On the computational completeness of equations over sets of natural numbers[*]

Artur Jeż[†] Alexander Okhotin[‡]

March 16, 2010

Abstract

Systems of equations of the form $\varphi_j(X_1, \ldots, X_n) = \psi_j(X_1, \ldots, X_n)$ with $1 \leqslant j \leqslant m$ are considered, in which the unknowns X_i are sets of natural numbers, while the expressions φ_j, ψ_j may contain singleton constants and the operations of union and pairwise addition $S + T = \{\, m + n \mid m \in S,\ n \in T\}$. It is shown that the family of sets representable by unique (least, greatest) solutions of such systems is exactly the family of recursive (r.e., co-r.e., respectively) sets of numbers. Basic decision problems for these systems are located in the arithmetical hierarchy. The same results are established for equations with addition and intersection.

1 Introduction

Consider equations, in which the variables assume values of sets of natural numbers, and the left- and right-hand sides use Boolean operations and pairwise addition of sets defined as $S + T = \{\, m + n \mid m \in S,\ n \in T\}$. The simplest example of such an equation is $X = (X + X) \cup \{2\}$, with the set of all even numbers as the least solution. On one hand, such equations constitute a basic mathematical object, which is closely related to *integer expressions* introduced in the seminal paper by Stockmeyer and Meyer [19] and later systematically studied by McKenzie and Wagner [11]. On the other hand, they can be regarded as *language equations* over a one-letter alphabet, with the sum of sets representing concatenation of such languages.

Language equations are equations with formal languages as unknowns, which recently became an active area of research, with numerous connections to computability established. Undecidability of the solution existence problem for language equations with concatenation and Boolean operations was shown by Charatonik [1]. Later it was determined by Okhotin [13, 15, 16] that the family of sets representable by unique (least,

---

[*] The extended abstract of this paper was presented at the ICALP 2008 conference held in Reykjavik, Iceland on July 6–13, 2008. Research supported by the Polish Ministry of Science and Higher Education under grant MNiSW N206 024 31/3826, 2006–2008, and by the Academy of Finland under grant 118540.

[†] Institute of Computer Science, University of Wrocław, 50–383 Wrocław, Poland. E-mail: `aje@ii.uni.wroc.pl`

[‡] Academy of Finland, *and* Department of Mathematics, University of Turku, *and* Turku Centre for Computer Science Turku FIN–20014, Finland. E-mail: `alexander.okhotin@utu.fi`

greatest) solutions of such equations is exactly the family of recursive languages (recursively enumerable, co-recursively enumerable, respectively). Shortly thereafter, Kunc [8] constructed a language equation of the form $XL = LX$, where $L \subseteq \{a, b\}^*$ is a finite constant language, with a computationally universal greatest solution. A survey of the area was recently given by Kunc [9].

The cited results essentially use languages over alphabets containing at least two symbols, and, until recently, the seemingly trivial case of language equations over a unary alphabet $\Sigma = \{a\}$ received little attention. Systems of the form

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

with union and concatenation represent context-free grammars, and their solutions over a unary alphabet are well-known to be regular. Constructing any equation with a non-regular unique solution is already not a trivial task. The first example of such an equation using the operations of concatenation and complementation was presented by Leiss [10], who however provided no insight into how his example was obtained. Not long ago Jeż [5] constructed a system (*) using concatenation, union and intersection with a non-regular solution, and this example had a clear explanation in terms of base-4 notation of numbers: concatenating and intersecting several unknown sets ultimately allowed representing the set of powers of 4. Using this method, a large class of unary languages was proved to be representable by Jeż and Okhotin [6], who showed that equations (*) with concatenation, union and intersection can simulate cellular automata of a certain simple kind [2] recognizing positional notation of numbers. The membership of a number in a solution can be straightforwardly tested in exponential time, and it was subsequently shown that a particular EXPTIME-complete set of numbers can be represented by a system of this form [7].

These recent advances suggest the task of understanding the exact limits of the expressive power of equations over sets of numbers (or language equations over a unary alphabet) of the general form

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

Unexpectedly, this paper establishes *computational completeness* of such systems, in which φ_j, ψ_j use only addition and union, as well as singleton constants. The same result is obtained for systems with addition and intersection. To be precise, it is proved that a set is representable as a component of a unique solution of a system (**) if and only if this set is recursive. Similar characterizations are obtained for least and greatest solutions (by r.e. and co-r.e. sets, respectively). This characterizes the notion of a computable set by extremely simple arithmetical equations.

The proof method can be described as a new kind of arithmetization of Turing machines, which uses addition of natural numbers as the only arithmetical operation. It is worth mentioning that constructing any system of this restricted form with a non-periodic unique, least or greatest solution is already a nontrivial task. No examples of their nontriviality have been shown up to date, and it would seem expected that their solutions are always periodic. The first examples of such systems representing non-regular sets constructed in this paper require quite extensive constructions.

The general line of the computational completeness argument presented in this paper models upon the existing computational completeness results for language equations [13], which are summarized in Section 2. The previous results on equations over sets of numbers [5, 6], explained in Section 3, are used as main building blocks. In the next Section 4, it is shown how the computational completeness arguments for language equations can be *remade* for a much more restricted object: equations over sets of numbers with *union, intersection and addition.* These results are improved in Section 5 to use systems (**) with either union or intersection, which requires re-implementing all core constructions using these more restricted equations. Finally, decision problems for these equations are considered in Section 6, and it is proved that, like in the case of language equations [13, 16], testing existence of a solution of a given system (**) is undecidable (complete for Π_1 in the arithmetical hierarchy), testing solution uniqueness is Π_2-complete, while testing whether a system has finitely many solutions is Σ_3-complete.

2 Language equations and their computational completeness

Let Σ be a finite alphabet and consider systems of equations of the form

$$\varphi_j(X_1, \ldots, X_n) = \psi_j(X_1, \ldots, X_n),$$

where the unknowns X_i are languages over Σ, while φ_j and ψ_j are expressions using union, intersection and concatenation, as well as singleton constants. The following computational completeness result is known:

Theorem 1 (Okhotin [13, 15]). *Let (**) be a system that has a unique (least, greatest) solution* $(L_1, \ldots, L_n)$. *Then each component* L_i *is recursive (r.e., co-r.e., respectively). Conversely, for every recursive (r.e., co-r.e.) language* $L \subseteq \Sigma^*$ *(with* $|\Sigma| \geqslant 2$*) there exists a system (**) with the unique (least, greatest, respectively) solution* $(L, \ldots)$.

As this paper considers a much more restricted family of equations, the first part of Theorem 1 will apply as it is, while the lower bound proofs will have to be entirely remade. The proof of the second part of Theorem 1 will serve as a model for the arguments for the case of equations over sets of numbers. The following sketch of this proof is useful for understanding the constructions presented later in this paper.

The main underlying device used for constructing such a system is the language of computation histories of a Turing machine, defined and used by Hartmanis [4] to give systematical undecidability proofs for context-free grammars. In short, for every TM T over an input alphabet Σ one can construct an alphabet Γ and an encoding of computations $C_T : \Sigma^* \to \Gamma^*$, so that for every $w \in L(T)$ the string $C_T(w)$ lists the configurations of T at each step of its accepting computation on w, and the language

$$\mathrm{VALC}(T) = \{\, w\natural C_T(w) \mid C_T(w) \text{ is an accepting computation}\},$$

where $\natural \notin \Sigma \cup \Gamma$, is an intersection of two linear context-free languages. Since equations (**) can directly simulate context-free grammars and are equipped with intersection, for every Turing machine it is easy to construct a system in variables $(X_1, \ldots, X_n)$ with a unique solution $(L_1, \ldots, L_n)$, so that $L_1 = \text{VALC}(T)$.

It remains to "extract" $L(T)$ out of $\text{VALC}(T)$ using a language equation. Let Y be a new variable and consider the inequality

$$\text{VALC}(T) \subseteq Y\natural\Gamma^*,$$

which can be formally rewritten as an equation $X_1 \cup Y\natural\Gamma^* = Y\natural\Gamma^*$. This inequality states that for every $w \in L(T)$, the string $w\natural C_T(w)$ should be in $Y\natural\Gamma^*$, that is, w should be in Y. This makes $L(T)$ the least solution of this inequality and proves the second part of Theorem 1 with respect to r.e. sets and least solutions. The construction for a co-r.e. set and a greatest solution is established by a dual argument, and these two constructions can then be combined to represent every recursive set [15]. Furthermore, the same construction is used to establish the undecidability of the solution existence and solution uniqueness problems for a given language equation [13, 16].

At the first glance, the idea that the above results could possibly hold if the alphabet consists of a single letter sounds odd. However, this is what will be proved in this paper, and, moreover, the general plan of the argument remains essentially the same.

3 Resolved systems with $\{\cup, \cap, +\}$

A formal language L over the alphabet $\Sigma = \{a\}$ can be regarded as a set of numbers $\{\, n \mid a^n \in L\}$, and so equations over sets of numbers represent a very special subclass of language equations. Very little was known about this case, until the recent results on *resolved systems* over sets of natural numbers, which are of the form

$$X_i = \varphi(X_1, \ldots, X_n) \quad (1 \leqslant i \leqslant n),$$

where the right-hand sides φ_i may contain union, intersection and addition, as well as singleton constants. To minimize the number of brackets, assume that the addition has the highest precedence, followed by intersection, while the precedence of union is the least.

If intersection is disallowed, such systems are basically context-free grammars over a one-letter alphabet, and hence their solutions are ultimately periodic. Equations with both union and intersection are equivalent to an extension of context-free grammars, the *conjunctive grammars* [12], and the question whether any non-periodic set can be specified by such a system of equations has been open for some years, until answered by the following example:

Example 1 (Jeż [5])**.** *The least solution of the system*

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

is $\big(\{\, 4^n \mid n \geqslant 0\}, \{\, 2 \cdot 4^n \mid n \geqslant 0\}, \{\, 3 \cdot 4^n \mid n \geqslant 0\}, \{\, 6 \cdot 4^n \mid n \geqslant 0\}\big)$.

To understand this construction, it is useful to consider positional notation of numbers. Let $\Sigma_k = \{\mathtt{0}, \mathtt{1}, \ldots, k-1\}$ be digits in base-k notation. For every $w \in \Sigma_k^*$, let $(w)_k$

be the number defined by this string of digits. 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 $\big((10^*)_4, (20^*)_4, (30^*)_4, (120^*)_4\big)$.

The following generalization of this example has been obtained:

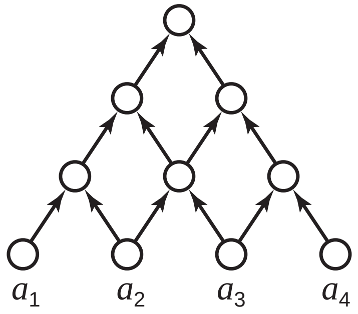
Theorem 2 (Jeż [5]). *For every $k \geqslant 2$ and for every regular language $L \subseteq \Sigma_k^+$ there exists a resolved system over sets of natural numbers in variables X, $Y_2, \ldots, Y_n$ with the least solution $X = (L)_k$ and $Y_i = K_i$ for some $K_i \subseteq \mathbb{N}$.*

A further extension of this result allows one to take a *trellis automaton* (one-way real-time cellular automaton) recognizing a positional notation of a set of numbers, and construct a system of equations representing this set of numbers.

A trellis automaton [2, 14], defined as a quintuple $(\Sigma, Q, I, \delta, F)$, processes an input string of length $n \geqslant 1$ using a uniform array of $\frac{n(n+1)}{2}$ nodes, as presented in the figure below. Each node computes a value from a fixed finite set Q. The nodes in the bottom row obtain their values directly from the input symbols using a function $I : \Sigma \to Q$. The rest of the nodes compute the function $\delta : Q \times Q \to Q$ of the values in their predecessors. The string is accepted if and only if the value computed by the topmost node belongs to the set of accepting states $F \subseteq Q$.

Definition 1. *A trellis automaton is a quintuple $M = (\Sigma, Q, I, \delta, F)$, in which:*

- *Σ is the input alphabet,*
- *Q is a finite non-empty set of states,*
- *$I : \Sigma \to Q$ is a function that sets the initial states,*
- *$\delta : Q \times Q \to Q$ is the transition function, and*
- *$F \subseteq Q$ is the set of final states.*



Extend δ to a function $\delta : Q^+ \to Q$ by $\delta(q) = q$ and

$$\delta(q_1, \ldots, q_n) = \delta(\delta(q_1, \ldots, q_{n-1}), \delta(q_2, \ldots, q_n)),$$

while I is extended to a homomorphism $I : \Sigma^* \to Q^*$.
Let $L_M(q) = \{ w \mid \delta(I(w)) = q\}$ and define $L(M) = \bigcup_{q \in F} L_M(q)$.

Theorem 3 (Jeż, Okhotin [6]). *For every $k \geqslant 2$ and for every trellis automaton M over Σ_k with $L(M) \cap 0\Sigma_k^* = \varnothing$ there exists a resolved system over sets of natural numbers in variables X, $Y_2, \ldots, Y_n$ with the least solution $X = (L(M))_k$ and $Y_i = K_i$ for some $K_i \subseteq \mathbb{N}$.*

An important example of a set representable according to this theorem is the numerical version of the set of computational histories of a given Turing machine. The symbols needed to represent the standard language of computations of a Turing machine are interpreted as digits, and then every string from this language is represented by a number. Since the standard language of computations can be recognized by a trellis automaton, by Theorem 3 there is a system of equations representing the corresponding set of numbers. This set can be used straightforwardly to infer some undecidability results on conjunctive grammars [6].

In the next section, such a set of numbers will be used for the same purpose as the standard language VALC in the computational completeness proofs for language equations [13, 15, 16].

4 Unresolved systems with $\{\cup, \cap, +\}$

Consider systems of equations of the form

$$\varphi_j(X_1, \ldots, X_n) = \psi_j(X_1, \ldots, X_n) \quad (1 \leqslant j \leqslant m),$$

where the unknowns X_i are sets of natural numbers and φ_j, ψ_j may use union, intersection and addition, as well as singleton constants.

The ultimate result of this paper is the computational completeness of such systems using *either* union *or* intersection, which is stated as follows:

Theorem 4. *The family of sets of natural numbers representable by unique (least, greatest) solutions of systems of equations of the form $\varphi_i(X_1, \ldots, X_n) = \psi_i(X_1, \ldots, X_n)$ with union and addition, is exactly the family of recursive (r.e., co-r.e., respectively) sets. The same result holds for systems with intersection and addition.*

These solutions are recursive (r.e., co-r.e., respectively) because so are the solutions of language equations with union, intersection and concatenation, as asserted by Theorem 1. So the task is to take any recursive (r.e., co-r.e.) set of numbers and to construct two systems of equations representing this set by a solution of the corresponding kind: one system with union and addition, and the other using intersection and addition. In this section, Theorem 4 is established in its weaker version, with the constructed systems using both union and intersection.

The case of only one Boolean operation presents additional challenges, mainly because the systems constructed in Theorems 2 and 3 already require both union and intersection, and thus have to be reproved for the cases of only union and only intesection. Theorem 4 will be established in its full later in Section 5.

The construction for Theorem 4 is based upon a rather complicated arithmetization of Turing machines, which proceeds in several stages. First, valid accepting computations of a Turing machine are represented as numbers, so that these numbers could be recognized by a trellis automaton working on base-6 positional notation of these numbers, which are regarded as strings over the alphabet $\Sigma_6 = \{\mathtt{0}, \mathtt{1}, \mathtt{2}, \mathtt{3}, \mathtt{4}, \mathtt{5}\}$. While trellis automata are rather flexible and could accept many different encodings of such computations, the subsequent constructions require a set of numbers of a very specific form. This form will now be defined.

Consider the following standard encoding of computations as strings:

Definition 2. *Let T be a Turing machine recognizing numbers given to it in base-6 notation. Let $V \supset \Sigma_6$ be its tape alphabet, let Q be its set of states, and define $\Gamma = V \cup Q \cup \{\sharp\}$. Let $S(T) \subseteq \mathbb{N}$ be the set of numbers accepted by T.*

For every number $n \in S(T)$, denote the instantaneous description of T after i steps of computation on n as a string $ID_i = \alpha q a \beta \subseteq V^* Q V V^*$, where T is in state q scanning $a \in \Gamma$ and the tape contains $\alpha a \beta$. Define

$$\widetilde{C}_T(n) = ID_0 \cdot \sharp \cdot ID_1 \cdot \sharp \cdot \ldots \cdot \sharp \cdot ID_{\ell-1} \cdot \sharp\sharp \cdot ID_\ell \cdot \sharp\sharp \cdot (ID_\ell)^R \cdot \sharp \cdot \ldots \cdot \sharp \cdot (ID_1)^R \cdot \sharp \cdot (ID_0)^R$$

Next, consider any code $h : \Gamma^* \to \Sigma_6^*$*, under which every codeword is in* $\{\mathtt{30}, \mathtt{300}\}^+$*. Define* $C_T(n) = h(\widetilde{C}_T(n))\mathtt{300}$*.*

The language $\{\, \widetilde{C}_T(n) \mid n \in S(T) \} \subseteq \Gamma^*$ is an intersection of two linear context-free languages and hence is recognized by a trellis automaton [2, 14]. By the known closure of trellis automata under codes [17], the language $\{\, C_T(n) \mid n \in S(T) \} \subseteq \Sigma_6^+$ is recognized by a trellis automaton as well.

Now the set of accepting computations of a Turing machine is represented as the following six sets of numbers:

Definition 3. *Let* T *be a Turing machine recognizing numbers given in base-6 notation. For every* $i \in \{\mathtt{1}, \mathtt{2}, \mathtt{3}, \mathtt{4}, \mathtt{5}\}$*, the valid accepting computations of* T *on numbers* $n \geqslant 6$ *with their base-6 notation beginning with the digit* i *is*

$$\mathrm{VALC}_i(T) = \{\, (C_T(n)\mathtt{1}w)_6 \mid n = (iw)_6,\ n \in S(T) \},$$

The computations of T *on numbers* $n \in \{0, 1, 2, 3, 4, 5\}$*, provided that they are accepting, are represented by the following finite set of numbers:*

$$\mathrm{VALC}_0(T) = \{\, (C_T(n))_6 + n \mid n \in \{0, 1, 2, 3, 4, 5\} \text{ and } n \in S(T) \}$$

For example, under this encoding, the accepting computation on a number $n = (\mathtt{543210})_6$ will be represented by a number $(\mathtt{30300300} \ldots \mathtt{30300143210})_6 \in \mathrm{VALC}_5(T)$, where the whole computation is encoded by blocks of digits `30` and `300`, the digit `1` acts as a separator and the lowest digits `43210` represent n with its leading digit cut.

A crucial property of this encoding is that it allows simulating *concatenation of strings of digits* representing the computation and the input number, which is simulated by adding these numbers to each other. Clearly, the number representing the computation of T on $(iw)_6 \in L(T)$ is representable as a sum of $(\mathtt{1}w)_6$ and an appropriate number in $(\{\mathtt{30}, \mathtt{300}\}^*\mathtt{3000}^*)_6$. What is important is that the converse statement holds as well: that is, whenever the sum of a number $(\mathtt{1}w)_6$ and any number in $(\{\mathtt{30}, \mathtt{300}\}^*\mathtt{3000}^*)_6$ is of the form $(x\mathtt{1}u)_6$ with $x \in \{\mathtt{30}, \mathtt{300}\}^*\mathtt{300}$, the string u must be equal to w. The following lemma rules out the hypothetical possibility that the number $(x\mathtt{1}u)_6$ could be obtained in any other way.

