consider the sequence $T^{(\alpha)}$ corresponding to this equation, see (3). Note, that $T^{(0)} = \mathbb{Z}$, $T^{(1)} = C$ and $T^{(\alpha)} \subseteq C$ for every ordinal α. Every step of this sequence contains the fathers of all elements occurring at the previous step:

Lemma 7. *For every countable ordinal* α, $x \in T^{(\alpha+1)}$ *if and only if* $x = (\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{10}s)_7$ *and there is* x_{k+1} *with* $(\mathtt{1}x_{k+1}\mathtt{1}x_k\mathtt{1}\ldots\mathtt{1}x_1\mathtt{10}s)_7 \in T^{(\alpha)}$.

Intuitively, the rank of an element specifies how many times this transformation can be applied until the element disappears. This is formalised as follows:

Lemma 8. *For every countable ordinal* α, $(\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{10}s)_7 \in T^{(\alpha)}$ *if and only if* $r((\mathtt{1}x_k\mathtt{1}x_{k-1}\mathtt{1}\ldots\mathtt{1}x_1\mathtt{10}s)_7) \geqslant \alpha$.

The proof is by an iterative application of Lemma 7 in a transfinite induction on α.

After ω_1 iterations, all elements with a countable rank are eliminated, and the greatest fixed point $T^{(\omega_1)}$ consist exactly of the elements with an infinite path, as they are invariant under φ.

Lemma 9. $(\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{10}s)_7 \in T^{(\omega_1)}$ *if and only if* $r((\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{10}s)_7) = \omega_1$.

Taking Lemma 6 into account, $(\mathtt{1}x_k \ldots \mathtt{1}x_1\mathtt{10}s)_7 \in T^{(\omega_1)}$ if and only if there exists an infinite sequence $x_{k+1}, \ldots x_{k+\ell}, \ldots$, such that for each $\ell \geqslant 0$, $(\mathtt{1}x_{k+\ell}\mathtt{1}x_{k+\ell-1} \ldots \mathtt{1}x_1\mathtt{10}s)_7 \in C$. It remains to extract the set S out of $T^{(\omega_1)}$. This is done using the expression $Remove_{\mathtt{10}}(F) = \{(w)_7 \mid (\mathtt{10}w)_7 \in F\}$ defined in Lemma 3. Consider a new variable Y with an new equation, which forms the following system:

$$\begin{cases} X = C \cap E(Remove_{\mathtt{1}}(X)) \\ Y = Remove_{\mathtt{10}}(X) \end{cases} \qquad (5)$$

Main Lemma. *The system* (5) *has a greatest solution with* $Y = S$.

The system constructed in this section uses a recursively enumerable constant set $C \subseteq \mathbb{N}$, as well as several constants required by Lemmata 1, 2 and 3. The former constant is representable by Theorem 1, while the rest of the constants are expressed as in the proof of that theorem. The method in the proof of Theorem 1 is also used to represent a set of integers from its positive and negative part. This yields the following result:

Theorem 2. *Every* Σ^1_1*-set* $S \subseteq \mathbb{Z}$ *is a unique solution of a resolved system of equations over sets of integers using union, intersection and addition, as well as singleton constants and the constants* $\mathbb{N}$, $-\mathbb{N}$.

The construction in the this section essentially used the infinite constants $\mathbb{N}$ and $-\mathbb{N}$. It turns out that at least one infinite constant is needed, as otherwise only trivial greatest solutions can be obtained.

Lemma 10. *For every solution of a resolved system of equations over* $\mathbb{Z}$ *using union, intersection, addition and finite constants, there is a greater solution with each component either finite or equal to* $\mathbb{Z}$.

5 Least solutions of resolved systems

As mentioned in Section 3, whenever a monotone operator is also continuous, reaching its least fixed point does not require a transfinite number of iterations: $S^{(\omega)}$ is always the least solution. In fact, this holds for a weaker property than continuity.

An operator φ is said to be $\cup$-*continuous* if $\varphi\big(\bigsqcup_{i\in\mathbb{N}} B_i\big) = \bigsqcup_{i\in\mathbb{N}} \varphi(B_i)$ holds for every *increasing* sequence B_i. A composition of $\cup$-continuous operators is $\cup$-continuous as well. It turns out that while addition of sets of integers is not continuous, it possesses this weaker property.

	least	unique	greatest
unresolved over $2^{\mathbb{N}}$, with $\{+, \cup\}$	Σ^0_1 (r.e.) [4]	Δ^0_1 (rec.) [4]	Π^0_1 (co-r.e.) [4]
resolved over $2^{\mathbb{Z}}$, with $\{+, \cup, \cap\}$	Σ^0_1	Σ^0_1	Σ^1_1
unresolved over $2^{\mathbb{Z}}$, with $\{+, \cup\}$	?	Δ^1_1 (HA) [6]	Σ^1_1

Table 1. Expressive power of solutions.

