Automata Giving Small Certificates for Large Solutions

Christoph Haase
University of Oxford, UK

based on joint work with
R. Defossez (ENS), A. Draghici (Oxford), F. Guepin (Imperial), F. Manea (Göttingen), A. Mansutti (IMDEA), G. Perez (Antwerp), J. Rozycki (Lille), J. Worrell (Oxford)

European Research Council
Established by the European Commission

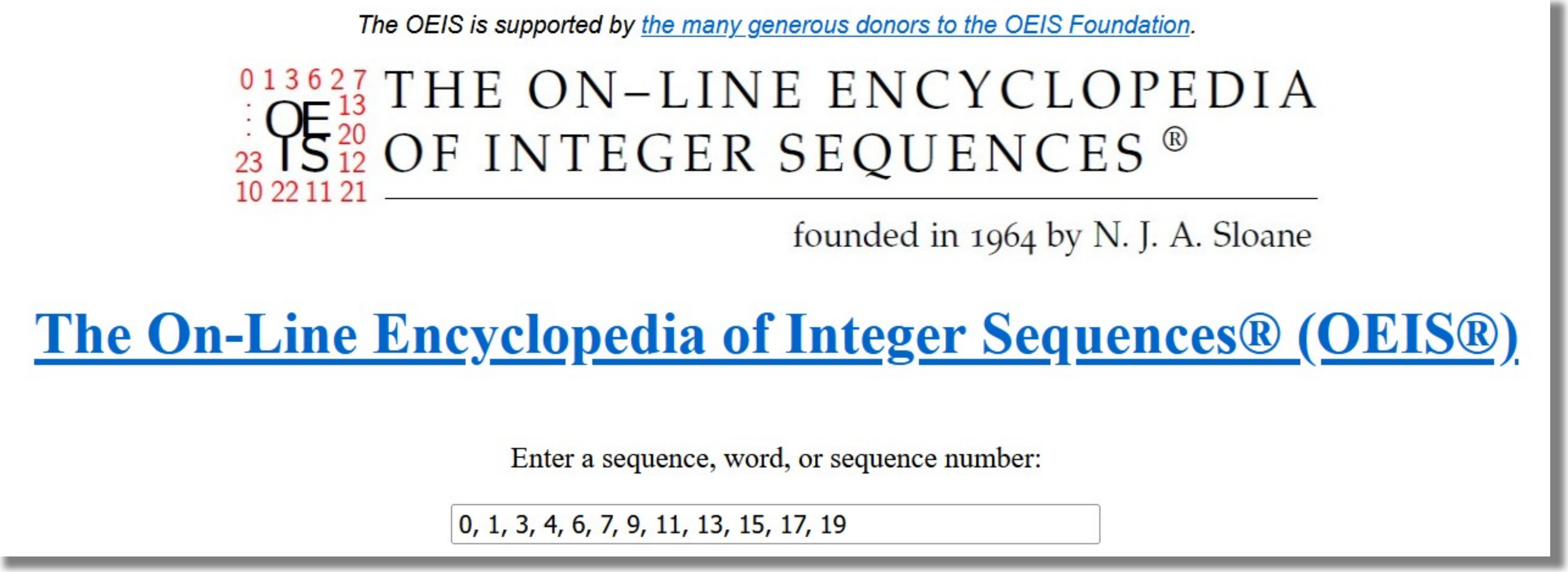
Theorietag, 5th October 2023

The Connoisseur of Number Sequences

(c) John Smock for Quanta Magazine

Neil Sloane (*1939)

OEIS


The OEIS is supported by the many generous donors to the OEIS Foundation.
THE ON–LINE ENCYCLOPEDIA OF INTEGER SEQUENCES ®
founded in 1964 by N. J. A. Sloane
The On-Line Encyclopedia of Integer Sequences® (OEIS®)
Enter a sequence, word, or sequence number:
0, 1, 3, 4, 6, 7, 9, 11, 13, 15, 17, 19

Numberphile

Problems with Powers of Two Problem

Given a set of integers S, denote by $b(S)$ the number of powers of two that can be obtained as the sum of two elements of S.