Lemma 1. *Let* $S \subseteq (\mathtt{1}\Sigma_6^+)_6$*. Then for all strings* $x \in \{\mathtt{30}, \mathtt{300}\}^*\mathtt{300}$ *and* $u \in \Sigma_6^+$*, if* $(x\mathtt{1}u)_6 \in (\{\mathtt{30}, \mathtt{300}\}^*\mathtt{3000}^*)_6 + S$*, then* $(\mathtt{1}u)_6 \in S$*.*

Proof. Let $(x\mathtt{1}u)_6 = (y\mathtt{0}^\ell)_6 + (\mathtt{1}v)_6$, where $y \in \{\mathtt{30}, \mathtt{300}\}^*\mathtt{300}$, $\ell \geqslant 0$ and $(\mathtt{1}v)_6 \in S$. The goal is to show that $u = v$, $x = y$ and $|\mathtt{1}v| = \ell$, that is, the only way by obtaining $(x\mathtt{1}u)_6$ is by adding $(x\mathtt{0}^{|u|+1})_6$ to $(\mathtt{1}u)_6$, and no other numbers from $(\{\mathtt{30}, \mathtt{300}\}^*\mathtt{3000}^*)_6$ and S could be used to fake this number. The proof is by the analysis of all "improper" cases of addition.

In each case it will be shown that the base-6 notation of $(y\mathtt{0}^\ell)_6 + (\mathtt{1}v)_6$ has a different structure than $x\mathtt{1}u$. Depending on the number of digits in $|\mathtt{1}v|$, consider the following cases:

1. $|\mathtt{1}v| < \ell$. Then $(y\mathtt{0}^\ell)_6 + (\mathtt{1}v)_6 = (y\mathtt{0}^{\ell-|\mathtt{1}v|}\mathtt{1}v)_6$, which is a number with a base-6 notation containing at least three consecutive zeroes to the left of the leftmost digit `1`. Since $(x\mathtt{1}u)_6$ has two zeroes to the left of the leftmost `1`, it follows that $(y\mathtt{0}^\ell)_6 + (\mathtt{1}v)_6 \neq (\mathtt{1}u)_6$, which makes this case impossible.

2. $|\mathtt{1}v| = \ell$. This is the case of addition done as intended. Then $(y\mathtt{0}^\ell)_6 + (\mathtt{1}v)_6 = (y\mathtt{1}v)_6$, and thus $(y\mathtt{1}v)_6 = (x\mathtt{1}u)_6$. The leftmost instance of $\mathtt{1}$ in $(y\mathtt{1}v)_6$ and in $(x\mathtt{1}u)_6$ is at the first position of $\mathtt{1}v$ and $\mathtt{1}u$, respectively. Therefore, $y = x$ and $\mathtt{1}v = \mathtt{1}u$.

3. $\ell < |\mathtt{1}v| \leqslant |y| + \ell$. Then the leading $\mathtt{1}$ from $\mathtt{1}v$ is at the same position as some digit of y in $y\mathtt{10}^\ell$. Let $y = y_1 i y_2$, where $|y_2| + \ell = |v|$. The digit i is either $\mathtt{0}$ or $\mathtt{3}$.

   - If $i = \mathtt{0}$, then y_1 ends with $\mathtt{3}$ or $\mathtt{30}$. The sum $(y_1 i y_2 \mathtt{0}^\ell)_6 + (\mathtt{1}v)_6$ is thus of the form $(y_1 i' z)_6$, where $i' \in \{\mathtt{1}, \mathtt{2}\}$ ($\mathtt{2}$ can appear due to a possible carry from the earlier position), and the prefix $y_1 i'$ is in $\{\mathtt{30}, \mathtt{300}\}^*\{\mathtt{31}, \mathtt{32}, \mathtt{301}, \mathtt{302}\}$. On the other hand, in $(x\mathtt{1}u)_6$, the leftmost occurrence of digits outside of $\{\mathtt{3}, \mathtt{0}\}$ must be of the form $\mathtt{3001}$.
   - If $i = \mathtt{3}$, then the sum $(y_1 i y_2 \mathtt{0}^\ell)_6 + (\mathtt{1}v)_6$ is of the form $(y_1 i' z)_6$, where $|z| = |v|$ and $i' \in \{\mathtt{4}, \mathtt{5}\}$ ($\mathtt{5}$ can appear due to a possible carry from the earlier position). Consider the leftmost digits of the numbers $(y_1 i' z)_6$ and $(x\mathtt{1}u)_6$ different from $\mathtt{0}$ and $\mathtt{3}$. For $(x\mathtt{1}u)_6$ it is $\mathtt{1}$, while for $(y_1 i' z)_6$ it is $\mathtt{4}$ or $\mathtt{5}$, and thus these numbers cannot be the same.

   In both cases it follows that $(y_1 i y_2 \mathtt{0}^\ell)_6 + (\mathtt{1}v)_6$ and $(x\mathtt{1}u)_6$ must be different, and the case is impossible.

4. $|\mathtt{1}v| > |y| + \ell$. Then the leading digit of $(y\mathtt{0}^\ell)_6 + (\mathtt{1}v)_6$ is $\mathtt{1}$ or $\mathtt{2}$ (due to a possible carry). As the leading digits are different, $(y\mathtt{0}^\ell)_6 + (\mathtt{1}v)_6 \neq (x\mathtt{1}u)_6$, which rules out this case.

It has thus been established that $y = x$ and $\mathtt{1}v = \mathtt{1}u$ in the only possible case, which yields the claim. □

A trellis automaton recognizing the base-6 notation of numbers in $\mathrm{VALC}_i(T)$, by Theorem 3, can be used to obtain a system of equations with union, intersection and addition representing $\mathrm{VALC}_i(T)$. The system given by Theorem 3 is actually resolved; casting away that property, this result can be proved in the following stronger form using only one Boolean operation:

Lemma 2. *For every TM T recognizing numbers there exists a system of equations*

$$\varphi_j(Y, X_1, \ldots, X_m) = \psi_j(Y, X_1, \ldots, X_m)$$

over sets of natural numbers using union and addition (intersection and addition), such that its least solution is $(S_0, S_1, \ldots, S_5, S_6, \ldots, S_n)$ with $S_i = \mathrm{VALC}_i(T)$ for $0 \leqslant i \leqslant 5$.

The lemma will be established in this form only in Section 5, but at the moment its weaker form is known, which asserts representability of the sets $\mathrm{VALC}_i(T)$ by equations with union, intersection and addition.

The next task is to use these sets of numbers as constants in order to construct equations representing $L(T)$. The first case to be established is the case of least solutions and r.e. sets.

Lemma 3. *For every TM T accepting a set $S_0 \subseteq \mathbb{N}$ there exists a system of equations of the form*

$$\varphi_j(Y, X_1, \ldots, X_m) = \psi_j(Y, X_1, \ldots, X_m)$$

with union and addition (or equally with intersection and addition), which has the set of solutions

$$\{\, (S, f_1(S), \ldots, f_m(S)) \mid S_0 \subseteq S \,\}$$

where $f_1, \ldots, f_m : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ are some monotone functions on sets of numbers defined with respect to S_0. In particular, there is a least solution with $Y = S_0$.

The below argument proves a weaker form of Lemma 3, with the constructed system using both union and intersection. Most of the given equations are stated as inclusions of the form $V \subseteq V'$ or $V \subseteq V' + V''$, and thus can be equally expressed using union and using intersection. There is only one equation in the proof that explicitly uses both union and intersection, and it will be shown in Section 5 that this equation can be rephrased using either union (Lemma 16) or intersection (Lemma 15). With that correction, the below proof will establish Lemma 3 in its full form stated above.

Proof of Lemma 3 (weaker form). The proof is by constructing a system in variables $(Y, Y_1, \ldots, Y_5, Y_0, X_7, \ldots, X_m)$, where the number m will be determined below, and the set of solutions of this system is defined by the following conditions, which ensure that the statement of the lemma is fulfilled:

$$S(T) \cap \{0,1,2,3,4,5\} \subseteq Y_0 \subseteq \{0,1,2,3,4,5\}, \tag{1a}$$

$$\{\, (\mathtt{1}w)_6 \mid w \in \Sigma_6^+,\ (iw)_6 \in S(T)\} \subseteq Y_i \subseteq (\mathtt{1}\Sigma_6^+)_6 \quad (1 \leqslant i \leqslant 5), \tag{1b}$$

$$Y = Y_0 \cup \bigcup_{i=1}^{5} \{\, (iw)_6 \mid (\mathtt{1}w)_6 \in Y_i\}, \tag{1c}$$

$$X_j = K_j \quad (7 \leqslant j \leqslant m). \tag{1d}$$

The sets $K_7, \ldots, K_m$ are some constants needed for the construction to work. These constants and the equations needed to specify them will be implicitly obtained in the proof. The constructed system will use inequalities of the form $\varphi \subseteq \psi$, which can be equivalently rewritten as equations $\varphi \cup \psi = \psi$ or $\varphi \cap \psi = \varphi$.

For each $i \in \{1,2,3,4,5\}$, consider the above definition of $\mathrm{VALC}_i(T)$, which can be constructed by Lemma 2, and define a variable Y_i with the equations

$$Y_i \subseteq (\mathtt{1}\Sigma_6^+)_6, \tag{2a}$$

$$\mathrm{VALC}_i(T) \subseteq (\{\mathtt{30}, \mathtt{300}\}^*\mathtt{3000}^*)_6 + Y_i. \tag{2b}$$

Both constants are given by regular languages of base-6 representations, and therefore can be specified by equations according to Theorem 2. It is claimed that this system is equivalent to (1b).

Suppose (1b) holds for Y_i. Then (2a) immediately follows. To check (2b), consider any $(C_T^i(iw)\mathtt{1}w)_6 \in \mathrm{VALC}_i(T)$. Since this number represents the computation of T on $(iw)_6$, this implies $(iw)_6 \in S(T)$, and hence $(\mathtt{1}w)_6 \in Y_i$ by (1b). Then $(C_T^i(iw)\mathtt{1}w)_6 \in (\{\mathtt{30}, \mathtt{300}\}^*\mathtt{3000}^{|\mathtt{1}w|})_6 + (\mathtt{1}w)_6 \subseteq (\{\mathtt{30}, \mathtt{300}\}^*\mathtt{3000})_6 + Y_i$, which proves the inclusion (2b).

Conversely, assuming (2), it has to be proved that for every $(iw)_6 \in S(T)$, where $w \in \Sigma_6^+$, the number $(\mathtt{1}w)_6$ is in Y_i. Since $(iw)_6 \in S(T)$, there exists an accepting computation of T: $(C_T^i(iw)\mathtt{1}w)_6 \in \mathrm{VALC}_i(T)$. Hence, $(C_T^i(iw)\mathtt{1}w)_6 \in (\{\mathtt{30},\mathtt{300}\}^*\mathtt{3000}^*)_6 + Y_i$ due to the inclusion (2b), and therefore $(\mathtt{1}w)_6 \in Y_i$ by Lemma 1.

Define one more variable Y_0 with the equations

$$Y_0 \subseteq \{0, 1, 2, 3, 4, 5\}, \tag{3a}$$

$$\mathrm{VALC}_0(T) \subseteq (\{\mathtt{30},\mathtt{300}\}^*\mathtt{300})_6 + Y_0. \tag{3b}$$

The claim is that (3) holds if and only if (1a).

Assume (1a) and consider any number $(C_T(n))_6 + n \in \mathrm{VALC}_0(T)$, where $n \in \{0, 1, 2, 3, 4, 5\}$ by definition. Then n is accepted by T, and, by (1a), $n \in Y_0$. Since $(C_T(n))_6 \in (\{\mathtt{30},\mathtt{300}\}^*\mathtt{300})_6$, the addition of n affects only the last digit, and $(C_T(w))_6 + n \in (\{\mathtt{30},\mathtt{300}\}^*\mathtt{300})_6 + n \subseteq (\{\mathtt{30},\mathtt{300}\}^*\mathtt{300})_6 + Y_0$, which proves (3b).

The converse claim is that (3) implies that every $n \in S(T) \cap \{0, 1, 2, 3, 4, 5\}$ must be in Y_0, The corresponding $(C_T(n))_6 + n \in \mathrm{VALC}_0(T)$ is in $(\{\mathtt{30},\mathtt{300}\}^*\mathtt{300})_6 + n$ by (3b). Since n is represented by a single digit, the number $(C_T(n))_6 + n$ ends with this digit. The set $(\{\mathtt{30},\mathtt{300}\}^*\mathtt{300})_6 + Y_0$ contains a number of such a form only if $n \in Y_0$.

Next, combine the above six systems together and add a new variable Y with the following equation:

$$Y = Y_0 \cup Y_1 \cup \bigcup_{\substack{i \in \{\mathtt{2},\mathtt{3},\mathtt{4},\mathtt{5}\} \\ i' \in \Sigma_6}} \Big((Y_i \cap (\mathtt{1}i'\Sigma_6^*)_6) + ((i-1)\mathtt{0}^*)_6 \cap (ii'\Sigma_6^*)_6 \Big). \tag{4}$$

This equation has been borrowed from the authors' previous paper [6, Lem.7], where it was proved equivalent to $Y = Y_0 \cup \{\, (iw)_6 \mid (\mathtt{1}w)_6 \in Y_i\}$, that is, to (1c). Note that this is the only equation in this proof that uses explicit union or intersection; it will be shown later, in Lemmata 15–16, that this equation can be equivalently represented using only one Boolean operation.

The final step of the construction is to express constants used in the above systems through singleton constants, which can be done by Theorem 2 and Lemma 2. The variables needed to specify these languages are denoted $(X_7, \ldots, X_n)$, and the equations for these variables have a unique solution $X_j = K_j$ for all j.

This completes the description of the set of solutions of the system. It is easy to see that there is a least solution in this set, with $Y = S(T)$, $Y_0 = S(T) \cap \{0, 1, 2, 3, 4, 5\}$, $Y_i = \{\, (\mathtt{1}w)_6 \mid w \in \Sigma_6^+,\ (iw)_6 \in S(T)\}$ and $X_j = K_j$. ☐

The representation of co-recursively enumerable sets by greatest solutions is dual to the case of least solutions and is established by an analogous argument.

Denote the complements of the languages $\mathrm{VALC}_i(T)$ $(0 \leqslant i \leqslant 5)$ by $\mathrm{INVALC}_i(T)$. Base-6 notations of numbers in these sets are recognized by trellis automata due to the closure of trellis automata under complementation. Therefore, analogously to Lemma 2, the sets $\mathrm{INVALC}_i(T)$ are representable by equations.

Lemma 4. *For every TM T recognizing a recursively enumerable set of numbers $S_0 \subseteq \mathbb{N}$ there exists a system of equations of the form*

$$\varphi_j(Z, X_1, \ldots, X_m) = \psi_j(Z, X_1, \ldots, X_m)$$

with union and addition (or equally with intersection and addition), which has the set of solutions

$$\left\{\, (S, f_1(S), \ldots, f_m(S)) \;\middle|\; S \subseteq \overline{S_0} \,\right\},$$

where $f_1, \ldots, f_m : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ *are some monotone functions on sets of numbers defined with respect to* S_0*. In particular, there is a greatest solution with* $Z = \overline{S_0}$*.*

As in the previous lemma, only a weaker form of Lemma 4 can be proved at the moment, with the system using both union and intersection. The full version will follow by improving one of the equations in the below argument according to the later established Lemmata 15–16.

Proof of the weaker version. The system has a set of variables $(Z, Z_1, \ldots, Z_5, Z_0, X_7, \ldots, X_m)$, and its set of solutions will be characterized by the following conditions:

$$Z_0 \subseteq \overline{S_0} \cap \{0, 1, 2, 3, 4, 5\} \tag{5a}$$

$$Z_i \subseteq \{\, (\mathtt{1}w)_6 \mid w \in \Sigma_6^+,\ (iw)_6 \notin S_0 \} \quad (1 \leqslant i \leqslant 5), \tag{5b}$$

$$Z = Z_0 \cup \bigcup_{i=1}^{5} \{\, (iw)_6 \mid (\mathtt{1}w)_6 \in Z_i \} \tag{5c}$$

$$X_j = K_j \quad (7 \leqslant j \leqslant n) \tag{5d}$$

The number m and the vector of languages $(K_7, \ldots, K_m)$ will be determined below. This set of solutions will satisfy the statement of the lemma.

The equations defining the value of each Z_i $(1 \leqslant i \leqslant 5)$ are as follows:

$$Z_i \subseteq (\mathtt{1}\Sigma_6^+)_6 \tag{6a}$$

$$(\{\mathtt{30}, \mathtt{300}\}^*\mathtt{3000}^*)_6 + Z_i \subseteq \mathrm{INVALC}_i(T), \tag{6b}$$

It is claimed that (6) holds if and only if (5b).

If Z_i satisfies (5b), then (6a) follows immediately, and in order to prove (6b), one has to consider any number not in $\mathrm{INVALC}_i(T)$ and show that it is not in $(\{\mathtt{30}, \mathtt{300}\}^*\mathtt{3000}^*)_6 + Z_i$. By definition, a number is not in $\mathrm{INVALC}_i(T)$ if it is in $\mathrm{VALC}_i(T)$, so take any number $n = (iw)_6 \in S_0$, for which $(C_T(n)\mathtt{1}w)_6 \in \mathrm{VALC}_i(T)$ with $C_T(iw) \in \{\mathtt{30}, \mathtt{300}\}^*\mathtt{300}$. Suppose $(C_T(iw)\mathtt{1}w)_6 \in (\{\mathtt{30}, \mathtt{300}\}^*\mathtt{3000}^*)_6 + Z_i$. Then, by Lemma 1, $(\mathtt{1}w)_6 \in Z_i$, hence $(iw)_6 \notin S_0$ by (5b), which yields a contradiction.

The converse is established as follows. Assuming (6), consider any number $n \in S_0$ and let $n = (iw)_6$ for some $i \in \{1, 2, 3, 4, 5\}$ and $w \in \Sigma_6^+$. It is sufficient to prove that $(\mathtt{1}w)_6 \notin Z_i$. Suppose $(\mathtt{1}w)_6 \in Z_i$, then $(C_T(n)w)_6 \in (\{\mathtt{30}, \mathtt{300}\}^*\mathtt{3000}^*)_6 + Z_i \subseteq \mathrm{INVALC}_i(T)$ by (6b). However, $(C_T(n)w)_6$ is in $\mathrm{VALC}_i(T)$ and thus cannot be in $\mathrm{INVALC}_i(T)$. The contradiction obtained proves this case.

Define the following equations for the variable Z_0:

$$Z_0 \subseteq \{0, 1, 2, 3, 4, 5\} \tag{7a}$$

$$(\{\mathtt{30}, \mathtt{300}\}^*\mathtt{300})_6 + Z_0 \subseteq \mathrm{INVALC}_0(T) \tag{7b}$$

Again, the claim is that these equations are equivalent to (5a).