Lemma 11. *A function over sets of integers defined as a composition of union, intersection, addition and any constants is* $\cup$*-continuous.*

Then it is known that the least fixpoint of any such function is reached in ω iterations. This leads to the following theorem:

Theorem 3. *The least solution of every resolved system of equations* $X_i = \varphi_i(X_1, \ldots, X_n)$ *over sets of integers using union, intersection, addition and r.e. constants is an r.e. set.*

For singleton constants, an algorithm constructs $S^{(\alpha)}$ for all $\alpha < \omega$, until the input number is found. The case of r.e. constants is reduced to the former case by encoding the constants as in Theorem 1.

Conversely, by Theorem 1, every r.e. set is represented by such a unique solution of a system with singleton constants and constants $\mathbb{N}$ and $-\mathbb{N}$, and hence by a least solution of such a system. Furthermore, the sets $\mathbb{N}$ and $-\mathbb{N}$ can be expressed as least solutions of the following equations:

$$X = (X + 1) \cup \{0\} \qquad X' = (X' + \{-1\}) \cup \{0\}.$$

Altogether, the following characterization is obtained:

Corollary 1. *Least solutions of resolved systems of equations* $X_i = \varphi_i(X_1, \ldots, X_n)$ *over sets of integers using union, intersection, addition and constants* $\{1\}$ *and* $\{-1\}$ *represent exactly the r.e. sets. If all r.e. constants are allowed, only r.e. sets can be represented.*

6 Conclusion

The new results on the expressive power of least and greatest solutions of equations over sets of integers are summarised and compared to related results in Table 1. The same results extend to a slightly different model: equations over sets of natural numbers with union, intersection, addition and subtraction: $A \dot{-} B = \{a - b \mid a \in A,\, b \in B,\, a \geqslant b\}$ their least solutions represent exactly the r.e. sets, while their greatest solutions represent all sets in Σ^1_1. These equations are isomorphic to language equations over a unary alphabet, with the operations of union, intersection, concatenation and quotient. Furthermore, the same results could be extended to language equations over multiple-letter alphabets, by a technically much simpler construction than presented in this paper.

Of the decision problems for these equations, solution existence is trivial (as there is always a least and a greatest solution), while the complexity of testing whether a system has a unique solution is left as an open problem.

References

1. S. Ginsburg, H. G. Rice, "Two families of languages related to ALGOL", *Journal of the ACM*, 9 (1962), 350–371.
2. A. Jeż, "Conjunctive grammars can generate non-regular unary languages", *International Journal of Foundations of Computer Science*, 19:3 (2008), 597–615.
3. A. Jeż, A. Okhotin, "Conjunctive grammars over a unary alphabet: undecidability and unbounded growth", *Theory of Computing Systems*, to appear.
4. A. Jeż, A. Okhotin, "On the computational completeness of equations over sets of natural numbers" *ICALP 2008* (Reykjavik, Iceland) LNCS 5126, 63–74.
5. A. Jeż, A. Okhotin, "Equations over sets of natural numbers with addition only", *STACS 2009* (Freiburg, Germany, 26–28 February, 2009), 577–588.
6. A. Jeż, A. Okhotin, "On equations over sets of integers", *STACS 2010* (Nancy, France, 4–6 March, 2010), 477–488.
7. S. C. Kleene, *Introduction to metamathematics*, North-Holland, Amsterdam, 1952.
8. M. Kunc, "What do we know about language equations?", *Developments in Language Theory* (DLT 2007, Turku, Finland, July 3–6, 2007), LNCS 4588, 23–27.
9. T. Lehtinen, A. Okhotin, "On language equations $XXK = XXL$ and $XM = N$ over a unary alphabet", *Developments in Language Theory* (DLT 2010, London, Ontario, Canada, August 17–20, 2010), LNCS 6224, to appear.
10. P. McKenzie, K. Wagner, "The complexity of membership problems for circuits over sets of natural numbers", *Computational Complexity*, 16:3 (2007), 211–244.
11. A. Okhotin, "Conjunctive grammars", *Journal of Automata, Languages and Combinatorics*, 6:4 (2001), 519–535.
12. A. Okhotin, "Conjunctive grammars and systems of language equations", *Programming and Computer Software*, 28:5 (2002), 243–249.
13. A. Okhotin, "On the equivalence of linear conjunctive grammars to trellis automata", *Informatique Théorique et Applications*, 38:1 (2004), 69–88.
14. A. Okhotin, "Strict language inequalities and their decision problems", *MFCS 2005* (Gdańsk, Poland, August 29–September 2, 2005), LNCS 3618, 708–719.
15. A. Okhotin, "Decision problems for language equations", *Journal of Computer and System Sciences*, 76:3–4 (2010), 251–266.
16. H. Rogers, Jr., *Theory of Recursive Functions and Effective Computability*, McGraw-Hill, 1967.
17. A. Tarski, "A lattice theoretical fixpoint theorem and its applications", *Pacific Journal of Mathematics*, 5 (1955), 285–310.
18. S. D. Travers, "The complexity of membership problems for circuits over sets of integers" *Theoretical Computer Science*, 369:1–3 (2006), 211–229.