Examples:

- $S = \{1, 3\},\ b(S) = 1$
- $S = \{-1, 3, 5\},\ b(S) = 3$
- $S = \{-3, -1, 3, 5\},\ b(S) = 4$

Problems with Powers of Two Problem

Denote by $a(n)$ the largest value of $b(S)$ that can be achieved for a set S with n elements.

n	1	2	3	4	5	6	7	8	9	10	11	12
$a(n)$	0	1	3	4	6	7	9	11	13	15	17	19

Largest known value: $a(18) = 34$

Upper bound: $$a(n) \leq \frac{n}{4} \cdot \sqrt{4n-3} + 1$$

OEIS A352178

The OEIS is supported by [the many generous donors to the OEIS Foundation](#).

THE ON–LINE ENCYCLOPEDIA OF INTEGER SEQUENCES ®

founded in 1964 by N. J. A. Sloane

Search Hints

(Greetings from The On-Line Encyclopedia of Integer Sequences!)

A352178 Let S = {t_1, t_2, ..., t_n} be a set of n distinct integers and consider the sums t_i + t_j (i<j); a(n) is the maximum number of such sums that are powers of 2, over all choices for S. 4

0, 1, 3, 4, 6, 7, 9, 11, 13, 15, 17, 19, 21, 24, 26, 29, 31, 34 (list; graph; refs; listen; history; text; internal format)

OFFSET 1,3

COMMENTS

Given distinct integers t_1, ..., t_n, form a graph G with n vertices labeled by the t_i, and with an edge from t_i to t_j, labeled t_i + t_j, whenever t_i + t_j is a power of 2.

See the Pratt link for the best lower bounds known, and examples of sets achieving these bounds, for 1 <= n <= 100. - N. J. A. Sloane, Sep 26 2022

The following remarkable theorem is due to M. S. Smith (email of Mar 06 2022).

Theorem: G contains no 4-cycles.

Proof. Suppose the contrary, and assume the vertices t_1, t_2, t_3, t_4 form a 4-cycle, with edges labeled b_1 = t_1+t_2, b_2 = t_2+t_3, b_3 = t_3+t_4, b_4 = t_4+t_1. The b_i are powers of 2.

Since the t_i are distinct, b_1 != b_4, b_2 != b_1, b_3 != b_2, and b_4 != b_3.

We also have

Open Problems and Challenges

- How does A352178 continue?

  Nobody knows! (but some lower and upper bounds are known)

- Is it possible to continue A352178, at least in theory?

  Nobody knew, iterating over all sets with n integers is not possible...

A Logicians View on the Problem

To determine whether $a(n) \geq k$:

- Find integers $z_1, z_2, \ldots, z_n$

  $\exists z_1, z_2, \ldots, z_n$

- For every pair $i < j$ an indicator variable $x_{i,j} \in \{0,1\}$ assigning 1 to $x_{i,j}$ exactly when $z_i + z_j$ is a power of 2

  $$\exists x_{1,2} x_{1,3} \ldots x_{n-1,n}\, P_2(z_1 + z_2) \rightarrow x_{1,2} = 1 \wedge$$
  $$\neg P_2(z_1 + z_2) \rightarrow x_{1,2} = 0 \wedge \cdots$$

- The sum of all indicator variables is at least k

  $x_{1,2} + x_{1,3} + \cdots + x_{n-1,n} \geq k$

Büchi Arithmetic

Logical formula obtained is statement in Büchi arithmetic, which is an automatic structure:

- Numbers are just sequences of digits
- Can define DFA for basic relations
- Use closure properties of regular languages under boolean operations, homomorphisms and inverse homomorphisms to decide logical theory

J.R. Büchi
(1924 – 1984)