Let Z_0 be a subset of $\{0,1,2,3,4,5\} \setminus S_0$, as stated in (5a). This immediately implies (7a). Consider any number not in $\mathrm{INVALC}_0(T)$; proving that it is not in $(\{\mathtt{30}, \mathtt{300}\}^*\mathtt{300})_6 + Z_0$ will establish (7b). A number not in $\mathrm{INVALC}_0(T)$ must be in $\mathrm{VALC}_0(T)$, so let $C_T(n) + n \in \mathrm{VALC}_0(T)$ for any $n \in \{0,1,2,3,4,5\}$, and suppose $C_T(n) + n \in (\{\mathtt{30}, \mathtt{300}\}^*\mathtt{300})_6 + Z_0$. The last digit of $C_T(n) + n$ is n, and hence $n \in Z_0$. Therefore, by (5a), $n \notin S_0$, which contradicts the accepting computation $C_T(n)$.

Conversely, assume (7) and suppose there exists $n \in \{0,1,2,3,4,5\}$, which is at the same time in S_0 and in Z_0. Then there exists an accepting computation $C_T(n) + n \in \mathrm{VALC}_0(T)$, that is, $C_T(n) + n \notin \mathrm{INVALC}_0(T)$. However, $C_T(n) + n \in (\{\mathtt{30}, \mathtt{300}\}^*\mathtt{300})_6 + Z_0$, because $C_T(n) \in (\{\mathtt{30}, \mathtt{300}\}^*\mathtt{300})_6$ and $w \in Z_0$ by assumption, which contradicts (7b). The contradiction obtained proves that no such w exists, which establishes (5a).

The equation for Z is the same as in Lemma 3:

$$Z = Z_0 \cup Z_1 \cup \bigcup_{\substack{i \in \{\mathtt{2},\mathtt{3},\mathtt{4},\mathtt{5}\} \\ i' \in \Sigma_6}} \Big((Z_i \cap (\mathtt{1}i'\Sigma_6^*)_6) + ((i-1)\mathtt{0}^*)_6 \cap (ii'\Sigma_6^*)_6 \Big). \tag{8}$$

As in the previous case, it is equivalent to (5c). Again, this is the only equation using union and intersection, which will later be replaced by simpler equations given in Lemmata 15–16.

To conclude the proof, linear conjunctive constants are expressed as in Theorem 3, using extra variables $(X_7, \ldots, X_n)$. The set of solutions has been described, and, clearly, the greatest of them is $Z_0 = \overline{S_0} \cap \{0,1,2,3,4,5\}$, $Z_i = \{ (\mathtt{1}w)_6 \mid w \in \Sigma_6^+,\ (iw)_6 \notin S_0 \}$, $Z = \overline{S_0}$. ☐

Finally, the case of recursive languages and unique solutions can be established by combining the constructions of Lemmata 3 and 4 as follows:

Lemma 5. *For every TM T halting on every input and recognizing a recursive set of numbers $S \subseteq \mathbb{N}$ there exists a system of equations of the form $\varphi_i(Y, Z, X_1, \ldots, X_n) = \psi_i(Y, Z, X_1, \ldots, X_n)$ with union, intersection and addition, such that its unique solution is $Y = Z = S$, $X_i = K_i$, where $(K_1, \ldots, K_n)$ is some vector of sets.*

Proof. A TM T' recognizing $\overline{S}$ is easily constructed out of T. Then Lemma 3 is applied to T and Lemma 4 is applied to T'. Consider both systems of language equations given by these lemmata, let Y be the variable from Lemma 3, let Z be the variable from Lemma 4, and let $X_1, \ldots, X_n$ be the rest of the variables in these systems combined. The set of solutions of the system obtained is

$$\big\{ (Y, Z, f_1(Y, Z), \ldots, f_n(Y, Z)) \bigm| S \subseteq Y \text{ and } Z \subseteq \overline{\overline{S}} \big\},$$

that is

$$\big\{ (Y, Z, f_1(Y, Z), \ldots, f_n(Y, Z)) \bigm| Z \subseteq S \subseteq Y \big\}.$$

Adding one more equation

$$Y = Z$$

to the system collapses the bounds $Z \subseteq S \subseteq Y$ to $Z = S = Y$, and the resulting system has the unique solution

$$\big\{ (S, S, f_1(S, S), \ldots, f_n(S, S)) \big\},$$

which completes the proof. ☐

The weaker form of the above three lemmata yields a weaker form of Theorem 4, which asserts computational completeness of equations with union, intersection and addition.

5 Unresolved systems with $\{\cup, +\}$ and $\{\cap, +\}$

All results so far have been established for equations with addition, union and intersection. In fact, the same results hold for equations using addition and *either* union *or* intersection. Establishing all results in this stronger form, in particular, requires rewriting the basic constructions of the previously known Theorems 2 and 3. The proof of the new Theorem 4 (to be specific, the constructions of Lemmata 3–4) also has to undergo some changes.

5.1 Two general translation lemmata

The first basic result is a simulation of a resolved system of a specific form using union, intersection and addition by an unresolved system that does not use intersection.

Consider resolved systems of equations over sets of numbers, as in Section 3. They are of the form

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

where φ_i may contain union, intersection and addition, as well as singleton constants.

This subsection defines a syntactical transformation of resolved equations of a particular kind into unresolved equations using only one Boolean operation (that is, either union or intersection).

A resolved system of equations is said to have a *chain dependency* of X from Y if the equation defining X is of the form $X = Y \cap \varphi$ or $X = Y \cup \varphi$, where φ is an arbitrary expression.

The following fact about solutions of systems of resolved equations without 0 in the constants can be easily proved using standard methods:

Proposition 1. *If* 0 *is an element of some component of the least solution of a resolved system of equations with only monotone and continuous operations, then at least one constant used in this system contains* 0*.*

Indeed, since the least solution of such a system is given by fixpoint iteration, the number 0 may only appear in this process if it is contained in one of the constants.

Lemma 6. *Let $X_i = \varphi_i(X_1, \ldots, X_n)$ be a resolved system of equations with union, intersection and addition and with constants from a set $\mathcal{C}$, where every constant contains only positive integers. Let $(S_1, \ldots, S_n)$ be its least solution. Assume that for every variable X_{i_0} there exists a subset of variables $\{X_i\}_{i \in I}$ containing X_{i_0}, such that*

- *the sets $\{S_i\}_{i \in I}$ are pairwise disjoint and their union is in $\mathcal{C}$, and*
- *the equations for all $\{X_i\}_{i \in I}$ are either all of the form $X_i = \bigcup_j \alpha_{ij}$, or all of the form $X_i = \bigcap_j \alpha_{ij} \cup C$, where C is a constant and $\alpha_{ij} = A_1 + \ldots + A_k$, with $k \geqslant 1$ and with each A_t being a constant or a variable.*

In addition, assume that there are no cyclic chain dependencies in the system. Then there exists an unresolved system with union and addition, with constants from $\mathcal{C}$, which has the unique solution $(S_1, \ldots, S_n)$.

Proof. Such a system is given directly by replacing each equation $X_i = \bigcap_j \alpha_{ij} \cup C_i$, where each α_i is a sum of constants and variables, by the following collection of inequalities:

$$X_i \subseteq \alpha_{ij} \cup C_i \quad (\text{for all } j) \tag{9}$$

In addition, for each group of variables $\{X_i\}_{i \in I}$, whose union of the group is a constant C_I, the following equation is added:

$$\bigcup_{i \in I} X_i = C_I. \tag{10}$$

The rest of the equations, which are of the form $X_i = \bigcup_j \alpha_{ij}$, with α_{ij} being a sum of variables and constants, are left as they are. Clearly, the least solution $(S_1, \ldots, S_n)$ of the former system is a solution of the new system. It remains to prove that no other solutions exist.

Assume for the sake of contradiction, that there is another solution $(S'_1, \ldots, S'_n)$. So there is a number $n \in S_i \,\Delta\, S'_i$ for some i. Such a number is called *wrong* or *wrong for* X_i. In particular, if $n \in S'_i \setminus S_i$, then n is said to be an *extra number* for X_i, and if $n \in S_i \setminus S'_i$, then n is a *missing number* for X_i.

Note that the supposed solution must have $0 \notin S'_i$ for all i. Indeed, every i belongs to some group of variables I, and then, by (10), $S'_i \subseteq C_I$. Since $0 \notin C_I$, zero may not be in S'_i. This, in particular, means that 0 cannot be a wrong number (as $0 \notin S_i$ by Proposition 1).

Fix $n > 0$ as the smallest wrong number. Then it can be proved that if this number is obtained as a nontrivial sum of variables and constants, it is equally obtained under the substitution of both solutions:

Claim 1. *If n is the smallest wrong number and $\alpha = A_1 + \ldots + A_k$, where $k \geqslant 2$ and all A_j are variables and constants, then $n \in \alpha(\ldots, S_i, \ldots)$ if and only if $n \in \alpha(\ldots, S'_i, \ldots)$.*

Proof. If $n \in \alpha(\ldots, S_i, \ldots)$, then $n = n_1 + \ldots + n_k$, with $n_j \in A_j(\ldots, S_i, \ldots)$. As all sets $A_j(\ldots, S_i, \ldots)$ are 0-free, each number n_j must be positive. Furthermore, each of them must be less than n because $k \geqslant 2$. Since n is the smallest wrong number, none of $n_1, \ldots, n_k$ is wrong for its respective variable, and hence $n_j \in A_j(\ldots, S'_i, \ldots)$. The same argument applies for the converse implication. ☐

Among all pairs (n, X_i), where n is the smallest wrong number and it is wrong for X_i, choose a pair such that n is an extra number for X_i, and if it is not possible, then a pair such that n is a missing number for X_i is chosen. Let us show that n must be wrong for another variable $X_{i'}$, with a chain dependency of $X_{i'}$ from X_i.

Suppose that X_i has an equation $X_i = \bigcup_j \alpha_{ij}$ in the original system, which is preserved in the new system. So $S_i = \bigcup_j \alpha_{ij}(\ldots, S_t, \ldots)$. Hence there exists α_{ij}, such that $n \in \alpha_{ij}(\ldots, S_t, \ldots) \,\Delta\, \alpha_{ij}(\ldots, S'_t, \ldots)$. Clearly this α_{ij} cannot be a constant. If it is a variable $X_{i'}$ then we replace S_i by $S_{i'}$. Note, that there is a chain dependency of S_i from $S_{i'}$ and n is wrong for $S'_{i'}$ and if n is an extra number, we can choose $S_{i'}$ so that n is still an extra number for $S_{i'}$. By Claim 1 α_{ij} cannot be a non-trivial sum of variables and constants.

Suppose now that the equation for X_i in the original system is of the form $X_i = \bigcap_j \alpha_{ij} \cup C_i$, and n is a missing number. We use (10) in this case—let $i \in I$ and $\bigcup_{j \in I} X_j = C_I$. Then by substituting S_i into those equations we obtain that $n \in C_I$. On the other hand by substituting S'_i into those equations we obtain that $n \in S'_{i'}$ for some $i' \in I$ and $i' \neq i$. As $n \in S_i$ then $n \notin S_{i'}$, as $i, i' \in I$ and by assumption sets in the same group are pairwise disjoint. Hence we obtain a contradiction, as n is an extra number for $X_{i'}$ and we are supposed to choose an extra number if there is any.

Let the equation for X_i in the original system be $X_i = \bigcap_j \alpha_{ij} \cup C$ and suppose that n is an extra number. So in the new system there are equations $X_i \subseteq \alpha_{ij} \cup C_i$ for $j \in I$, hence $n \in \alpha_{ij}(\ldots, S'_t, \ldots) \cup C_i$ for $j \in I$. On the other hand $n \notin S_i = \bigcap_{j \in I} \alpha_{ij}(\ldots, S_i, \ldots) \cup C_i$. And so there is $j' \in I$ such that $n \notin \alpha_{ij'}(\ldots, S_t, \ldots) \cup C_i$. Hence $n \in \alpha_{ij'}(\ldots, S'_t, \ldots) \setminus \alpha_{ij'}(\ldots, S_t, \ldots)$. Clearly $\alpha_{ij'}$ cannot be a constant, assuming that it is a non-trivial sum would again derive a contradiction by Claim 1. And so $\alpha_{ij'}$ is a variable $X_{i'}$. We replace X_i by $X_{i'}$ and continue the process. Note, that there is a chain dependency of X_i from $X_{i'}$ and n is an extra number for $X_{i'}$.

Now the same argument applies to the pair $(n, X_{i'})$, and in this way an infinite sequence of variables with a chain dependency to their successors is obtained. This is a contradiction, as there are no cyclic chain dependencies in the system. ☐

A similar construction produces equations with intersection instead of union. The next lemma is very similar in spirit and proof technique to Lemma 6, but some technical details are different, therefore it has to be proved separately.

Lemma 7. *Under the assumptions of Lemma 6, there exists an unresolved system with intersection and addition and with constants from $\mathcal{C}$, which has a unique solution that coincides with the least solution of the given system.*

Proof. Here the new system is obtained by the following transformation. For every equation $X_i = \bigcup_j \alpha_{ij}$ in the original system, where each α_{ij} is a sum of constants and variables, the new system contains inequalities

$$\alpha_{ij} \subseteq X_i \quad \text{for each } j. \tag{11}$$

For every subset of variables $\{X_i\}_{i \in I}$, with union C_I, the following equations are added:

$$X_i \cap X_j = \varnothing \qquad \text{for each } i, j \in I \text{ with } i \neq j, \tag{12}$$

$$X_i \subseteq C_I \qquad \text{for each } i \in I. \tag{13}$$

The rest of the equations are of the form $X_i = \bigcap_j \alpha_{ij} \cup C_i$, where C_i is a constant and $\alpha_{ij} = A_1 + \ldots + A_k$, with $k \geqslant 1$ and with each A_t being a constant or a variable. They are changed in the way similar to the equations for union, i.e. are replaced by inequalities

$$C_i \subseteq X_i \text{ and } \bigcap_j \alpha_{i,j} \subseteq X_i.$$

Clearly, the least solution $(\ldots, S_i, \ldots)$ of the former system is still a solution. It should be proved that no other solution exists.

As in Lemma 6, Proposition 1 is used to show that the least solution $(\ldots, S_i, \ldots)$ of the resolved system is 0-free. Also, since the assumptions of the lemma are the same as those of Lemma 6, then Claim 1 holds.

Suppose that there is another solution $(\ldots, S'_i, \ldots)$. Note that 0 may not be in any S'_i by the equation (13).

Define wrong numbers, missing numbers and extra numbers as in the proof of Lemma 6. Let n be the smallest wrong number with $n \in S_i \,\Delta\, S'_i$ for some i. By the above arguments, n must be positive. Among all pairs (n, X_i), such that n is the smallest wrong number and it is wrong for X_i, choose the one in which n is a missing number, if there is any such pair. If there is none, then choose a pair (n, X_i), where n is an extra number for X_i. As in the proof of the previous lemma, the idea is to show that there must be another variable $X_{i'}$ which has a chain dependence on X_i, so that n is a wrong number of $X_{i'}$.

Suppose first that n is an extra number. We use the (12) and (13) in this case: substituting $(\ldots, S_t, \ldots)$ into (13) one obtains that $n \in C_I$, where $i \in I$. On the other hand, by the assumption of the Lemma $\bigcup_{j \in I} S_j = C_I$, hence there exists $i' \neq i$ such that $n \in S_{i'}$. But by (12): $n \notin S'_{i'}$, as $S'_i \cap S'_{i'} = \varnothing$. Hence n is a missing number for i', a contradiction, as we were supposed to choose a missing number if there was any.

Assume now that n is a missing number and in the original resolved system the equation defining S_i is of the from $X_i = \bigcup_j \alpha_{ij}$. By the construction there are equations $\alpha_{ij} \subseteq X_i$ for $j \in I$, hence $n \notin \alpha_{ij}(\ldots, S'_t, \ldots)$ for $j \in I$. On the other hand $n \in S_i = \bigcup_{j \in I} \alpha_{ij}(\ldots, S_t, \ldots)$. Hence there is $i' \in I$ such that $n \in \alpha_{i'j}(\ldots, S_t, \ldots)$ and therefore $n \in \alpha_{i'j}(\ldots, S_t, \ldots) \setminus \alpha_{i'j}(\ldots, S'_t, \ldots)$. By Claim 1 $\alpha_{i'j}$ cannot be a non-trivial sum. Clearly it cannot be a constant, hence it is a variable. And so $\alpha_{i'j} = X_{i'}$. We swap S_i for $S_{i'}$. Note that there is a chain dependency of X_i from $X_{i'}$ and n is a missing number for $X'_{i'}$.

Suppose now that n is a missing number and in the original system the equation for S_i is of the form $X_i = \bigcap \alpha_{ij} \cup C_i$. The construction assures that there are equations $\bigcap_j \alpha_{ij} \subseteq X_i$ and $C_i \subseteq X_i$ in the new system. Then $n \notin \bigcap_j \alpha_{ij}(\ldots, S'_t, \ldots)$ and $n \notin C_i$. On the other hand, as $n \in L_i$, it holds that $n \in \bigcap_j \alpha_{ij}(\ldots, S_t, \ldots) \cup C_i$. Since $n \notin C_i$ by previous observation, $n \in \bigcap_j \alpha_{ij}(\ldots, S_t, \ldots)$. Thus there exists $i' \in I$ such that $n \in \alpha_{i'j}(\ldots, S_t, \ldots) \setminus \alpha_{i'j}(\ldots, S'_t, \ldots)$. Similarly to the analysis in the previous case, $\alpha_{i'j}$ cannot be a constant and by Claim 1 it cannot be a non-trivial sum as well. Hence $\alpha_{i'j} = X_{i'}$, for some variable $X_{i'}$, i.e. there is a chain dependency of X_i from $X_{i'}$. We replace S_i with $S_{i'}$, nota that n is a missing number for $X_{i'}$.

And so for every n and X_i for which it is wrong we are able of finding another $X_{i'}$ such that n is wrong for it as well and there is a chain dependency of S_i from $S_{i'}$. As there are no cyclic chain dependencies in the system we obtain a contradiction. ☐

The next task is to apply Lemmata 6 and 7 to resolved systems constructed in the proofs of Theorems 2 and 3. For the lemmata to be applicable, the equations given by Jeż [5] and Jeż and Okhotin [6] need to be decomposed into smaller parts and slightly changed. Then the variables can be grouped into subsets, as required by the lemmata.

5.2 Sets with a regular positional notation

Using the lemmata from the previous section, the resolved equations of Jeż [5] and Jeż and Okhotin [6] will now be converted to unresolved equations with sum and either union or intersection. The first task is to reformulate them so that Lemmata 6 and 7 are applicable.

The following known properties of equations over sets of numbers will be used in the constructions:

Lemma 8 ([6, Lem.3]). *Let $S \subseteq \mathbb{N}$ be a set of numbers, let k and k^m (with $k \geqslant 2$, $m \geqslant 2$) be two bases of positional notation. Then the language $L \subseteq \Sigma_k^* \setminus 0\Sigma_k^*$ of base-k notations* *of numbers in S is regular (linear conjunctive) if and only if the language $L' \subseteq \Sigma_{k^m}^* \setminus 0\Sigma_{k^m}^*$ of their base-k^m notations is regular (linear conjunctive, respectively).*

Lemma 9 ([6, Lem.4]). *Let $\varphi(X)$ be an expression defined as a composition of the following operations: (i) the variable X; (ii) constant sets; (iii) union; (iv) intersection with a constant set; (v) addition of a constant set. Then φ is distributive over infinite union, that is, $\varphi(X) = \bigcup_{n \in X} \varphi(\{n\})$.*

In addition, two transformations of systems of equations, which are intuitively obvious meta-theorems, will be used to convert equations over sets of numbers to the form required by Lemmata 6 and 7. One of them states that any components of a least solution may be replaced by constants with the same value:

Proposition 2. *Let a system*

$$\varphi_i(X_1, \ldots, X_m, Y_1, \ldots, Y_n) = \psi_i(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$$

have a least solution $X_i = K_i$, $Y_j = L_j$. Then the system

$$\varphi_i(K_1, \ldots, K_m, Y_1, \ldots, Y_n) = \psi_i(K_1, \ldots, K_m, Y_1, \ldots, Y_n)$$

in variables $\{Y_1, \ldots, Y_n\}$ has the least solution $Y_j = L_j$.

The other transformation is a decomposition of complex right-hand sides by introducing extra variables:

Proposition 3. *Let $(\ldots, S_t, \ldots)$ be the least solution of a system of equations in variables $(\ldots, X_t, \ldots)$ using union, intersection and addition, and let*

$$\varphi(\ldots, X_t, \ldots; \psi(\ldots, X_t, \ldots)) = \eta(\ldots, X_t, \ldots)$$

be one of its equations. Then a system with a new variable Y, a new equation $Y = \psi(\ldots, X_t, \ldots)$, and with the above equation replaced by

$$\varphi(\ldots, X_t, \ldots; Y) = \eta(\ldots, X_t, \ldots)$$

has the least solution $(\ldots, S_t, \ldots; \psi(\ldots, S_t, \ldots))$.

Note that the property of having a least solution is maintained because the subexpression ψ contains only monotone operations on sets.

Now the first result on the expressive power of equations with one Boolean operation asserts representability of finite and co-finite sets of numbers.

Lemma 10. *Every finite or co-finite subset of $\mathbb{N}$ is representable by a unique solution of a resolved system with union and addition, as well as by a unique solution of an unresolved system with intersection and addition.*

Proof. The case of union follows from the fact that every ultimately periodic unary language can be specified by a resolved system of language equations with union, one-sided concatenation and constants $\{a\}$ and $\{\varepsilon\}$.

Let us prove the lemma in the case of intersection, where the use of unresolved equations becomes essential. Let $K = \{n_1, n_2, \ldots, n_m\}$, with $0 \leqslant n_1 < \ldots < n_m$, be any finite set of numbers. First define the following equations for a variable X:

$$n_m + 1 \subseteq X \tag{14a}$$
$$X + 1 \subseteq X \tag{14b}$$
$$n_m \cap X = \varnothing \tag{14c}$$

Here (14b) ensures that the solution is of the form $\{\, n \mid n \geqslant k\}$ for some k (or empty), (14a) states that $n_m + 1$ is in X, while (14c) ensures that n is not in X. Thus the unique solution of these equations is $X = \{\, n \mid n > n_m\}$. Using this variable, define three more equations for a new variable Y:

$$X \cap Y = \varnothing \tag{14d}$$
$$n_i \subseteq Y \quad \text{for } i \in \{1, 2, \ldots, m\} \tag{14e}$$
$$n \cap Y = \varnothing \quad \text{for each } n < n_m \text{ with } n \notin K \tag{14f}$$

By (14d), Y must be a subset of $\{0, \ldots, n_m\}$. The next two equations state the membership of every number between 0 and n_m in Y: it should be in Y if and only if it is in K. Hence, the unique solution is $Y = K$. Finally, define one more variable Z, with the following equations:

$$X \subseteq Z \tag{14g}$$
$$n_i \cap Z = \varnothing \quad \text{for } i = 1, 2, \ldots, m \tag{14h}$$
$$n_i \subseteq Z \quad \text{for } n_i < n_m, n_i \notin \{n_1, \ldots, n_m\} \tag{14i}$$

The equation (14g) states that every number greater than n_i must be in Z. The next two equations define, similarly to the equations for Y, for each number not exceeding n_m, that it should be in Z if and only if it is not in K. Altogether these equations specify $Z = \mathbb{N} \setminus K$, which completes the proof. □

Consider a set of natural numbers with base-k notation $ij\mathtt{0}^*$ for $i \neq \mathtt{0}$. It is known that such sets are representable by resolved systems with union, intersection and sum [5]. This result will now be reconstructed to use only one Boolean operation, at the expense of turning the resolved equations into unresolved ones. The new construction is based upon a slightly modified version of equations from the original paper [5]. The proof that they have a stated solution is omitted, as it is exactly the same as the original one.

Theorem 5. *For every $k \geqslant 9$, there exists an unresolved system with union (intersection), sum and singleton constants, which has a unique solution with some of its components being*

$$(ij\mathtt{0}^*)_k \quad \text{(for all } i, j \in \Sigma_k \text{ with } i > 0\text{)}.$$

Proof. It is known that there exists a resolved system of equations with union, intersection and addition representing the sets $S_{ij} = (ij\mathtt{0}^*)_k$ through each other [5, Thm.14]. However, this system would not be sufficient for the present paper, since these sets cannot be grouped to match the conditions of Lemmata 6 and 7. The proposed construction relies on representing both these sets and the complementary sets $\widetilde{S}_{ij} = (ij(\Sigma_k^* \setminus \mathtt{0}^*))_k$. Then all sets S_{ij} and $\widetilde{S}_{ij}$ will be pairwise disjoint and their union will be co-finite, making the lemmata applicable.

Define the set of variables $X_{i,j}$, $Y_{i,j}$, $X_{i,j,\ell}$ and $Y_{i,j,\ell}$, with $i, j, \ell \in \Sigma_k$ and $i \neq \mathtt{0}$, and consider the following resolved system of equations:

$$\begin{aligned}
X_{1,j} &= \bigcap_{n=1}^{2} X_{k-n,0} + X_{j+n,0} \cup (\mathtt{1}j)_k && \text{for } j = \mathtt{0}, \mathtt{1}, \mathtt{2} \\
X_{i,j} &= \bigcap_{n=1}^{2} X_{i-1,k-n} + X_{j+n,0} \cup (ij)_k && \text{for } j = \mathtt{0}, \mathtt{1}, \mathtt{2},\ i \geqslant 2 \\
X_{i,j} &= \Big(\bigcap_{n=1}^{2} X_{i,j-n} + X_{n,0}\Big) \\
&\quad \cap\ X_{i,0} + X_{j,0} \cup (ij)_k && \text{for } j \geqslant 3 \\
X_{i,j,\ell} &= \bigcap_{n=0}^{3} X_{i,n} + X_{j-n,\ell} && \text{for } j \geqslant 4,\ i \neq \mathtt{0},\ \ell \in \Sigma_k, \\
X_{i,j,\ell} &= \bigcap_{n=1}^{4} X_{i-1,j+n} + X_{k-n,\ell} && \text{for } j \leqslant 3,\ i \neq \mathtt{0}, \mathtt{1},\ \ell \in \Sigma_k, \\
X_{1,j,\ell} &= \bigcap_{n=1}^{4} X_{k-n,0} + X_{j+n,\ell} && \text{for } j \leqslant 3,\ \ell \in \Sigma_k, \\
Y_{i,j} &= \bigcup_{\ell \neq \mathtt{0}} X_{i,j,\ell} \cup \bigcup_{\ell \in \Sigma_k} Y_{i,j,\ell} && \text{for } j \in \Sigma_k,\ i \neq \mathtt{0}, \\
Y_{i,j,\ell} &= \bigcap_{n=0}^{3} X_{i,n} + Y_{j-n,\ell} && \text{for } j \geqslant 4,\ i \neq \mathtt{0},\ \ell \in \Sigma_k, \\
Y_{i,j,\ell} &= \bigcap_{n=1}^{4} X_{i-1,j+n} + Y_{k-n,\ell} && \text{for } j \leqslant 3,\ i \neq \mathtt{0}, \mathtt{1},\ \ell \in \Sigma_k, \\
Y_{1,j,\ell} &= \bigcap_{n=1}^{4} X_{k-n,0} + Y_{j+n,\ell} && \text{for } j \leqslant 3,\ \ell \in \Sigma_k.
\end{aligned}$$

It is claimed that the least solution of those equations is:

$$\begin{aligned}
X_{i,j} &= (ij\mathtt{0}^*)_k, \\
X_{i,j,\ell} &= (ij\ell\mathtt{0}^*)_k, \\
Y_{i,j} &= (ij(\Sigma_k^* \setminus \mathtt{0}^*))_k, \\
Y_{i,j,\ell} &= (ij\ell(\Sigma_k^* \setminus \mathtt{0}^*))_k.
\end{aligned}$$

The equations for $X_{i,j}$ are already known [5, Thm.14] and their least solution has been proved to be $(ij\mathtt{0}^*)_k$, as claimed. The rest of the equations implicitly occur in a proof of an earlier more general result [5, Lem.17], yet some explanations are due in order to recognize them in the above system. This is in fact the construction of Theorem 2, used to represent a set of numbers $(L(M))_k$ for a given finite automaton M. Consider that

$\Sigma_k^* \setminus \mathtt{0}^*$ is a regular language recognized by a finite automaton reading the string of digits from the right to the left. The automaton has two states, q_0 and q_1; it is in state q_0 while all digits encountered so far are zeroes, and once any non-zero digit is read, it enters state q_1 and remains there. Applying the known construction [5, Lem.17] to this automaton gives a system in variables $X_{i,j}$ and $Y_{i,j}$, with the X-variables corresponding to the state q_0 and with the Y-variables representing the state q_0. A straightforward transformation of that system

It remains to show that these equations satisfy the assumptions of Lemmata 6 and 7, with the variables separated into the following two groups:

$$\{\, X_{i,j}, Y_{i,j} \mid i,j \in \Sigma_k,\ i \neq \mathtt{0}\},\ \{\, X_{i,j,\ell}, Y_{i,j,\ell} \mid i,j,\ell \in \Sigma_k,\ i \neq \mathtt{0}\}.$$

The unions of the corresponding sets in the least solution for the former group is $\{\, n \mid n \geqslant k\}$, and for the latter group it is $\{\, n \mid n \geqslant k^2\}$; both are co-finite sets. Clearly, in either group all the components are pairwise disjoint. The only chain dependencies are those of variables $X_{i,j}$ on (some) variables $X_{i,j,\ell}$, as well as of $Y_{i,j}$ on some $Y_{i,j,\ell}$; hence there are no cyclic chain dependencies. And so by Lemma 6 and Lemma 7 there exist unresolved systems with union (intersection), sum and finite and co-finite constants, whose unique solution has the requested components. Co-finite and finite constants are eliminated by expressing them according to Lemma 10. ☐

Now the construction of Theorem 2 can be remade using unresolved equations using only one Boolean operation.

Lemma 11. *For every deterministic finite automaton $M = (\Sigma, Q, q_0, \delta, F)$ there exists an unresolved system of equations using union (intersection), sum and singleton constants, in which some of the components of the unique solution are*

$$S_{i,j,q} := \{\, (ijw)_k \mid \delta(q_0, w^R) = q\} \qquad \text{for } i,j \in \Sigma_k, i \neq \mathtt{0}, q \in Q.$$

Proof. Consider the following resolved language equations [5, Lem. 17] with constants of the form $(ij\mathtt{0}^*)_k$:

$$\begin{aligned}
X_{i,j,q} &= \bigcup_{(x,q'):\delta(q',\ell,q)} X_{i,j,\ell,q'} \cup \{\, (ij)_k \mid \text{ if } q = q_0\} && \text{for } j \geqslant 4,\ i \geqslant 1,\\
X_{i,j,\ell,q} &= \bigcap_{n=0}^{3} (in\mathtt{0}^*)_k + X_{j-n,\ell,q} && \text{for } j \geqslant 4,\ i \geqslant 1,\\
X_{i,j,q} &= \bigcup_{(\ell,q'):\delta(q',\ell,q)} X_{i,j,\ell,q'} \cup \{\, (ij)_k \mid \text{ if } q = q_0\} && \text{for } j \leqslant 3,\ i \geqslant 2,\\
X_{i,j,\ell,q} &= \bigcap_{n=1}^{4} ((i-1)(j+n)\mathtt{0}^*)_k + X_{k-n,\ell,q} && \text{for } j \leqslant 3,\ i \geqslant 2,\\
X_{1,j,q} &= \bigcup_{(\ell,q'):\delta(q',\ell,q)} X_{1,j,\ell,q'} \cup \{\, (ij)_k \mid \text{ if } q = q_0\} && \text{for } j \leqslant 3,\\
X_{1,j,\ell,q} &= \bigcap_{n=1}^{4} ((k-n)\mathtt{0}\mathtt{0}^*)_k + X_{j+n,\ell,q} && \text{for } j \leqslant 3
\end{aligned}$$

For these equations, it was proved that their least solution is

$$X_{i,j,q} = \{\, (ijw)_k \mid \delta(q_0, w^R) = q\}, \quad X_{i,j,\ell,q} = \{\, (ij\ell w)_k \mid \delta(q_0, w^R) = q\}.$$

The rest of the proof shows how to obtain an unresolved system of equations with the same unique solution, which is done similarly to the proof of Theorem 5.

The plan is to apply Lemmata 6 and 7 to the above system. To this end, the variables of the system have to be grouped. Again, there will be two groups,

$$\{\, X_{i,j,q} \mid i, j \in \Sigma_k, i \neq 0, q \in Q\} \quad \text{and} \quad \{\, X_{i,j,\ell,q} \mid i, j, \ell \in \Sigma_k, i \neq 0, q \in Q\}.$$

The union of the least solution in the first group is $\{\, n \mid n \geqslant k\}$, and $\{\, n \mid n \geqslant k^2\}$ for the second group. The sets within each group are clearly disjoint.

The resulting system uses two co-finite constants obtained as unions of the groups, as well as constants of the form $(ij\mathtt{0}^*)_k$. The former are expressed as in Lemma 10, while the latter are replaced by references to equations from Theorem 5. ☐

Theorem 6. *For every $k \geqslant 2$ and for every regular language $L \subseteq \Sigma_k^* \setminus \mathtt{0}\Sigma_k^*$ there exists an unresolved system with union (intersection), addition and singleton constants, which has a unique solution with $(L)_k$ as one of its components.*

Proof. First consider the case of $2 \leqslant k < 9$. Then, by Lemma 8, there exists a regular language $L' \subseteq \Sigma_{k'}^*$ for $k' = k^4 > 9$, such that $(L')_{k'} = (L)_k$. Hence it is sufficient to establish the theorem for $k \geqslant 9$.

Let $M = (\Sigma, Q, q_0, \delta, F)$ be a deterministic finite automaton recognizing L^R. By Lemma 11, there exists an unresolved system of the specified form, in which every variable $X_{i,j,q}$ in the unique solution equals $\{\, (ijw)_k \mid \delta(q_0, w^R) = q\}$. Then the set $(L)_k$ can be obtained as the following union:

$$(L)_k = \Big(\underbrace{(L)_k \cap \{\, n \mid n < k\}}_{\text{finite constant}}\Big) \cup \bigcup_{\substack{i,j,q: \\ \delta(q,ji) \in F}} \underbrace{\{\, (ijw)_k \mid \delta(q_0, w^R) = q\}}_{X_{i,j,q}}. \qquad (15)$$

In the case of unresolved equations with union, the equality (15) can be directly specified by introducing a new variable Y and adding the following equation:

$$Y = \big((L)_k \cap \{\, n \mid n < k\}\big) \cup \bigcup_{\substack{i,j,q: \\ \delta(q,ji) \in F}} X_{i,j,q}.$$

The finite constant $\{\, n \mid n < k\}$ is expressed according to Lemma 10.

For the case of intersection, consider that the sets $\{\, (ijw)_k \mid \delta(q_0, w^R) = q\}$, along with the finite set $\{\, n \mid n < k\}$, form a partition of $\mathbb{N}$. Then a new variable Y is added, and its intersection with every element of this partition is expressed:

$$\begin{aligned} Y \cap \{\, n \mid n < k\} &= (L \cap \Sigma_k^{\leqslant 1})_k \\ Y \cap X_{i,j,q} &= \varnothing && \text{for } (i, j, q) \text{ with } \delta(q, ji) \notin F \\ Y \cap X_{i,j,q} &= X_{i,j,q} && \text{for } (i, j, q) \text{ with } \delta(q, ji) \in F \end{aligned}$$

Because these equalities state the membership of *every* natural number in Y, this representation is equivalent to (15), and hence the system has a unique solution with $Y = (L)_k$. Both finite constants are again replaced according to Lemma 10. ☐

5.3 Any linear conjunctive language

The next task is to remake another key construction of a system of equations using only one Boolean operation. As stated in Theorem 3, for every trellis automaton M with $L(M) \subseteq \Sigma_k^+ \setminus \mathtt{0}\Sigma_k^*$, there exists a resolved system of equations over sets of natural numbers with $(L(M))_k$ as one of the components of its least solution. This construction essentially uses both union and intersection, and the goal is again to refine the known construction [6] so that Lemmata 6 and 7 could be applied to it.

This construction essentially uses the operations of symbolic addition and subtraction of 1 on positional notations of numbers. For every base $k \geqslant 2$ and for every string $w \in \Sigma_k^* \setminus (k-1)^*$, the string $w' = w \boxplus \mathtt{1}$ is defined as the unique string with $|w| = |w'|$ and $(w)_k + 1 = (w')_k$. Similarly, for every $w \in \Sigma_k^* \setminus \mathtt{0}^*$, define $w' = w \boxminus \mathtt{1}$ as the unique string with $|w| = |w'|$ and $(w)_k - 1 = (w')_k$.

For example, in decimal notation, $\mathtt{0099} \boxplus \mathtt{1} = \mathtt{0100}$. and $\mathtt{0100} \boxminus \mathtt{1} = \mathtt{0099}$. This notation shall never be used for strings on which it is undefined, such as $\mathtt{999} \boxplus \mathtt{1}$ and $\mathtt{000} \boxminus \mathtt{1}$. This notation is extended to languages in the natural way:

$$L \boxplus \mathtt{1} = \{\, w \boxplus \mathtt{1} \mid w \in L \setminus (k-1)^* \}$$
$$L \boxminus \mathtt{1} = \{\, w \boxminus \mathtt{1} \mid w \in L \setminus \mathtt{0}^* \}$$

This operation obviously preserves regularity, hence it can be used inside regular expressions for sets of positional notations, and the sets thus defined remain regular.

The original construction of a resolved system simulating a trellis automaton went in three stages: first, the set $(\mathtt{1}(L(M) \boxminus \mathtt{1})\mathtt{10}^*)_k$ was represented [6, Lem.5]; next, $(\mathtt{1} \cdot L(M))_k$ [6, Lem.6]; and finally, a system for $(L(M))_k$ was obtained [6, Lem.7]. This composition will be followed in the below proof, and each part of the known construction will be carefully remade.