Véronique
Bruyère

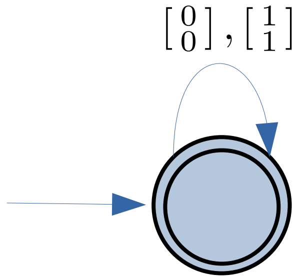
Presburger Arithmetic

First-order theory of $(\mathbb{N}, 0, 1, +, =)$

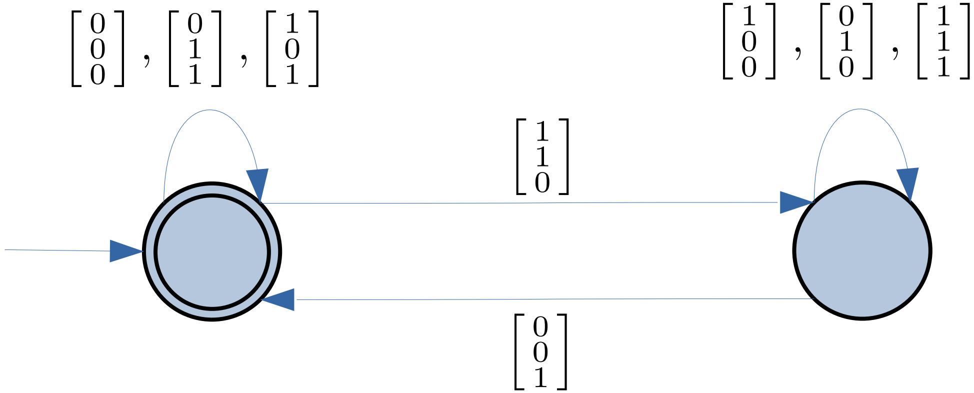
Represent $\boldsymbol{x} \in \mathbb{N}^d$ as strings over the alphabet

$$\Sigma_d = \left\{ \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \right\}$$

Gadget for $x = y$:



Gadget for $x + y = z$:

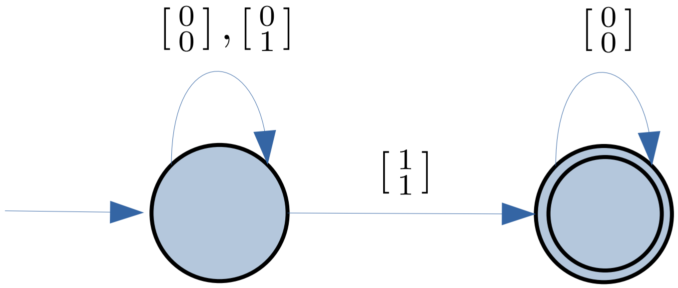


Büchi Arithmetic

First-order theory of $(\mathbb{N}, 0, 1, +, V_p, =)$ for fixed $p > 1$:

$V_p(x, y) \Leftrightarrow$ x is the largest power of p dividing y without remainder

Gadget for $V_2(x, y)$:



<u>Theorem</u> (Büchi, 1960, Bruyere 1985; H., Różycki , 2021)
Sets definable in Büchi arithmetic coincide with regular languages. Büchi arithmetic is not model-complete.

Dealing with Negative Numbers

Sloane's problem requires looking for integer solutions:

- Encode numbers in base -2:

$$23 = 1 \cdot (-2)^0 + 1 \cdot (-2)^1 + 0 \cdot (-2)^2 + 1 \cdot (-2)^3 + 0 \cdot (-2)^4 + 1 \cdot (-2)^5 + 1 \cdot (-2)^6$$

- DFA for addition becomes a bit more complicated:

A Partial Answer to the Power-of-Two Problem

The constructed NFA shrink the search space:

- There is a constant c such that to check whether $a(n) \geq k$, it suffices to "only" consider sets with integers in the interval

$$\{-2^{2^{c \cdot n}}, \ldots, 2^{2^{c \cdot n}}\}$$

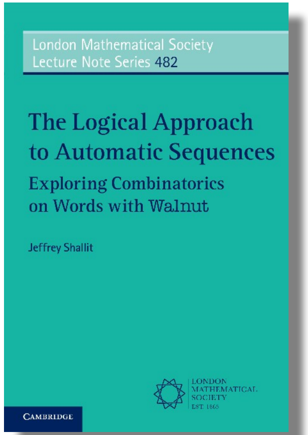
The NFA become huge – $a(3) \geq 3$

Practical Aspects

The vanilla logical approach not truly advantageous in practice:

- Walnut quickly runs out of resources
- SMT-solvers do not support power-of-two assertions, but ad-hoc approach finds in reasonable time $a(18) = 34$
- Still $a(19)$ hits the wall

Jeffrey Shallit
(*1957)



Solutions May Have Super-Polynomial Bit Length

Theorem (Matthews, 1982)
There are infinitely many primes p such that the multiplicative order of two is at least $\sqrt{p}$.