Lemma 12. *For every $k \geqslant 4$ and for every trellis automaton M over $\Sigma_k = \{\mathtt{0}, \ldots, k-1\}$ with $L(M) \cap \mathtt{0}\Sigma_k^* = \varnothing$, there exists and can be effectively constructed an unresolved system of equations over sets of natural numbers using union and addition (or intersection and addition) and singleton constants, such that the unique solution of this system contains a component*

$$(\mathtt{1}(L_M(q) \boxminus \mathtt{1})\mathtt{10}^*)_k = \{\, (\mathtt{1}w\mathtt{10}^\ell)_k \mid \ell \geqslant 0, w \notin (k-1)^*, w \boxplus \mathtt{1} \in L_M(q)\}.$$

Proof. Let $M = (\Sigma_k, Q, I, \delta, F)$ be any trellis automaton and consider the known resolved system of equations representing the given sets of numbers [6, Lem.5]. It uses variables X_q over all $q \in Q$ and contains the equations

$$X_q = R_q \ \cup \bigcup_{\substack{q,q' : \delta(q',q'')=q \\ i,j \in \Sigma_k}} \lambda_i(X_{q''}) \cap \rho_j(X_{q'}) \quad (\text{for all } q \in Q)$$

where

$$\begin{aligned}
R_q &= \{\, (\mathtt{1}(w \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k \mid w \in \mathtt{0}^*(\Sigma_k \setminus \mathtt{0}) \cup (\Sigma_k \setminus \mathtt{0})\mathtt{0}^*,\ w \in L_M(q)\} \\
\kappa_{i'}(X) &= \big(X \cap (\mathtt{1}i'\Sigma_k^*\mathtt{1}\mathtt{0}^*)_k\big) + (\mathtt{1}\mathtt{0}^*)_k \ \cap\ (\mathtt{2}i'\Sigma_k^*)_k, \quad \text{for all } i' \in \Sigma_k \\
\lambda_i(X) &= \bigcup_{i' \in \Sigma_k} \big(\kappa_{i'}(X) + ((k+i-2)\mathtt{0}^*)_k \cap (\mathtt{1}i\Sigma_k^*)_k\big), \quad \text{for } i = 0, 1 \\
\lambda_i(X) &= \bigcup_{i' \in \Sigma_k} \big(\kappa_{i'}(X) + (\mathtt{1}(i-2)\mathtt{0}^*)_k \cap (\mathtt{1}i\Sigma_k^*)_k\big), \quad \text{for } i \geqslant 2 \\
\pi_{j'}(X) &= \big(X \cap (\mathtt{1}\Sigma_k^*j'\mathtt{1}\mathtt{0}^*)_k\big) + (\mathtt{1}\mathtt{0}^*)_k \ \cap\ (\mathtt{1}\Sigma_k^*j'\mathtt{2}\mathtt{0}^*)_k, \quad \text{for all } j' \in \Sigma_k \\
\rho_j(X) &= \bigcup_{j' \in \Sigma_k} \big(\pi_{j'}(X) + ((k+j-2)\mathtt{1}\mathtt{0}^*)_k \cap (\mathtt{1}\Sigma_k^*j\mathtt{1}\mathtt{0}^*)_k\big), \quad \text{for } j = 0, 1 \\
\rho_j(X) &= \bigcup_{j' \in \Sigma_k} \big(\pi_{j'}(X) + (\mathtt{1}(j-2)\mathtt{1}\mathtt{0}^*)_k \cap (\mathtt{1}\Sigma_k^*j\mathtt{1}\mathtt{0}^*)_k\big), \quad \text{for } 2 \leqslant j \leqslant k-2 \\
\rho_{k-1}(X) &= \bigcup_{j' \in \Sigma_k} \big(\pi_{j'}(X) + ((k-3)\mathtt{1}\mathtt{0}^*)_k \cap (\mathtt{1}\Sigma_k^*(k-1)\mathtt{1}\mathtt{0}^*)_k\big)
\end{aligned}$$

All constants used in the system have regular base-k notation.

The least solution is $X_q = S_q$ [6, Main Claim], where

$$S_q = (\mathtt{1}((L_M(q) \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k = \{\, (\mathtt{1}w\mathtt{1}\mathtt{0}^\ell)_k \mid \ell \geqslant 0, w \notin (k-1)^*, w \boxplus \mathtt{1} \in L_M(q)\}.$$

These sets are pairwise disjoint and their union is a set with a regular base-k notation. In order to prove this, let us establish a more general statement that will be used several times in the following:

Claim 2. *Let $x \in \Sigma_k^* \setminus \mathtt{0}\Sigma_k^*$ and $y \in \Sigma_k^+ \setminus \mathtt{0}^*$ be strings of digits, let $K_1, \ldots, K_m \subseteq \Sigma_k^+$ be any pairwise disjoint languages. Let $S_1, \ldots, S_m$ be sets of numbers defined by*

$$S_t = \{\, (xuy\mathtt{0}^\ell)_k \mid \ell \geqslant 0,\ u \in K_t\}.$$

Then these sets are pairwise disjoint and their union is

$$\bigcup_{t=1}^m S_t = (x(\textstyle\bigcup_{t=1}^m K_t)y\mathtt{0}^*)_k.$$

Proof. Consider any two sets S_t and $S_{t'}$ with $t \neq t'$, and suppose there is a number n belonging to both sets. Then $n = (xuy\mathtt{0}^\ell)_k$ for some $u \in K_t$ and $n = (xu'y\mathtt{0}^{\ell'})_k$ with $u' \in K_{t'}$. Since y contains a non-zero digit, the length of the tail of zeroes in n is independent of u and u', and therefore $\ell = \ell'$. Then u and u' must be the same string, which is impossible since $K_t \cap K_{t'} = \varnothing$ by assumption. This proves that $S_t \cap S_{t'} = \varnothing$.

The union of these sets is

$$\bigcup_t S_t = \bigcup_t (xK_ty\mathtt{0}^*)_k = (x(\textstyle\bigcup_t K_t)y\mathtt{0}^*)_k,$$

as stated. □

Now Claim 2 can be applied to the particular case of the sets S_q to obtain the following result:

Claim 3. *The sets of numbers* S_q *with different* $q \in Q$ *are pairwise disjoint, and their union is*

$$\bigcup_q S_q = (\mathtt{1}(\Sigma_k^+ \setminus (k-1)^*)\mathtt{1}\mathtt{0}^*)_k.$$

Proof. As a trellis automaton computes a uniquely determined state $\delta(I(w)) \in Q$ on each string $w \in \Sigma_k^+$, it induces a partition of Σ_k^+ into classes corresponding to different states. Define $K_q = (L_M(q) \setminus \mathtt{0}^*) \boxminus \mathtt{1}$; these sets are pairwise disjoint and their union for all $q \in Q$ is $\Sigma_k^* \setminus (k-1)^*$, since every string $w \in \Sigma_k^+$ belongs to some $L_M(q)$. The rest is given by Claim 2 with $x = y = \mathtt{1}$. ☐

Though the values of the variables X_q as they are already satisfy Lemma 6 and Lemma 7, the right-hand sides of the above equations are not of the required simple form. Now the goal is to transform the system, splitting the existing equations into smaller parts and introducing new variables, so that it satisfies the assumptions of the lemmata.

The first step is to construct equations of the required form representing λ and ρ. Each occurrence of $\lambda_i(X_q)$ will be replaced by a new variable $Z^\lambda_{i,q}$, and similarly $\kappa_{i'}(X_q)$ is replaced by $W^\lambda_{i',q}$, where the new variables have the following equations:

$$U^\lambda_{i',q} = X_q \cap (\mathtt{1}i'\Sigma_k^*\mathtt{1}\mathtt{0}^*)_k \tag{16}$$

$$W^\lambda_{i',q} = U^\lambda_{i',q} + (\mathtt{1}\mathtt{0}^*)_k \cap (\mathtt{2}i'\Sigma_k^*)_k \tag{17}$$

$$Y^\lambda_{i,i',q} = W^\lambda_{i',q} + (\mathtt{1}(i-2)\mathtt{0}^*)_k \cap (\mathtt{1}i\Sigma_k^*)_k \qquad \text{for } i \geqslant 3 \tag{18}$$

$$Y^\lambda_{i,i',q} = W^\lambda_{i',q} + ((k+i-2)\mathtt{0}^*)_k \cap (\mathtt{1}i\Sigma_k^*)_k \qquad \text{for } i \leqslant 2 \tag{19}$$

$$Z^\lambda_{i,q} = \bigcup_{i'} Y^\lambda_{i,i',q} \tag{20}$$

Since the equation for $Z^\lambda_{i,q}$ represents the expression $\lambda_i(X_q)$ broken into pieces, the "old variables" $\{X_q\}$ have the same values in the least solution of the new system as in the least solution of the old system. The newly introduced variables are arranged into the following four groups:

$$\{\, U^\lambda_{i',q} \mid i' \in \Sigma_k, q \in Q\}, \quad \{\, W^\lambda_{i',q} \mid i' \in \Sigma_k, q \in Q\},$$
$$\{\, Y^\lambda_{i,i',q} \mid i, i' \in \Sigma_k, q \in Q\}, \quad \{\, Z^\lambda_{i,q} \mid i \in \Sigma_k, q \in Q\}.$$

Let us calculate the values of these variables in the least solution. For every variable V, let $\mu(V)$ be the set corresponding to V in the least solution of the new system of equations.

Claim 4. *For all* $(i'_1, q_1) \neq (i'_2, q_2)$ *sets* $\mu(U^\lambda_{i'_1,q_1})$ *and* $\mu(U^\lambda_{i'_2,q_2})$ *are disjoint and and their union is*

$$\bigcup_{i',q} \mu(U^\lambda_{i',q}) = (\mathtt{1}(\Sigma_k^+ \setminus (k-1)^*)\mathtt{1}\mathtt{0}^*)_k.$$

Proof. It is already known [6, Eq. (3)] that

$$\mu(U^\lambda_{i',q}) = \{\, (\mathtt{1}i'w\mathtt{1}\mathtt{0}^\ell)_k \mid \ell \geqslant 0, i'w \notin (k-1)^*, i'w \boxplus \mathtt{1} \in L_M(q)\}.$$

These sets are obtained from the languages $K^{\lambda}_{i',q} = ((L_M(q) \setminus \mathtt{0}^*) \boxminus \mathtt{1}) \cap i'\Sigma_k^*$ with $i' \in \Sigma_k$ and $q \in Q$ as in the statement of Claim 2. To see that the sets $K^{\lambda}_{i',q}$ are pairwise disjoint, consider $K^{\lambda}_{i_1,q_1}$ and $K^{\lambda}_{i_2,q_2}$: if $i_1 \neq i_2$, then the words in these sets start from different digits, and if $q_1 \neq q_2$, then $K^{\lambda}_{i_1,q_1} \subseteq L_M(q_1) \boxminus \mathtt{1}$ and $K^{\lambda}_{i_2,q_2} \subseteq L_M(q_2) \boxminus \mathtt{1}$. In both cases, $K^{\lambda}_{i_1,q_1} \cap K^{\lambda}_{i_2,q_2} = \varnothing$.

Therefore, Claim 2 with $x = y = \mathtt{1}$ asserts that $\mu(U^{\lambda}_{i',q})$ are pairwise disjoint and the union of this group of sets is

$$\bigcup_{i',q} \mu(U^{\lambda}_{i',q}) = (\mathtt{1}(\textstyle\bigcup_{i',q} K_{i',q})\mathtt{1}\mathtt{0}^*)_k = (\mathtt{1}\big(\textstyle\bigcup_{i',q}((L_M(q) \setminus \mathtt{0}^*) \boxminus \mathtt{1}) \cap i'\Sigma_k^*\big)\mathtt{1}\mathtt{0}^*)_k =$$

$$= (\mathtt{1}((\textstyle\bigcup_q L_M(q) \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k = (\mathtt{1}((\Sigma_k^* \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k = (\mathtt{1}(\Sigma_k^+ \setminus (k-1)^*)\mathtt{1}\mathtt{0}^*)_k,$$

which completes the proof. ☐

Similar statements will now be proved for the other three groups of variables.

Claim 5. *For all $(i_1, q_1) \neq (i_2, q_2)$ the sets $\mu(W^{\lambda}_{i_1,q_1})$ and $\mu(W^{\lambda}_{i_1,q_1})$ are disjoint and the union of all sets in the group is:*

$$\bigcup_{i',q} W^{\lambda}_{i',q} = (\mathtt{2}(\Sigma_k^+ \setminus (k-1)^*)\mathtt{1}\mathtt{0}^*)_k.$$

Proof. It is known [6, Eq. (4)] that

$$\mu(W^{\lambda}_{i',q}) = \{\, (\mathtt{2}i'w\mathtt{1}\mathtt{0}^{\ell})_k \mid \ell \geqslant 0, i'w \notin (k-1)^*, i'w \boxplus \mathtt{1} \in L_M(q)\}.$$

These sets are induced by $K^{\lambda}_{i',q} = ((L_M(q) \setminus \mathtt{0}^*) \boxminus \mathtt{1}) \cap i'\Sigma_k^*$ with $i' \in \Sigma_k$ and $q \in Q$ as in Claim 2 with $x = \mathtt{2}$ and $y = \mathtt{1}$. It has been proved in Claim 4 that $K^{\lambda}_{i',q}$ are pairwise disjoint and their union is $\Sigma_k^+ \setminus (k-1)^+$. Both statements of the present claim follow. ☐

Claim 6. *For all $(i_1, i'_1, q_1) \neq (i_2, i'_2, q_2)$ it holds that $\mu(Y^{\lambda}_{i_1,i'_1,q_1}) \cap \mu(Y^{\lambda}_{i_1,i'_1,q_1}) = \varnothing$. Their union is:*

$$\bigcup_{i,i',q} \mu(Y^{\lambda}_{i,i',q}) = (\mathtt{1}\Sigma_k(\Sigma_k^+ \setminus (k-1)^*)\mathtt{1}\mathtt{0}^*)_k.$$

Proof. It is known [6, Eqs. (5,6)] that

$$\mu(Y^{\lambda}_{i,i',q}) = \{\, (\mathtt{1}ii'w\mathtt{1}\mathtt{0}^{\ell})_k \mid \ell \geqslant 0, i'w \notin (k-1)^*, i'w \boxplus \mathtt{1} \in L_M(q)\}.$$

Then, for each fixed i, those sets are obtained from the languages $K^{\lambda}_{i',q} = ((L_M(q) \setminus \mathtt{0}^*) \boxminus \mathtt{1}) \cap i'\Sigma_k^*$ as in Claim 2 with $x = \mathtt{1}i$ and $y = \mathtt{1}$. It was shown in Claim 4 that $K^{\lambda}_{i',q}$ are pairwise disjoint and their union is $\Sigma_k^+ \setminus (k-1)^*$. Thus, for each i,

$$\bigcup_{i',q} \mu(Y^{\lambda}_{i,i',q}) = (\mathtt{1}i(\Sigma_k^+ \setminus (k-1)^*)\mathtt{1}\mathtt{0}^*)_k,$$

and for all $(i'_1, q_1) \neq (i'_2, q_2)$ the sets $\mu(Y^{\lambda}_{i,i'_1,q_1})$ and $\mu(Y^{\lambda}_{i,i'_2,q_2})$ are disjoint. Then, clearly,

$$\bigcup_{i,i',q} \mu(Y^{\lambda}_{i,i',q}) = \bigcup_{i} (\mathtt{1}i(\Sigma_k^+ \setminus (k-1)^*)\mathtt{1}\mathtt{0}^*)_k = (\mathtt{1}\Sigma_k(\Sigma_k^+ \setminus (k-1)^*)\mathtt{1}\mathtt{0}^*)_k,$$

What is left to show is that for $(i_1, i'_1, q_1) \neq (i_2, i'_2, q_2)$, the sets $\mu(Y_{i_1,i'_1,q_1})$ and $\mu(Y_{i_2,i'_2,q_2})$ are disjoint. If $i_1 = i_2$, then $(i'_1, q_1) \neq (i'_2, q_2)$, and such sets were already shown to have empty intersection. If $i_1 \neq i_2$, then these sets consist of numbers with a different second leading digit, and are bound to be disjoint as well. ☐

Claim 7. *For all* $(i_1, q_1) \neq (i_2, q_2)$*, the sets* $\mu(Z^\lambda_{i_1,q_1})$ *and* $\mu(Z^\lambda_{i_2,q_2})$ *are disjoint, and their union equals*

$$\bigcup_{i,q} \mu(Z^\lambda_{i,q}) = (\mathtt{1}\Sigma_k(\Sigma_k^+ \setminus (k-1)^*)\mathtt{10}^*)_k.$$

Proof. The equation (20) defines $Z^\lambda_{i,q}$ as the union of $Y^\lambda_{i,i',q}$ for all i', and the values of the latter variables are known from Claim 6. Then the value of $Z^\lambda_{i,q}$ is calculated as follows:

$$\bigcup_{i,q} \mu(Z^\lambda_{i,q}) = \bigcup_{i,q} (\textstyle\bigcup_{i'} \mu(Y^\lambda_{i,i',q})) = \displaystyle\bigcup_{i,i',q} \mu(Y^\lambda_{i,i',q}) = (\mathtt{1}\Sigma(\Sigma_k^+ \setminus (k-1)^*)\mathtt{10}^*)_k.$$

The sets $\mu(Z^\lambda_{i,q})$ are pairwise disjoint as unions of pairwise disjoint sets. ☐

The equations for ρ will now undergo a similar reconstruction. Every $\rho_j(X_q)$ is replaced by $U^\rho_{j,q}$ and each $\pi_{j'}(X_q)$ by $W^\rho_{j',q}(X_q)$. The new variables are defined by the following resolved equations:

$$\begin{aligned}
U^\rho_{j',q} &= X_q \cap (\mathtt{1}\Sigma_k^* j'\mathtt{10}^*)_k && (21)\\
W^\rho_{j',q} &= U^\rho_{j',q} + (\mathtt{10}^*)_k \;\cap\; (\mathtt{1}\Sigma_k^* j'\mathtt{20}^*)_k && (22)\\
Y^\rho_{j,j',q} &= W^\rho_{j',q} + (\mathtt{1}(k+j-2)\mathtt{10}^*)_k \;\cap\; (\mathtt{1}\Sigma_k^* j\mathtt{10}^*)_k \quad \text{for } j < 2 && (23)\\
Y^\rho_{j,j',q} &= W^\rho_{j',q} + (\mathtt{1}(j-2)\mathtt{10}^*)_k \;\cap\; (\mathtt{1}\Sigma_k^* j\mathtt{10}^*)_k \quad \text{for } 2 \leqslant j < k-1 && (24)\\
Y^\rho_{k-1,j',q} &= W^\rho_{j',q} + ((k-3)\mathtt{10}^*)_k \;\cap\; (\mathtt{1}\Sigma_k^*(k-1)\mathtt{10}^*)_k && (25)\\
Z^\rho_{j,q} &= \bigcup_{j'} Y^\rho_{j,j',q} && (26)
\end{aligned}$$

As the new equations represent the subexpressions of $\rho_j(X_q)$, the values of the variables X_q in the least solution of the new system are the same as in the least solution of the old system.

These variables are grouped as follows:

$$\begin{gathered}
\{\, U^\rho_{j',q} \mid j' \in \Sigma_k, q \in Q\}, \quad \{\, W^\rho_{j',q} \mid j' \in \Sigma_k, q \in Q\},\\
\{\, Y^\rho_{j,j',q} \mid j, j' \in \Sigma_k, q \in Q\}, \quad \{\, Z^\rho_{j,q} \mid j \in \Sigma_k, q \in Q\}.
\end{gathered}$$

As in the case of λ, the values of the variables in each group are pairwise disjoint, and the union of each group is a set with a regular notation.