Multiples of multiplicative order of two definable in Büchi arithmetic:

$$\Phi_p(x) \equiv x > 1 \wedge P_2(x) \wedge \exists y\, x - p \cdot y = 1$$

Theorem (Guepin, H., Worrell, 2019)
Existential Büchi arithmetic is NP-complete.

An Alternative DFA construction

DFA accepting solutions of system of equations $\boldsymbol{a} \cdot \boldsymbol{x} = c$

$M = (Q, \{0,1\}^d, \delta, q_0, F)$:

- $Q = \mathbb{Z} \cup \{\perp\}$
- $q_0 = 0$
- $\delta(z, \boldsymbol{u}) = 2z + \boldsymbol{a} \cdot \boldsymbol{u}$ for all $z \in \mathbb{Z}$
- $\delta(\perp, \boldsymbol{u}) = \perp$ for all $\boldsymbol{u} \in \{0,1\}^d$
- $F = \{c\}$

After reading $\boldsymbol{u}_{n-1} \cdots \boldsymbol{u}_0 \in \{0,1\}^d$, automaton is in state

$$\sum_{i=0}^{n-1} \boldsymbol{a} \cdot \boldsymbol{u}_i$$

Certificates Witnessing Existence of Solutions

Given configurations $\boldsymbol{q}, \boldsymbol{r} \in \mathbb{Z}^m$ of DFA for $A \cdot \boldsymbol{x} = \boldsymbol{c}$ reading solutions in base p, have

$\boldsymbol{q}$ reaches $\boldsymbol{r}$

$$\iff$$

there exists $k \in \mathbb{N}$ and $\boldsymbol{u} \in \{0, \ldots, p^k - 1\}^d$ such that

$$\boldsymbol{r} = p^k \cdot \boldsymbol{q} + A \cdot \boldsymbol{u}$$

To decide existential Büchi arithmetic in NP, guess polynomially many configurations and check reachability

p-Universality

- p-universality: Given existential formula of Büchi arithmetic, is it satisfiable in every (prime) base p?
- Number of states of DFA for $A \cdot \boldsymbol{x} = \boldsymbol{c}$ independent of base p

Theorem (H., Mansutti, 2021)
Deciding p-universality is coNEXP-complete.

Existential Presburger Arithmetic with Divisbility

p-universality cornerstone of decidability of existential Presburger arithmetic with divisibility:

- Atomic formulas:

$$(a_1 \cdot x_1 + \cdots + a_d \cdot x_d + a_0) \mid (b_1 \cdot x_1 + \cdots + b_d \cdot x_d + b_0)$$

- Generalizes systems of linear congruences

- Smallest solutions can be large: $x_n \geq 2^{2^n}$ for

$$\Phi_n \equiv x_n > 1 \wedge \bigwedge_{i=0}^{n-1} x_i > 1 \wedge (x_i \mid x_{i+1}) \wedge (x_i + 1 \mid x_{i+1})$$

Existential Presburger Arithmetic with Divisbility

Lipshitz (1978) showed certain local-to-global property:

- Every formula Φ is equi-satisfiable with some Ψ in *increasing form* such that Ψ has a solution in $\mathbb{Z}$ iff

  Ψ has a solution modulo every prime p, i.e., for every $f(\boldsymbol{x}) \mid g(\boldsymbol{x})$ in Ψ, find $\boldsymbol{x}_p$ such that
  $$f(\boldsymbol{x}_p) \neq 0, \ v_p(f(\boldsymbol{x}_p)) \leq v_p(g(\boldsymbol{x}_p))$$

- Allows to deduce NEXP upper bound

Complexity of Increasing Formulas

Theorem (Defossez, H., Mansutti, Perez, 2023)
Increasing formulas of existential Presburger arithmetic with divisibility are decidable in NP.