Claim 8. *For all* $(j_1', q_1) \neq (j_2', q_2)$*, the sets* $\mu(U^\rho_{j_1',q_1})$ *and* $\mu(U^\rho_{j_2',q_2})$ *are disjoint, and their union is*

$$\bigcup_{j',q} \mu(U^\rho_{j',q}) = (\mathtt{1}(\Sigma_k^+ \setminus (k-1)^*)\mathtt{10}^*)_k.$$

Proof. It was proved [6, Eq. (8)] that

$$\mu(U^\rho_{j',q}) = \{\, (\mathtt{1}wj'\mathtt{10}^\ell)_k \mid \ell \geqslant 0, wj' \notin (k-1)^*, wj' \boxplus \mathtt{1} \in L_M(q)\}.$$

These sets can be obtained from the languages $K^\rho_{j,q} = (L_M(q) \boxminus \mathtt{1}) \cap \Sigma_k^* j$ as in Claim 2 with $x = \mathtt{1}$ and $y = \mathtt{1}$. The languages $K^\rho_{j_1,q_1}$ and $K^\rho_{j_2,q_2}$ are disjoint for all $(j_1, q_1) \neq (j_2, q_2)$, as for $j_1 \neq j_2$ their last digits are different, while for $q_1 \neq q_2$ it holds that $K^\rho_{j_1,q_1} \subseteq$

$L_M(q_1) \boxminus \mathtt{1}$ and $K^{\rho}_{j_2,q_2} \subseteq L_M(q_2) \boxminus \mathtt{1}$, and the supersets are disjoint. Then, by Claim 2, $\mu(U^{\rho}_{j_1,q_1}) \cap \mu(U^{\rho}_{j_2,q_2}) = \varnothing$ for $(j_1, q_1) \neq (j_2, q_2)$, while the union of these sets is

$$\begin{aligned}\bigcup_{j\in\Sigma_k, q\in Q} \mu(U^{\rho}_{j,q}) &= (\mathtt{1}(\textstyle\bigcup_{j\in\Sigma_k, q\in Q} K^{\rho}_{j,q})\mathtt{10}^*)_k = (\mathtt{1}(\textstyle\bigcup_{j\in\Sigma_k, q\in Q}(L_M(q) \boxminus \mathtt{1}) \cap \Sigma_k^* j)\mathtt{10}^*)_k = \\ &= (\mathtt{1}(\textstyle\bigcup_{q\in Q}(L_M(q) \boxminus \mathtt{1}) \cap \Sigma_k^+)\mathtt{10}^*)_k = (\mathtt{1}((\Sigma_k^* \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{10}^*)_k = (\mathtt{1}(\Sigma_k^+ \setminus (k-1)^*)\mathtt{10}^*)_k,\end{aligned}$$

and the claim follows. $\square$

Claim 9. *For $(j'_1, q_1) \neq (j'_2, q_2)$, the sets $\mu(W^{\rho}_{j'_1,q_1})$ abd $\mu(W^{\rho}_{j'_2,q_2})$ are disjoint, and*

$$\bigcup_{j',q} \mu(W^{\rho}_{j',q}) = (\mathtt{1}(\Sigma_k^+ \setminus (k-1)^*)\mathtt{20}^*)_k.$$

Proof. It was proved [6, Eq. (9)] that

$$\mu(W^{\rho}_{j',q}) = \{\, (\mathtt{1}wj'\mathtt{20}^{\ell})_k \mid wj' \boxplus \mathtt{1} \in L_M(q),\ wj' \notin (k-1)^*,\ \ell \geqslant 0\}.$$

These sets are induced by the languages $K^{\rho}_{j,q} = (L_M(q) \boxminus \mathtt{1}) \cap \Sigma_k^* j$ as in Claim 2 with $x = \mathtt{1}$ and $y = \mathtt{2}$. These languages appeared already in Claim 8, where it was shown that they are pairwise disjoint and their union is $\Sigma_k^+ \setminus (k-1)^*$. Then, by Claim 2, for all $(j'_1, q'_1) \neq (j'_2, q'_2)$, the sets $\mu(W^{\rho}_{j'_1,q_1})$ and $\mu(W^{\rho}_{j'_2,q_2})$ are disjoint, and

$$\bigcup_{j',q} \mu(W^{\rho}_{j',q}) = (\mathtt{1}(\textstyle\bigcup_{j',q} K^{\rho}_{j',q})\mathtt{20}^*)_k = (\mathtt{1}(\Sigma_k^+ \setminus (k-1)^*)\mathtt{20}^*)_k,$$

which completes the proof. $\square$

Claim 10. *The sets $\mu(Y^{\rho}_{j_1,j'_1,q_1})$ and $\mu(Y^{\rho}_{j_2,j'_2,q_2})$ are disjoint for all $(j_1, j'_1, q_1) \neq (j_2, j'_2, q_2)$, and the union in the group equals*

$$\bigcup_{j,j',q} \mu(Y^{\rho}_{j,j',q}) = (\mathtt{1}((\Sigma_k^* \setminus \mathtt{0}^*)\Sigma_k \boxminus \mathtt{1})\mathtt{10}^*)_k.$$

Proof. It is known [6, Eqs. (10, 11, 12)] that

$$\mu(Y^{\rho}_{j,j',q}) = \{\, (\mathtt{1}(w'j' \boxplus \mathtt{1})j\mathtt{10}^{\ell-1})_k \mid \ell \geqslant 1,\ w'j' \notin (k-1)^*,\ w'j' \boxplus \mathtt{1} \in L_M(q)\}$$

for all $j \neq k-1$, and

$$\mu(Y^{\rho}_{k-1,j',q}) = \{\, (\mathtt{1}w'j'(k-1)\mathtt{10}^{\ell-1})_k \mid \ell \geqslant 1,\ w'j' \notin (k-1)^*,\ w'j' \boxplus \mathtt{1} \in L_M(q)\}.$$

Fix any $j \neq k-1$. Then the sets $\mu(Y^{\rho}_{j,j',q})$ are obtained from the languages $K^{\rho}_{j,j',q} = (L_M(q) \setminus \mathtt{0}^*) \cap (\Sigma_k^* j' \boxplus \mathtt{1})$ as in Claim 2 with $x = \mathtt{1}$ and $y = j\mathtt{1}$. Then, for all $(j'_1, q_1) \neq (j'_2, q_2)$, the languages K^{ρ}_{j,j'_1,q_1} and K^{ρ}_{j,j'_2,q_2} are disjoint, as for $q_1 \neq q_2$ $K^{\rho}_{j,j'_1,q_1} \subseteq L_M(q_1)$ and $K^{\rho}_{j,j'_1,q_1} \subseteq L_M(q_2)$, and the supersets are disjoint. If $j'_1 \neq j'_2$, then the strings from these languages differ in the last digit. Therefore, by Claim 2,

$$\begin{aligned}\bigcup_{j',q} \mu(Y^{\rho}_{j,j',q}) &= (\mathtt{1}(\textstyle\bigcup_{j',q} K^{\rho}_{j,j',q})j\mathtt{10}^*)_k = (\mathtt{1}(\textstyle\bigcup_{j',q}(L_M(q) \setminus \mathtt{0}^*) \cap (\Sigma_k^* j' \boxplus \mathtt{1}))j\mathtt{10}^*)_k = \\ &= (\mathtt{1}(\textstyle\bigcup_{j'}(\Sigma_k^+ \setminus \mathtt{0}^*) \cap (\Sigma_k^* j' \boxplus \mathtt{1}))j\mathtt{10}^*)_k = (\mathtt{1}((\Sigma_k^+ \setminus \mathtt{0}^*) \cap (\Sigma_k^+ \boxplus \mathtt{1}))j\mathtt{10}^*)_k = \\ &= (\mathtt{1}(\Sigma_k^+ \setminus \mathtt{0}^*)j\mathtt{10}^*)_k = (\mathtt{1}((\Sigma_k^+ \setminus \mathtt{0}^*)(j+1) \boxminus \mathtt{1})\mathtt{10}^*)_k,\end{aligned}$$

and $\mu(Y^\rho_{j,j'_1,q_1}) \cap \mu(Y^\rho_{j,j'_2,q_2}) = \varnothing$ for all $(j'_1, q_1) \neq (j'_2, q_2)$.

Next, consider the case of $j = k-1$ and recall the languages $K^\rho_{j',q} = (L_M(q) \boxminus \mathtt{1}) \cap \Sigma_k^* j$ introduced in Claim 8, where it was shown that these languages are pairwise disjoint and their union is

$$\bigcup_{j',q} K^\rho_{j',q} = \Sigma_k^+ \setminus (k-1)^*.$$

Now the sets $\mu(Y^\rho_{k-1,j'_1,q_1})$ can be obtained from the languages $K^\rho_{j',q}$ by the method of Claim 2 with $x = \mathtt{1}$ and $y = (k-1)\mathtt{1}$. Therefore,

$$\bigcup_{j',q} \mu(Y^\rho_{k-1,j'_1,q_1}) = (\mathtt{1}(\Sigma_k^+ \setminus (k-1)^*)(k-1)\mathtt{10}^*)_k = (\mathtt{1}((\Sigma_k^+ \setminus \mathtt{0}^*)\mathtt{0} \boxminus \mathtt{1})\mathtt{10}^*)_k,$$

where the the first equality comes from Claim 2 and the second one is a simple calculation. Also, for different $(j'_1, q_1) \neq (j'_2, q_2)$, the sets $\mu(Y^\rho_{k-1,j'_1,q_1})$ and $\mu(Y^\rho_{k-1,j'_1,q_1})$ are disjoint.

Finally, in order to prove the claim, consider any two sets $\mu(Y^\rho_{j_1,j'_1,q_1})$ and $\mu(Y^\rho_{j_2,j'_2,q_2})$ with $(j_1, j'_1, q_1) \neq (j_2, j'_2, q_2)$. If $j_1 \neq j_2$, then these sets are disjoint, as their elements differ in the second from the last non-zero digit. If $j_1 = j_2$ and $(j'_1, q_1) \neq (j'_2, q_2)$, then these two sets have been proved to be disjoint in one of the cases above.

The union of all these sets is

$$\bigcup_{j,j',q} \mu(Y^\rho_{j,j',q}) = \bigcup_{j \neq k-1} \bigcup_{j',q} \mu(Y^\rho_{j,j',q}) \cup \bigcup_{j',q} \mu(Y^\rho_{k-1,j',q}) =$$
$$\bigcup_{j \neq k-1} (\mathtt{1}((\Sigma_k^+ \setminus \mathtt{0}^*)(j+1) \boxminus \mathtt{1})\mathtt{10}^*)_k \cup (\mathtt{1}((\Sigma_k^+ \setminus \mathtt{0}^*)\mathtt{0} \boxminus \mathtt{1})\mathtt{10}^*)_k = (\mathtt{1}((\Sigma_k^+ \setminus \mathtt{0}^*)\Sigma_k \boxminus \mathtt{1})\mathtt{10}^*)_k,$$

which establishes the claim. □

Claim 11. *The sets* $\mu(Z^\rho_{j_1,q_1})$ *and* $\mu(Z^\rho_{j_2,q_2})$ *are disjoint for* $(j_1, q_1) \neq (j_2, q_2)$. *Their union equals*

$$\bigcup_{j,q} \mu(Z^\rho_{j,q}) = (\mathtt{1}((\Sigma_k^* \setminus \mathtt{0}^*)\Sigma_k \boxminus \mathtt{1})\mathtt{10}^*)_k.$$

Proof. The variable $Z^\rho_{j,q}$ is defined by the equation (26) as the union of $Y^\rho_{j,j',q}$ for all j'. Then

$$\bigcup_{j,q} \mu(Z^\rho_{j,q}) = \bigcup_{j,q} \bigcup_{j'} \mu(Y^\rho_{j,j',q}) = \bigcup_{j,j',q} \mu(Y^\rho_{j,j',q}) = (\mathtt{1}((\Sigma_k^* \setminus \mathtt{0}^*)\Sigma_k \boxminus \mathtt{1})\mathtt{10}^*)_k,$$

where the second equality is given by Claim 10. The latter claim also states that the sets $\mu(Y^\rho_{j,j',q})$ are pairwise disjoint, and hence so are the sets $\mu(Z^\rho_{j,q})$. □

Thus the expressions $\lambda_i(X_q)$ and $\rho_j(X_q)$ have been expressed by equations of the form satisfying the assumptions of Lemma 6 and Lemma 7. It remains to transform the equation defining X_q to the same form. The original equation [6] was

$$X_q = R_q \cup \bigcup_{\substack{q',q'' \colon \delta(q',q'')=q \\ i,j \in \Sigma_k}} \lambda_i(X_{q''}) \cap \rho_j(X_{q'}),$$

The subexpression corresponding to every i, q'', j and q' shall be represented by a new variable $X_{i,q'',j,q'}$ with the equation

$$X_{i,q'',j,q'} = Z^{\lambda}_{i,q''} \cap Z^{\rho}_{j,q'}, \tag{27}$$

while the equation for X_q is accordingly replaced by

$$X_q = R_q \cup \bigcup_{\substack{q',q'' \colon \delta(q',q'')=q \\ i,j \in \Sigma_k}} X_{i,q'',j,q'} \tag{28}$$

The variables are divided into two groups,

$$\{\, X_{i,q'',j,q'} \mid i, j \in \Sigma_k, q', q'' \in Q\}, \quad \{\, X_q \mid q \in Q\}$$

and it remains to show the required properties of the variables in each group.

Claim 12. *For all* $(i_1, q''_1, j_1, q'_1) \neq (i_2, q''_2, j_2, q'_2)$*, the sets* $\mu(X_{i_1,q''_1,j_1,q'_1})$ *and* $\mu(X_{i_2,q''_2,j_1,q'_2})$ *are disjoint, and the union of all these sets is*

$$\bigcup_{i,q'',j,q'} \mu(X_{i,q'',j,q'}) = (\mathtt{1}((\Sigma_k^* \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k \setminus \textstyle\bigcup_q R_q.$$

Proof. According to the equation (27), $\mu(X_{i,q'',j,q'}) = \mu(Z^{\lambda}_{i,q''}) \cap \mu(Z^{\rho}_{j,q'})$. By Claim 7, $\mu(Z^{\lambda}_{i_1,q''_1}) \cap \mu(Z^{\lambda}_{i_2,q''_2}) = \varnothing$ for $(i_1, q''_1) \neq (i_2, q''_2)$. Similarly, by Claim 11, $\mu(Z^{\rho}_{j_1,q'_1}) \cap \mu(Z^{\rho}_{j_2,q'_2}) = \varnothing$ for $(j_1, q'_1) \neq (j_2, q'_2)$. Thus for $(i_1, q''_1, j_1, q'_1) \neq (i_2, q''_2, j_2, q'_2)$ it holds that $\mu(X_{i_1,q''_1,j_1,q'_1}) \cap \mu(X_{i_2,q''_2,j_2,q'_2}) = \varnothing$.

By the equation (27), the union of all these sets is

$$\bigcup_{i,q'',j,q'} \mu(X_{i,q'',j,q'}) = \bigcup_{i,q'',j,q'} \mu(Z^{\lambda}_{i,q''}) \cap \mu(Z^{\rho}_{j,q'}) = \Big(\bigcup_{i,q''} \mu(Z^{\lambda}_{i,q''})\Big) \cap \Big(\bigcup_{j,q'} \mu(Z^{\rho}_{j,q'})\Big),$$

and using the values of both unions given by Claim 7 and Claim 11, this can be calculated as follows:

$$\begin{aligned}
&(\mathtt{1}\Sigma_k(\Sigma_k^+ \setminus (k-1)^*)\mathtt{1}\mathtt{0}^*)_k \cap (\mathtt{1}((\Sigma_k^+ \setminus \mathtt{0}^*)\Sigma_k \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k = \\
&\qquad = (\mathtt{1}(\Sigma_k(\Sigma_k^+ \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k \cap (\mathtt{1}((\Sigma_k^+ \setminus \mathtt{0}^*)\Sigma_k \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k = \\
&\qquad = ((\mathtt{1}((\Sigma_k^+ \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*))_k \setminus (((\mathtt{1}(\Sigma_k\mathtt{0}^* \cup \mathtt{0}^*\Sigma_k) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*))_k = \\
&\qquad\qquad = ((\mathtt{1}((\Sigma_k^+ \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*))_k \setminus \textstyle\bigcup_{q \in Q} R_q,
\end{aligned}$$

which concludes the proof. □

Since the new equations represent the subexpressions of the original system, the value of the least solution of the common variables (i.e., X_q) remains the same, that is $\mu(X_q) = S_q$. Moreover, Claim 3 asserts that the sets S_q are pairwise disjoint and that their union is a set with a regular notation. Thus the only thing remaining to be checked is that there are no cyclic chain dependencies in the defined system.

Claim 13. *There are no cyclic chain dependencies in the equations* (16)–(28)*.*

Proof. The constructed system contains the following chain dependencies:

- there may be a chain dependency of $U^\lambda_{i',q}$ from X_q or $U^\rho_{j',q}$ from X_q
- of X_q from (some) $X_{i,q'',j,q'}$
- of $X_{i,q'',j,q'}$ from $Z^\lambda_{i,q''}$ and from $Z^\rho_{i,q'}$
- of $Z^\lambda_{i,q}$ from $Y^\lambda_{i,i',q}$
- of $Z^\rho_{j,q}$ from $Y^\rho_{j,j',q}$

Consider the following groups of variables:

$$\begin{aligned}
\mathcal{G}_1 &= \{\, X_q \mid q \in Q\} \\
\mathcal{G}_2 &= \{\, U^\lambda_{i',q}, U^\rho_{j',q} \mid q \in Q;\ i', j'' \in \Sigma_k\} \\
\mathcal{G}_3 &= \{\, Z^\lambda_{i,q}, Z^\rho_{j,q} \mid q \in Q;\ i, j \in \Sigma_k\} \\
\mathcal{G}_4 &= \{\, X_{i,q'',j,q'} \mid q', q'' \in Q;\ i, j \in \Sigma_k\} \\
\mathcal{G}_5 &= \{\, Y^\lambda_{i,i',q}, Y^\rho_{j,j',q} \mid q \in Q;\ i, j, i', j' \in \Sigma_k\}
\end{aligned}$$

Then it can be easily seen that if a variable from a group $\mathcal{G}_m$ depends on a variable in a group $\mathcal{G}_n$, then $m < n$. Therefore, there are no chain dependencies in the system. $\square$

According to the above claims, there exists a resolved system of equations satisfying the assumption of Lemma 6 and Lemma 7, such that one of the components in its least solution is

$$(\mathtt{1}(L_M(q) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k = \{\, (\mathtt{1}w\mathtt{1}\mathtt{0}^\ell)_k \mid \ell \geqslant 0, w \notin (k-1)^*, w \boxplus \mathtt{1} \in L_M(q)\}.$$

Then, by the aforementioned lemmata, there exist unresolved systems either with union and sum, or with intersection and sum, which have the same unique solution. Finally, using Theorem 6, regular constants used in these systems are replaced by singleton constants, which completes the proof of Lemma 12. $\square$

The next task is to represent the set $(\mathtt{1}L_M(q))_k$ for any trellis automaton M and its state q. Similarly to Lemma 12, this will be done by transforming an existing construction [6, Lem.6].

Lemma 13. *For every $k \geqslant 4$ and for every trellis automaton M over Σ_k there exists and can be effectively constructed an unresolved system of equations over sets of numbers using the operations of union (or intersection) and addition, as well as singleton constants, such that its unique solution contains a component $(\mathtt{1}L_M(q))_k$ for each state q of this automaton.*

Proof. The argument will use a simple technical claim, similar to Claim 2 in the proof of Lemma 12.

Claim 14. *Let $x \in \Sigma_k^* \setminus \mathtt{0}\Sigma_k^*$ and $y \in \Sigma_k^*$ be a string of digits (possibly empty), let $K_1, \ldots, K_m \subseteq \Sigma_k^+$ be any pairwise disjoint languages, and let $S_1, \ldots, S_m$ be sets of numbers defined by*

$$S_t = \{\, (xuy)_k \mid u \in K_t\}.$$

Then these sets are pairwise disjoint and their union is

$$\bigcup_{t=1}^m S_t = (x \textstyle\bigcup_{t=1}^m K_t y)_k.$$

The proof is nearly obvious and is omitted. A stronger statement will be proved in the following as Claim 17.

Consider the trellis automaton M over Σ_k. For every state q and for every digit $j \in \Sigma_k$, construct a trellis automaton $M_{q,j}$ with the set of states $Q_{q,j}$ recognizing the language $L_M(q)\{j\}^{-1}$ using the known transformation [14]. Then, by Lemma 12, there is a system of equations using addition and either union or intersection, which contains a variable $Y_{q,j,p}$ for each state p of $M_{q,j}$, and has a unique solution with $Y_{q,j,p} = (\mathtt{1}((L_{M_{q,j}}(p) \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k$.

The first goal is to combine these systems into a larger system of equations containing variables $Y_{q,j}$ for each state q of M and for each digit j, so that it has $Y_{q,j} = (\mathtt{1}((L(M_{q,j}) \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k$ in its unique solution.

When union and addition are allowed, the construction is immediate: if $F_{q,j}$ is the set of accepting states of $M_{q,j}$, then

$$Y_{q,j} = \bigcup_{p \in F_{q,j}} Y_{q,j,p} \tag{29}$$

merged with subsystems defining $Y_{q,j,p}$ satisfies the goal.

If the allowed operations are intersection and addition, then the following system is constructed:

$$\begin{aligned} Y_{q,j} \cap Y_{q,j,p} &= \varnothing && \text{for } p \notin F_{q,j} && (30) \\ Y_{q,j} \cap Y_{q,j,p} &= Y_{q,j,p} && \text{for } p \in F_{q,j} && (31) \\ Y_{q,j} \cap \big[\mathbb{N} \setminus (\mathtt{1}((\Sigma_k^* \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*))_k\big] &= \varnothing, && && (32) \end{aligned}$$

where the variables $Y_{q,j,p}$ are defined in subsystems. As the sets $\{(\mathtt{1}((L_{M_{q,j}}(p) \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k\}_{p \in Q_{q,j}}$ together with $\mathbb{N} \setminus (\mathtt{1}((\Sigma_k^+ \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*))_k$ form a partition of natural numbers, these equations effectively represent the union of $Y_{q,j,p}$ for all p. The additional constant $(\mathtt{1}((\Sigma_k^* \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*))_k$ used in the construction is a set of numbers with a regular base-k positional notation, and hence it can be expressed by Theorem 6.

The sets $(\mathtt{1}(((L_M(q)\{j\}^{-1}) \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k$ are used in a known construction [6, Lem.6] of an equation representing the set $(\mathtt{1} \cdot L_M(q))_k$. This equation is of the form

$$Z_q = C_q \cup \bigcup_{j=0}^{k-1} (Y_{q,j} \cap (\mathtt{1}\Sigma_k^*\mathtt{1})_k) + (\mathtt{1}j \boxminus \mathtt{1})_k,$$

which uses the constant $C_q = (\mathtt{1}L_M(q))_k \cap (\mathtt{1}\mathtt{0}^*\Sigma_k)_k$ with a regular base-k notation. These constants are similar to the constants R_q in Lemma 12, in the sense that they represent strings of digits of a simple form not handled by the main formula. This equation also refers to variables $Y_{q,j}$ defined in their own subsystems, so that their least solution satisfies

$$\mu(Y_{q,j}) = (\mathtt{1}(((L_M(q)\{j\}^{-1}) \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k.$$

It is already known [6, Eq. (15)] that this equation, together with the aforementioned subsystems for variables $Y_{q,j}$, has a least solution with

$$\mu(Z_q) = (\mathtt{1} \cdot L_M(q))_k.$$

Then, by Proposition 2, the equation for Z_q with variables $Y_{q,j}$ replaced by constants $Y_{q,j} = (\mathtt{1}(((L_M(q)\{j\}^{-1}) \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k$ has the least solution with $\mu(Z_q) = (\mathtt{1} \cdot L_M(q))_k$.

The equations for Z_q for all $q \in Q$ can be turned into a system satisfying the assumption of Lemma 6 and Lemma 7 by introducing new variables $Z_{q,j}$ and rewriting the equations as:

$$\begin{aligned} Z_{q,j} &= Y_{q,j} \cap (\mathtt{1}\Sigma_k^*\mathtt{1})_k \\ Z_q &= C_q \cup \bigcup_{j=0}^{k-1} Z_{q,j} + (\mathtt{1}j \boxminus \mathtt{1})_k \end{aligned}$$

The grouping of variables required by Lemmata 6 and 7 is

$$\{\, Z_q \mid q \in Q\}, \quad \{\, Z_{q,j} \mid q \in Q\}_{j \in \Sigma_k}.$$

It has to be proved that the sets in each group form a disjoint partition of a certain set with a regular notation.

Claim 15. *For every $j \in \Sigma_k$ and $q_1 \neq q_2$, the sets $\mu(Z_{q_1,j})$ and $\mu(Z_{q_2,j})$ are disjoint and*

$$\bigcup_q \mu(Z_{q,j}) = (\mathtt{1}(\Sigma_k^* \setminus (k-1)^*)\mathtt{1})_k$$

Proof. The value of $Z_{q,j}$ is determined from its equation as follows:

$$\begin{aligned} \mu(Z_{q,j}) = \mu(Y_{q,j}) \cap (\mathtt{1}\Sigma_k^*\mathtt{1})_k = (\mathtt{1}(((L_M(q)\{j\}^{-1}) \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^*)_k \cap (\mathtt{1}\Sigma_k^*\mathtt{1})_k = \\ (\mathtt{1}(((L_M(q)\{j\}^{-1}) \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1}\mathtt{0}^* \cap \mathtt{1}\Sigma_k^*\mathtt{1})_k = (\mathtt{1}(((L_M(q)\{j\}^{-1}) \setminus \mathtt{0}^*) \boxminus \mathtt{1})\mathtt{1})_k \end{aligned}$$

Fix any digit j. The sets $\mu(Z_{q,j})$ satisfy the assumption of Claim 14 with $K_q = ((L_M(q)\{j\}^{-1}) \setminus \mathtt{0}^*) \boxminus \mathtt{1}$ and $x = y = \mathtt{1}$. For $q_1 \neq q_2$ the languages $L_M(q_1)$ and $L_M(q_2)$ are disjoint, and hence the sets also $K_{q_1} = ((L_M(q_1)\{j\}^{-1}) \setminus \mathtt{0}^*) \boxminus \mathtt{1}$ and $K_{q_2} = ((L_M(q_2)\{j\}^{-1}) \setminus \mathtt{0}^*) \boxminus \mathtt{1}$ are disjoint as well. Thus $\mu(Z_{q_1,j}) \cap \mu(Z_{q_2,j}) = \varnothing$ by Claim 14. Also

$$\bigcup_q \mu(Z_{q,j}) = (\mathtt{1}(\textstyle\bigcup_q K_q)\mathtt{1})_k = (\mathtt{1}\Sigma_k^+ \setminus (k-1)^*\mathtt{1})_k,$$

since every string not in $(k-1)^*$ belongs to some K_q. $\square$

Claim 16. *For all $q_1 \neq q_2$, the sets $\mu(Z_{q_1})$ and $\mu(Z_{q_2})$ are disjoint and*

$$\bigcup_q \mu(Z_q) = (\mathtt{1}\Sigma_k^+)_k.$$

Proof. By Proposition 3, $\mu(Z_q)$ remains the same as in the original system, hence $\mu(Z_q) = (\mathtt{1}L_M(q))_k$. Thus $\mu(Z_q)$ satisfy the assumption of Claim 14 with $K'_q = L_M(q)$, $x = \mathtt{1}$ and $y = \varepsilon$. Clearly, the languages $\{K'_q\}$ are pairwise disjoint, as trellis automata are deterministic. Also each non-empty string belongs to some K'_q, hence $\bigcup_{q \in Q} K'_q = \Sigma_k^+$. Therefore, by Claim 14, $\mu(Z_{q_1}) \cap \mu(Z_{q_2}) = \varnothing$ for $q_1 \neq q_2$ and

$$\bigcup_q \mu(Z_q) = (\mathtt{1}(\textstyle\bigcup_{q \in Q} K'_q))_k = (\mathtt{1}\Sigma_k^+)_k.$$

as claimed. $\square$

Lemma 6 and Lemma 7 require that there are no chain cyclic dependencies in the constructed system. As the only chain dependencies are those of Z_q from (some) $Z_{q,j}$, there are no cycles among them.

Therefore, the new system satisfies the assumption of the Lemma 6 and Lemma 7, and accordingly, there exists an unresolved system using addition and either union or intersection, which has a unique solution with $(\mathtt{1} \cdot L_M(q))_k$ as one of its components. The system uses regular constants, which can be eliminated using Theorem 6, and constants $Y_{q,j}$, which are represented in (29) in the case of union and addition, and in (30–32) using intersection and addition. □

The final step of the known construction [6] was to specify the set $(L(M))_k$ with minimal assumptions on the language $L(M)$. This step will now be similarly replicated using unresolved systems.

Lemma 14. *For every $k \geqslant 4$ and for every trellis automaton M over Σ_k, such that $L(M) \cap \mathtt{0}\Sigma_k^* = \varnothing$, there exists and can be effectively constructed an unresolved system of equations over sets of numbers using the operations of union (or intersection) and addition, as well as singleton constants, such that its unique solution contains a component $(L(M))_k$.*

Proof. The following slightly more complicated version of Claim 14 will be used in the proof:

Claim 17. *Let $x \in \Sigma_k^+ \setminus \mathtt{0}\Sigma_k^*$ and $y, z \in \Sigma_k^*$ be strings of digits (possibly empty), let $K_1, \ldots, K_m \subseteq \Sigma_k^+$ be any pairwise disjoint languages, and let $S_1, \ldots, S_m$ be sets of numbers defined by*

$$S_t = \{\, (x(z^{-1}u)y)_k \mid u \in K_t\}.$$

Then these sets are pairwise disjoint and their union is

$$\bigcup_{t=1}^{m} S_t = (x(z^{-1}(\textstyle\bigcup_{t=1}^{m} K_t))y)_k.$$

Proof. Let S_t and $S_{t'}$ be any two sets with $t \neq t'$ and suppose there is a number n belonging to both of them. Then $n = (x(z^{-1}u)y)_k$ for some $u \in K_t$ and $n = (x(z^{-1}u')y)_k$ with $u' \in K_{t'}$. Clearly, z is a prefix of both u and u', that is, $u = zv$ and $u' = zv'$. Then $n = (x(z^{-1})uy)_k = (xvy)_k$ and $n = (x(z^{-1})u'y)_k = (xv'y)_k$, and therefore $v' = v$ and $u = u'$. It is a contradiction, as K_t and $K_{t'}$ are disjoint. This proves that $S_t \cap S_{t'} = \varnothing$.

The union of these sets is

$$\bigcup_t S_t = \bigcup_t (x(z^{-1}K_t)y)_k = (x(z^{-1} \textstyle\bigcup_t K_t)y)_k,$$

as desired. □

The proof of Lemma 14 begins with the following system of equations [6, Eqs. (16,19)]:

$$T_q = (L_M(q) \cap \Sigma_k)_k \cup Z_{1,p} \cup \bigcup_{i \in \Sigma_k \setminus \{\mathtt{0},\mathtt{1}\}} \tau_i(Z_{i,q}), \quad \text{where}$$

$$\tau_i(X) = \bigcup_{i' \in \Sigma_k} \Big((X \cap (\mathtt{1}i'\Sigma_k^*)_k) + ((i-1)\mathtt{0}^*)_k \,\cap\, (ii'\Sigma_k^*)_k \Big) \quad (\text{for } i \neq \mathtt{0}, \mathtt{1}).$$

The system refers to the variables $Z_{i,p}$; the values of these variables are defined in their own subsystems with the solution $\mu(Z_{i,q}) = (\mathtt{1}\{i\}^{-1}L_M(q))_k$.

It is known [6, Eqs. (16,19)], that

$$\tau_i(\{n\}) = \begin{cases} \{(iw)_k\}, & \text{if } n = (\mathtt{1}w)_k \ w \in \Sigma_k^+, \\ \varnothing, & \text{otherwise,} \end{cases}$$

and that the system of equations formed by the above equation for T_q and the subsystems for all variables $Z_{i,p}$ has a least solution with

$$\mu(T_q) = (L_M(q) \setminus \mathtt{0}^*)_k.$$

Consider the following decomposition of this equation:

$$\begin{aligned} U_{i,i',q} &= Z_{i,q} \cap (\mathtt{1}i'\Sigma_k^*)_k && \text{for } i \geqslant 2 \\ W_{i,i',q} &= U_{i,i',q} + ((i-1)\mathtt{0}^*)_k \cap (ii'\Sigma_k^*)_k && \text{for } i \geqslant 2 \\ T_q &= \bigcup_{i \geqslant 2, i'} W_{i,i',q} \cup Z_{1,q} \cup (L_M(q) \cap \Sigma_k)_k \end{aligned}$$

Let the set of variables be split into the following $2k-3$ groups:

$$\{\, U_{i,i',q} \mid i' \in \Sigma_k, q \in Q\}_{2 \leqslant i < k}, \quad \{\, W_{i,i',q} \mid i, i' \in \Sigma_k, q \in Q\}_{2 \leqslant i < k}, \quad \{\, T_q \mid q \in Q\}.$$

As in the previous proofs, it is claimed that the union of each group is a set with a regular base-k notation, and that the sets in each group are pairwise disjoint.

It it known from the previous work [6, Eq. (16)] that

$$\mu(U_{i,i',q}) = \{\, (\mathtt{1}i'w)_k \mid ii'w \in L_M(q)\}.$$

Fix $i \geqslant 2$. Then the sets $\{\mu(U_{i,i',q})\}$ for all $i' \in \Sigma_k$ and q satisfy the assumption of Claim 17 with $K_{i',q} = ii'\Sigma_k^* \cap L_M(q)$, $x = \mathtt{1}$, $y = \varepsilon$ and $z = i$. The intersection $K_{i'_1,q_1} \cap K_{i'_2,q_2}$ is empty, as for $i'_1 \neq i'_2$ it holds that $\mathtt{1}i'_1\Sigma_k^* \cap \mathtt{1}i'_2\Sigma_k^* = \varnothing$, and $L_M(q_1) \cap L_M(q_2) = \varnothing$ for $q_1 \neq q_2$, because trellis automata is deterministic. Hence,

$$\bigcup_{i',q} \mu(U_{i,i',q}) = (\mathtt{1}(i^{-1}\bigcup_{i',q} K_{i',q}))_k = (\mathtt{1}(i^{-1}i\Sigma_k^+))_k = (\mathtt{1}\Sigma_k^+)_k \quad \text{for each } i \geqslant 2.$$

It is also known [6, Eq. (17)] that

$$\mu(W_{i,i',q}) = \{\, (ii'w)_k \mid ii'w \in L_M(q)\} = (L_M(q))_k \cap (ii'\Sigma_k^*)_k.$$

Consider any two variables W_{i_1,i'_1,q_1} and W_{i_2,i'_2,q_2} with $(i_1, i'_1, q_1) \neq (i_2, i'_2, q_2)$. If $q_1 \neq q_2$, then $L_M(q_1) \cap L_M(q_2) = \varnothing$. If $(i_1, i'_1) \neq (i_2, i'_2)$ then $i_1i'_1\Sigma_k^* \cap i_2i'_2\Sigma_k^* = \varnothing$. In both cases $\mu(W_{i_1,i'_1,q_1}) \cap \mu(W_{i_2,i'_2,q_2}) = \varnothing$. The union of these sets equals:

$$\bigcup_{i,i',q} \mu(W_{i,i',q}) = \bigcup_{i,i',q} (L_M(q))_k \cap (ii'\Sigma_k^*)_k = (\Sigma_k^{\geqslant 2})_k.$$

The proof any sets T_{q_1} and T_{q_2} with $q_1 \neq q_2$ are disjoint is immediate, as the new equations represent subexpressions of the former system, hence $\mu(T_q) = (L_M(q) \setminus \mathtt{0}^*)_k$. Thus for all $q_1 \neq q_2$

$$\mu(T_{q_1}) \cap \mu(T_{q_2}) \subseteq (L_M(q_1))_k \cap (L_M(q_2))_k = \varnothing,$$

while the union of all these sets is

$$\bigcup_q \mu(T_q) = \bigcup_q (L_M(q) \setminus \mathtt{0}^*)_k = (\Sigma_k^+ \setminus \mathtt{0}^*)_k.$$

The only chain dependency in the constructed system is that of T_q from $W_{i,i',q}$. Hence there are no cyclic chain dependencies and the system obtained satisfies the assumptions of Lemma 6 and Lemma 7 with constants $Z_{i,q}$ and regular constants.

Hence there exists an unresolved system of the required form with one of the components of its unique solution equal to $(L(M))_k$. This system uses constants $Z_{i,q}$ and regular constants. The former are expressed using Lemma 13 and the latter by Theorem 6. $\square$

5.4 Universality

Using the above Lemma 12 instead of Theorem 3 immediately proves Lemma 2 in its full form stated above, that the sets $\mathrm{VALC}_i(T)$ are representable by equations using, along with addition, *either* union *or* intersection. The final step of the argument is to modify the systems defined in the proofs of Lemmata 3 and 4 to use these sets of operations.