- Show that only a polynomial number of prime numbers are essential for local-to-global property
- Then perform polynomial number of queries to existential Büchi arithmetic

From Local to Global Solutions

Going from local to global solutions requires combining solutions modulo p via the Chinese remainder theorem

Theorem (Defossez, H., Mansutti, Perez, 2023)

If a system of congruences and non-congruences

$$x \equiv b_m \pmod{m} \qquad m \in M$$

$$x \not\equiv c_{q,i} \pmod{q} \qquad q \in Q \subseteq \mathbb{P}, 1 \leq i \leq d$$

has a solution then it has a solution bounded by

$$\prod M \cdot ((d+1) \cdot \#Q)^{4(d+1)^2(3+\ln\ln(\#Q+1))}$$

Integer Programming with GCD constraints

With further technical developments, can show small-model property for generalisation of integer programming:

$$\begin{array}{lll} \text{minimize} & \boldsymbol{c} \cdot \boldsymbol{x} & \\ \text{subject to} & A \cdot \boldsymbol{x} \leq \boldsymbol{b} & \\ & \gcd(f_i(\boldsymbol{x}), g_i(\boldsymbol{x})) \sim_i d_i, & 1 \leq i \leq k \end{array}$$

Theorem (Defossez, H., Mansutti, Perez, 2023)
If an instance of IP-GCD is feasible then it has a solution (and an optimal solution, if one exists) of polynomial bit length. Hence, IP-GCD feasibility is NP-complete.

Semenov arithmetic

First-order theory of the structure $(\mathbb{N}, 0, 1, +, 2^x)$

- Shown decidable by Semenov in 1984
- Model-complete and admits quantifier eliminiation (Cherlin and Point, 1986)
- Non-elementary complexity in general:

$$p\colon (x, y) \mapsto 2^{2^x} + 2^{2y+1}$$

- Existential fragment decidable in NEXP (Benedikt et al. 2023), but also decidable if regular predicates allowed?

Semenov arithmetic is not automatic

<u>Theorem</u> (Khoussainov, Nerode, 1995)
If $f\colon \mathbb{U}^n \to \mathbb{U}$ is a function whose graph is a regular relation then there is a constant $C > 0$ such that for all $u_1, \ldots, u_n \in \mathbb{U}$:

$$|f(u_1, \ldots, u_n)| \leq \max(|u_1|, \ldots, |u_n|) + C$$

- No such constant exists for $f\colon x \mapsto 2^x$
- Quantifier-elimination incompatible with regular predicates, since theory undecidable if V_2 included

Power functions

Suppose we have $x = 2^y$ then:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} \cdots \\ \cdots \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- Length of string of tailing zeros of x equals value of y
- Keep two counters when reading tuple of numbers:
  - First counter tracks number of tailing zeros of x
  - Second counter computes value of y
- Accept if both counters have the same value

Affine Vector Addition Systems with States (VASS)

- Finite-state automata with finite number of counters taking values from $\mathbb{N}$
- Transitions update counters by affine functions

$$f : x \mapsto a \cdot x + b, \;\; a, b \in \mathbb{Z}$$

- Languages closed under union and intersection

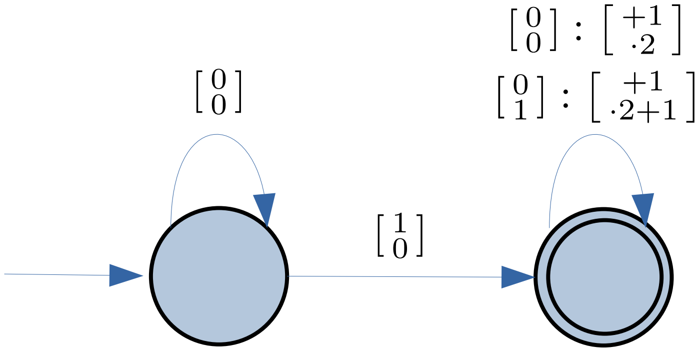
Theorem (Finkel, Göller, H., 2013; Reichert, 2015; Jaax, Kiefer, 2020)
Reachability in affine VASS with a single counter is PSpace-complete and undecidable for two counters.