The only equations using Boolean operations in those proofs are (4) and (8), and since they are identical, it is sufficient to rephrase a single equation (4). Its reformulation using addition and intersection is immediate:

Lemma 15. *Let* $Y_i \subseteq (\mathtt{1}\Sigma_6^+)_6$ *for* $1 \leqslant i \leqslant 5$ *and let* $Y_0 \subseteq \{0,1,2,3,4,5\}$*. Then, for every set* $Y \subseteq \mathbb{N}$*,*

$$Y = Y_0 \cup Y_1 \cup \bigcup_{\substack{i \in \{\mathtt{2},\mathtt{3},\mathtt{4},\mathtt{5}\} \\ j \in \Sigma_6}} \Big((Y_i \cap (\mathtt{1}j\Sigma_6^*)_6) + ((i-1)\mathtt{0}^*)_6 \ \cap \ (ij\Sigma_6^*)_6 \Big) \tag{33}$$

if and only if

$$\begin{aligned} Y \cap (ij\Sigma_6^*)_6 &= (Y_i \cap (\mathtt{1}j\Sigma_6^*)_6) + ((i-1)\mathtt{0}^*)_6 \ \cap \ (ij\Sigma_6^*)_6 \qquad (i,j \in \Sigma_6,\ i \neq \mathtt{0},\mathtt{1}), \\ Y_0 &= Y \cap \{0,1,2,3,4,5\}, \\ Y_1 &= Y \cap (\mathtt{1}\Sigma_6^+)_6. \end{aligned}$$

Proof. $\Leftarrow$ Assume that the sets Y_i satisfy the latter three equations. Then, since $\mathbb{N} = \{0,\ldots,5\} \cup (\mathtt{1}\Sigma_6^+)_6 \cup \bigcup_{i>1,j} (ij\Sigma_6^*)_6$,

$$\begin{aligned} Y = (Y \cap \{0,\ldots,5\}) \cup (Y \cap (\mathtt{1}\Sigma_6^+)_6) \cup \bigcup_{i>1,j} (Y \cap (ij\Sigma_6^*)_6) = \\ Y_0 \cup Y_1 \cup \bigcup_{\substack{i \in \{\mathtt{2},\mathtt{3},\mathtt{4},\mathtt{5}\} \\ j \in \Sigma_6}} \Big((Y_i \cap (\mathtt{1}j\Sigma_6^*)_6) + ((i-1)\mathtt{0}^*)_6 \ \cap \ (ij\Sigma_6^*)_6 \Big). \end{aligned}$$

$\Rightarrow$ Conversely, assume that (33) holds. Then, intersecting both sides of (33) with $(ij\Sigma_6^*)_6$, $\{0,\ldots 5\}$ and $(\mathtt{1}\Sigma_6^+)_6$, one obtains:

$$\begin{aligned} Y \cap (ij\Sigma_6^*)_6 &= (Y_i \cap (\mathtt{1}j\Sigma_6^*)_6) + ((i-1)\mathtt{0}^*)_6 \cap (ij\Sigma_6^*)_6 \\ Y \cap \{0,\ldots 5\} &= Y_0 \\ Y \cap (\mathtt{1}\Sigma_6^+)_6 &= Y_1. \end{aligned}$$

$\square$

An analogous result for addition and union requires introducing new variables, and so the statement looks more complicated:

Lemma 16. *There exist monotone functions* $f_{i,j}, g_{i,j}, h_{i,j} : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$, *with* $i \in \{2, \dots, 5\}$ *and* $j \in \{0, \dots, 5\}$ *and a system of equations in variables* $\{Y, Y_0, \dots, Y_5\} \cup \{\, Y_{i,j}, Y'_{i,j}, Y''_{i,j} \,|\, 2 \leqslant i \leqslant 5,\, 0 \leqslant j \leqslant 5\}$ *using the operations of union and addition, such that* $Y = S$, $Y_i = S_i$, $Y_{i,j} = S_{i,j}$, $Y'_{i,j} = S'_{i,j}$, $Y''_{i,j} = S''_{i,j}$ *with* $i \in \{2, \dots, 5\}$ *and* $j \in \{0, \dots, 5\}$ *is a solution of that system if and only if* $S_0 \subseteq \{0, 1, 2, 3, 4, 5\}$, $S_1, S_2, S_3, S_4, S_5 \subseteq (\mathtt{1}\Sigma_6^+)_6$, *and* $Y = S$, $Y_i = S_i$ *is a solution of the equation*

$$Y = Y_0 \cup Y_1 \cup \bigcup_{\substack{i \in \{\mathtt{2,3,4,5}\} \\ j \in \Sigma_6}} \Big((Y_i \cap (\mathtt{1}j\Sigma_6^*)_6) + ((i-1)\mathtt{0}^*)_6 \ \cap\ (ij\Sigma_6^*)_6 \Big),$$

and $S_{i,j} = f_{i,j}(S_i)$, $S'_{i,j} = g_{i,j}(S_i)$ *and* $S''_{i,j} = h_{i,j}(S_i)$ *for* $i \in \{2, \dots, 5\}$ *and* $j \in \{0, \dots, 5\}$.

Proof. Define

$$\begin{aligned} f_{i,j}(X) &= X \cap (\mathtt{1}j\Sigma_6^*)_6, \\ g_{i,j}(X) &= f_{i,j}(X) + ((i-1)\mathtt{0}^*)_6 \ \cap\ (ij\Sigma_6^*)_6, \\ h_{i,j}(X) &= f_{i,j}(X) + ((i-1)\mathtt{0}^*)_6 \ \cap\ (\Sigma_6^* \setminus ij\Sigma_6^*)_6. \end{aligned}$$

Note that these are monotone functions. The system of equations is constructed as follows:

$$Y = Y_0 \cup Y_1 \cup \bigcup_{i,j} Y'_{i,j} \tag{34}$$

$$Y_0 \subseteq \{0, 1, 2, 3, 4, 5\} \tag{35a}$$

$$Y_1 \subseteq (\mathtt{1}\Sigma_6^+)_6 \tag{35b}$$

$$\bigcup_{j=0}^{5} Y_{i,j} = Y_i \tag{36a}$$

$$Y_{i,j} \subseteq (\mathtt{1}j\Sigma_6^*)_k \tag{36b}$$

$$Y'_{i,j} \subseteq Y_{i,j} + ((i-1)\mathtt{0}^*)_6 \tag{37a}$$

$$Y'_{i,j} \subseteq (ij\Sigma_6^*)_6 \tag{37b}$$

$$Y''_{i,j} \subseteq Y_{i,j} + ((i-1)\mathtt{0}^*)_6 \tag{37c}$$

$$Y''_{i,j} \subseteq (\Sigma_6^* \setminus ij\Sigma_6^*)_6 \tag{37d}$$

$$Y'_{i,j} \cup Y''_{i,j} = Y_{i,j} + ((i-1)\mathtt{0}^*)_6 \tag{37e}$$

The statement of the lemma is proved separately in two directions.

$\textcircled{\Rightarrow}$ Suppose $(S, S_0 \dots, S_5, \dots, S_{i,j}, S'_{i,j}, S''_{i,j}, \dots)$ is a solution of the system (34–37e). Then, by (36b), for each $i \in \{2, \dots, 5\}$ and $j \in \{0, \dots, 5\}$,

$$S_{i,j} \subseteq (\mathtt{1}j\Sigma_6^*)_k,$$

and taking into account that $\bigcup_{j=0}^{5} S_{i,j} = S_i$ for $2 \leqslant i \leqslant 5$, it follows that $S_i \subseteq (\mathtt{1}\Sigma_6^+)_k$ holds for $S_2, \ldots, S_5$. The inclusions $S_0 \subseteq \{0,1,2,3,4,5\}$ and $S_1 \subseteq (\mathtt{1}\Sigma_6^+)_6$ are explicitly stated in the system as (35a) and (35b).

To see that $S_{i,j} = f_{i,j}(S_i)$, consider that, by (36a), $S_i = \bigcup_j S_{i,j}$, and further, by (36a) and (36b), for each j it holds that $S_{i,j} \subseteq S_i \cap (\mathtt{1}j\Sigma_6^*)_6$. Taking the union over j $\bigcup_j S_{i,j} \subseteq \bigcup_j S_i \cap (\mathtt{1}j\Sigma_6^*)_6$. The latter is, clearly, a subset of S_i, and thus

$$S_i = \bigcup_j S_{i,j} \subseteq \bigcup_j S_i \cap (\mathtt{1}j\Sigma_6^*)_6 \subseteq S_i.$$

hence the inequalities are in fact equalities. Since $S_{i,j} \subseteq (\mathtt{1}j\Sigma_6^*)_6$ and for $j \neq j'$ the sets $(\mathtt{1}j\Sigma_6^*)_6$ and $(\mathtt{1}j'\Sigma_6^*)_6$ are disjoint, for each j it holds that $S_{i,j} = S_i \cap (\mathtt{1}j\Sigma_6^*)_6$, that is $S_{i,j} = f_{i,j}(S_i)$.

The proof of $S'_{i,j} = g_{i,j}(S_i)$ and $S''_{i,j} = h_{i,j}(S_i)$ is by a similar chain of inclusions:

$$\begin{aligned} S_{i,j} + ((i-1)\mathtt{0}^*)_6 &\overset{(37\text{e})}{=} S'_{i,j} \cup S''_{i,j} \overset{(37\text{a}–37\text{d})}{\subseteq} \\ &\subseteq \Big(S_{i,j} + ((i-1)\mathtt{0}^*)_6 \,\cap\, (ij\Sigma_k^*)_6\Big) \cup \Big(S_{i,j} + ((i-1)\mathtt{0}^*)_6 \,\cap\, (\Sigma_6^* \setminus ij\Sigma_k^*)_6\Big) = \\ &\qquad = S_{i,j} + ((i-1)\mathtt{0}^*)_6. \end{aligned}$$

Therefore, the inequalities turn into equalities:

$$\begin{aligned} S'_{i,j} &= S_{i,j} + ((i-1)\mathtt{0}^*)_6 \cap (ij\Sigma_6^*)_6 = g_{i,j}(S_i) \\ S''_{i,j} &= S_{i,j} + ((i-1)\mathtt{0}^*)_6 \cap (\Sigma_6^* \setminus ij\Sigma_6^*)_6 = h_{i,j}(S_i). \end{aligned}$$

Since $(S, \ldots, S_i, \ldots, S_{i,j}, \ldots, S'_{i,j}, \ldots, S''_{i,j}, \ldots)$ satisfies (34),

$$S = S_0 \cup S_1 \cup \bigcup_{\substack{i \in \{2,3,4,5\} \\ j \in \Sigma_6}} S'_{i,j},$$

and it can be concluded that

$$\begin{aligned} S = S_0 \cup S_1 \cup \bigcup_{\substack{i \in \{2,3,4,5\} \\ j \in \Sigma_6}} S'_{i,j} = S_0 \cup S_1 \cup \bigcup_{\substack{i \in \{2,3,4,5\} \\ j \in \Sigma_6}} g_{i,j}(S_i) = \\ = S_0 \cup S_1 \cup \bigcup_{\substack{i \in \{2,3,4,5\} \\ j \in \Sigma_6}} \Big((S_i \cap (\mathtt{1}j\Sigma_6^*)_6) + ((i-1)\mathtt{0}^*)_6 \cap (ij\Sigma_6^*)_6\Big). \end{aligned}$$

Hence $(S, S_0, \ldots, S_5)$ is a solution of the equation.

$\Leftarrow\!\!\!\!\bigcirc$ Conversely, assume that $(S, S_0 \ldots, S_5)$ with $S_0 \subseteq \{0,1,2,3,4,5\}$, $S_1, S_2, S_3, S_4, S_5 \subseteq (1\Sigma_6^+)_6$. is a solution of the equation. To show that $(S, S_0 \ldots, S_5, \ldots, f_{i,j}(S_i), \ldots, g_{i,j}(S_i), \ldots, h_{i,j}(S_i), \ldots)$ is a solution of the former system, these values should be substituted into (34)–(37). For (36), the equality holds by the following calculations:

$$\begin{aligned} f_{i,j}(S_i) &= S_i \cap (\mathtt{1}j\Sigma_6^*)_6 \subseteq (\mathtt{1}j\Sigma_6^*)_6 \\ \bigcup_j f_{i,j}(S_i) &= \bigcup_j S_i \cap (\mathtt{1}j\Sigma_6^*)_6 = S_i \cap \bigcup_j (\mathtt{1}j\Sigma_6^*)_6 = S_i \cap (\mathtt{1}j\Sigma_6^*)_6 = S_i. \end{aligned}$$

In the same manner, all five equations in (37) hold true:

$$\begin{aligned}
g_{i,j}(S_i) &= f_{i,j}(S_i) + ((i-1)\mathtt{0}^*)_6 \cap (ij\Sigma_6^*)_6 \\
&\subseteq f_{i,j}(S_i) + ((i-1)\mathtt{0}^*)_6 \\
g_{i,j}(S_i) &= f_{i,j}(S_i) + ((i-1)\mathtt{0}^*)_6 \cap (ij\Sigma_6^*)_6 \\
&\subseteq (ij\Sigma_6^*)_6 \\
h_{i,j}(S_i) &= f_{i,j}(S_i) + ((i-1)\mathtt{0}^*)_6 \cap (\Sigma_6^* \setminus ij\Sigma_6^*)_6 \\
&\subseteq f_{i,j}(S_i) + ((i-1)\mathtt{0}^*)_6 \\
h_{i,j}(S_i) &= f_{i,j}(S_i) + ((i-1)\mathtt{0}^*)_6 \cap (\Sigma_6^* \setminus ij\Sigma_6^*)_6 \\
&\subseteq (\Sigma_6^* \setminus ij\Sigma_6^*)_6 \\
g_{i,j}(S_i) \cup h_{i,j}(S_i) &= \Big(f_{i,j}(S_i) + ((i-1)\mathtt{0}^*)_6 \cap (ij\Sigma_6^*)_6\Big) \\
&\cup \Big(f_{i,j}(S_i) + ((i-1)\mathtt{0}^*)_6 \cap (\Sigma_6^* \setminus ij\Sigma_6^*)_6\Big) \\
&= f_{i,j}(S_i) + ((i-1)\mathtt{0}^*)_6
\end{aligned}$$

The equality (34) follows by the assumption that $(S, S_0, \ldots, S_5)$ is a solution of the original system:

$$\begin{aligned}
S_0 \cup S_1 \cup \bigcup_{i,j} g_{i,j}(S_i) &= S_0 \cup S_1 \cup \bigcup_{i,j} f_{i,j}(S_i) + ((i-1)\mathtt{0}^*)_6 \cap (ij\Sigma_6^*)_6 = \\
&= S_0 \cup S_1 \cup \bigcup_{i,j} (S_i \cap (\mathtt{1}j\Sigma_6^*)_6) + ((i-1)\mathtt{0}^*)_6 \cap (ij\Sigma_6^*)_6 = S.
\end{aligned}$$

Finally, (35) is explicitly stated in the former system, so it clearly holds. □

Using these equivalent reformulations of equations (4) and (8), the constructions in the proofs of Lemmata 3 and 4 can be modified to use either union only or intersection only, thus proving those lemmata in their full form. This completes the proof of Theorem 4.

6 Decision problems

Consider basic properties of equations, such as the existence and the uniqueness of solutions. For the more general case of language equations it is known that these and a few other properties are undecidable [13, 15, 16], and their exact position in the arithmetical hierarchy has been determined. These results will now be re-created for equations over sets of numbers, based upon the constructions from the previous section.

Theorem 7. *The problem of whether a system of equations* $\varphi_i(X_1, \ldots, X_n) = \psi_i(X_1, \ldots, X_n)$ *over sets of natural numbers has a solution is* Π_1*-complete. It remains* Π_1*-hard if the allowed operations are union and addition, or intersection and addition.*

Proof. The problem is in Π_1 in the more general case of language equations [13].

Its Π_1-hardness is proved by a reduction from the emptiness problem for Turing machines. Let T be a TM and construct a system of equations in variables $(Y_0, \ldots, Y_5, X_1, \ldots, X_m)$ with the unique solution $Y_i = \mathrm{VALC}_i(T)$, $X_j = K_j \subseteq \mathbb{N}$. Since $S(T) = \varnothing$ if and only $\bigcup_{i=0}^{5} \mathrm{VALC}_i(T) = \varnothing$, it is sufficient to add six new equations $Y_i = \varnothing$ for $i \in \{0, 1, \ldots, 5\}$, so that the resulting system has a solution if and only if $S(T) = \varnothing$. □

Theorem 8. *Testing whether a system* $\varphi_i(X_1, \ldots, X_n) = \psi_i(X_1, \ldots, X_n)$ *over sets of natural numbers has a unique solution is a* Π_2*-complete problem. It is still* Π_2*-hard if the operations are limited to union (intersection) and addition.*

Proof. The Π_2 upper bound is known from the case of language equations [13].

Π_2-hardness is proved by a reduction from the known Π_2-complete Turing machine universality problem, which can be stated as follows: "Given a TM M working on natural numbers, determine whether it accepts every $n \in \mathbb{N}_0$". Given M, construct the system of equations as in Lemma 3. It has a unique solution if and only if the bounds $S(T) \subseteq S \subseteq \mathbb{N}$ are tight, that is, if and only if the TM accepts every number. This completes the reduction. $\square$

Theorem 9. *The problem whether a system* $\varphi_i(X_1, \ldots, X_n) = \psi_i(X_1, \ldots, X_n)$ *over sets of natural numbers has finitely many solutions is* Σ_3*-complete. Its* Σ_3*-hardness is maintained for the operations of union (intersection) and addition.*

Proof. The problem is in Σ_3 for language equations [16].

To prove Σ_3-hardness, consider the co-finiteness problem for Turing machines, which is stated as "Given a TM T working on natural numbers, determine whether $\mathbb{N} \setminus S(T)$ is finite", which is known to be Σ_3-complete [18, Cor. 14-XVI]. Given M, use Lemma 3 to construct the system of equations with the set of solutions $\{\, (S, f_1(S), \ldots, f_k(S)) \mid S(T) \subseteq S\}$. This set is finite if and only if $\mathbb{N} \setminus S(T)$ is finite, which completes the reduction. $\square$

7 Conclusion

The equations considered in this paper are a pure mathematical object and apparently a rather simple one: constructing any system with a non-periodic solution is a challenging task in itself. Unexpectedly, it turned out to be equivalent to the notion of effective computability.

This can be compared to Diophantine equations, which have been proved to be computationally complete by Matiyasevich. Due to this result, it is known, for instance, that there is a Diophantine equation for which the range of admissible values of a certain variable x is exactly the set of primes. Similarly, our Lemma 3 allows one to construct a system of equations over sets of natural numbers, which has a unique solution with one of its components being exactly the set of primes.

Among the applications of this result, it settles the expressive power of a generalization of integer circuits [11], as well as shows that language equations are computationally complete even in the seemingly trivial case of a unary alphabet.

References

[1] W. Charatonik, "Set constraints in some equational theories", *Information and Computation*, 142:1 (1998), 40–75.

[2] K. Culik II, J. Gruska, A. Salomaa, "Systolic trellis automata", I and II, *International Journal of Computer Mathematics*, 15 (1984), 195–212; 16 (1984), 3–22.

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

[4] J. Hartmanis, "Context-free languages and Turing machine computations", *Proceedings of Symposia in Applied Mathematics*, Vol. 19, AMS, 1967, 42–51.

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

[6] A. Jeż, A. Okhotin, "Conjunctive grammars over a unary alphabet: undecidability and unbounded growth", *Theory of Computing Systems*, preliminary version at CSR 2007.

[7] A. Jeż, A. Okhotin, "Complexity of solutions of equations over sets of natural numbers", *STACS 2008*.

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

[9] M. Kunc, "What do we know about language equations?", *Developments in Language Theory* (DLT 2007, Turku, Finland, July 3–6, 2007), LNCS 4588, 23–27.

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

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

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

[13] A. Okhotin, "Decision problems for language equations", *Journal of Computer and System Sciences*, to appear; preliminary version at ICALP 2003.

[14] A. Okhotin, "On the equivalence of linear conjunctive grammars to trellis automata", *Informatique Théorique et Applications*, 38:1 (2004), 69–88.

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

[16] A. Okhotin, "Strict language inequalities and their decision problems", *Mathematical Foundations of Computer Science* (MFCS 2005, Gdańsk, Poland, August 29–September 2, 2005), LNCS 3618, 708–719.

[17] A. Okhotin, "Homomorphisms preserving linear conjunctive languages", *Journal of Automata, Languages and Combinatorics*, to appear.

[18] H. Rogers, Jr., *Theory of Recursive Functions and Effective Computability*, 1967.

[19] L. J. Stockmeyer, A. R. Meyer, "Word problems requiring exponential time", *STOC 1973*, 1–9.