Affine VASS for power functions

Suppose we have $x = 2^y$ then:

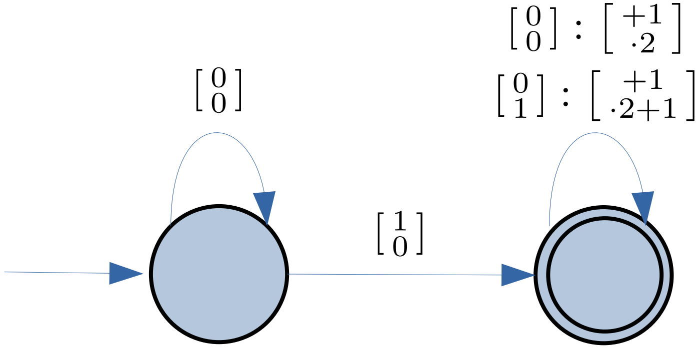
$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{matrix} \cdots \\ \cdots \end{matrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Affine 2-VASS accepting such strings



Accept if in accepting state and both counters have same value

Restricted affine VASS



- Counters can be grouped into pairs
- Once first counter of pair has value 1, it gets incremented and only incremented at every transition
- Languages of restricted affine VASS closed under union and intersection

A counter elimination procedure

For deciding emptiness with arbitrary number of pairs:

- Guess ordering in which counters become non-zero
- Successively eliminate counters until finite-state automaton is obtained with language equi-nonempty
- Number of control states squares in every iteration

Theorem (Draghici, H., Manea, 2023)
Deciding language emptiness of restricted affine VASS is in EXPSpace.

A Decision Procedure

To decide existential $(\mathbb{N}, 0, 1, +, 2^x, (R_k)_{k>0})$

- Formula given as positive Boolean combination of
  - linear equations $a_1 \cdot x_1 + \cdots + a_n \cdot x_n = b$
  - $R_i(x_1, \ldots, x_{d_i})$
  - applications of powering function $x = 2^y$

- Yields equi-nonempty exponential-size restricted affine VASS

Theorem (Draghici, H., Manea, 2023)
Existential $(\mathbb{N}, 0, 1, +, 2^x, (R_k)_{k>0})$ is in EXPSpace.

String Constraints

Two-sorted logic with (subset of) consisting of:

- Systems of equations of the form:

$$x \cdot y = z, \ \ x, y, z \in \{0,1\}^*$$

- Length function $\ell \colon \{0,1\}^* \to \mathbb{N}$:

$$\ell \colon b_0 \cdots b_k \mapsto k + 1$$

- Number-to-string conversions $\mathrm{nr2str} \colon \mathbb{N} \to \{0,1\}^*$:

$$\mathrm{nr2str} \colon n \mapsto \left\{ b_0 \cdots b_k : n = \sum_{i=0}^{k} 2^i \cdot b_i \right\}$$

- Regular membership: $R(x)$
- Presburger constraints: $\Phi(u_1, \ldots, u_n)$

Word equations

Theorem (Makanin, 1977; Jez 2017)
Simple word equations are decidable in non-deterministic linear space.

Theorem (Berzish et al., 2021)
String constraints with length constraints, number-to-string functions and regular language membership are undecidable.

A decidable fragment

Theorem (Draghici, H., Manea, 2023)
The existential theory of string constraints with length constraints

- number-to-string constraints, and
- regular language membership

polynomial-time reduces to $(\mathbb{N}, 0, 1, +, 2^x, (R_k)_{k>0})$ and is hence decidable in EXPSpace.

Idea: Map $s \in \{0,1\}^*$ to $1 \cdot s$