Universal Skolem Sets

Florian Luca[*], Joël Ouaknine[†] and James Worrell[‡]
[*]School of Mathematics, University of Witwatersrand, South Africa
Research Group in Algebraic Structures and Applications, King Adulaziz University, Jeddah, Saudi Arabia
Centro de Ciencas Matemaáticas UNAM, Morelia, Mexico
Email: florian.luca@wits.ac.za
[†] Max Planck Institute for Software Systems, Saarland Informatics Campus, Saarbrücken, Germany
Email: joel@mpi-sws.org
[‡] Department of Computer Science, University of Oxford, United Kingdom
Email: jbw@cs.ox.ac.uk

***Abstract*—It is a longstanding open problem whether there is an algorithm to decide the Skolem Problem for linear recurrence sequences, namely whether a given such sequence has a zero term. In this paper we introduce the notion of a *Universal Skolem Set*: an infinite subset $\mathscr{S}$ of the positive integers such that there is an effective procedure that inputs a linear recurrence sequence $\boldsymbol{u} = (u(n))_{n\geq 0}$ and decides whether $u(n) = 0$ for some $n \in \mathscr{S}$. The main technical contribution of the paper is to exhibit such a set.**

I. Introduction

A linear recurrence sequence (LRS) $\boldsymbol{u} = (u(n))_{n\geq 0}$ is a sequence of integers satisfying a recurrence of the form

$$u(n+k) = a_1 u(n+k-1) + \cdots + a_k u(n) \qquad (n \in \mathbb{N}), \qquad (1)$$

where the coefficients $a_1, \ldots, a_k$ are integers. The celebrated theorem of Skolem, Mahler, and Lech (see [6]) describes the structure of the set $Z(\boldsymbol{u}) := \{n \in \mathbb{N} : u(n) = 0\}$ of zero terms of such a recurrence:

Theorem 1: Given an integer linear recurrence sequence $(u(n))_{n\geq 0}$, the set $Z(\boldsymbol{u})$ is a union of finitely many arithmetic progressions together with a finite set.

The statement of Theorem 1 can be refined by considering the notion of *non-degeneracy* of an LRS. An LRS is non-degenerate if in its minimal recurrence the quotient of no two distinct roots of the characteristic polynomial is a root of unity. A given LRS can be effectively decomposed as an interleaving of finitely many non-degenerate sequences, some of which may be identically zero. The core of the Skolem-Mahler-Lech Theorem is the fact that a non-zero non-degenerate linear recurrence sequence has finitely many zero terms. Unfortunately, all known proofs of this last result are ineffective: it is not known how to compute the finite set of zeros of a given non-degenerate linear recurrence sequence. It is readily seen that the existence of a procedure to do so is equivalent to the existence of a procedure to decide whether an arbitrary given LRS has a zero term. The problem of deciding whether an LRS has a zero term is variously known as Skolem's Problem or the Skolem-Pisot Problem.

978-1-6654-4895-6/21/$31.00

Decidability of Skolem's Problem is known only for certain special cases, based on the relative order of the absolute values of the characteristic roots. Say that a characteristic root λ is *dominant* if its absolute value is maximal among all the characteristic roots. Decidability is known in case there are at most 3 dominant characteristic roots, and also for recurrences of order at most 4 [12], [18]. However for LRS of order 5 it is not currently known how to decide Skolem's Problem—the hard case being that all characteristic roots are simple, with four dominant roots.

In computer science, Skolem's Problem has been studied in the context of formal power series [15], [2], stochastic model checking [14], control theory [3], [8], and loop termination [13]. The problem is often used as a reference to show hardness of other open decision problems.

Our aim in this paper is to initiate an alternative approach to the decidability of Skolem's Problem. Rather than place restrictions on sequences (e.g., on the order of the recurrence or dominance pattern of the characteristic roots), the idea is to restrict the domain in which to search for zeros. To this end, we give the following definition.

Definition 2: We say that an infinite set $\mathscr{S} \subseteq \mathbb{N}$ is a *Universal Skolem Set* if there is an effective procedure that, given any integer linear recurrence sequence $\boldsymbol{u}$, outputs whether or not there exists $n \in \mathscr{S}$ with $u(n) = 0$.

The main technical contribution of the paper is to exhibit a Universal Skolem Set. Specifically, we have:

Theorem 3: Define $f : \mathbb{N} \setminus \{0\} \to \mathbb{N}$ by $f(n) = \lfloor \sqrt{\log n} \rfloor$, and define the sequence $(s_n)_{n\geq 0}$ inductively by $s_0 = 1$ and $s_n = n! + s_{f(n)}$ for $n > 0$. Then $\mathscr{S} := \{s_n : n \in \mathbb{N}\}$ is a Universal Skolem Set.

To prove the theorem, the key idea is to show that given an integer LRS $\boldsymbol{u}$, one can compute an effective threshold N such that for all $n \geq N$, if $u(s_n) = 0$ then also $u(s_{f(n)}) = 0$. Since the function f is strictly decreasing, this entails that if $\boldsymbol{u}$ has a zero term in $\mathscr{S}$ then already $u(s_n) = 0$ for some $n < N$. This suffices to prove Theorem 3.

In addition to Theorem 3, we show that one can decide whether a given LRS has infinitely many zeros in $\mathscr{S}$ and, in case the number of zeros is finite, we obtain an effective upper bound on n such that $u(s_n) = 0$. For the latter we use a result of Schlickewei and Schmidt [17] stating that for a non-zero non-degenerate LRS $\boldsymbol{u}$ of order k, one has $\#Z(\boldsymbol{u}) \le (2k)^{35k^3}$.

A. Related Work

A forerunner of the present paper is [10], which described how to decide the existence of prime powers in the set $Z(\boldsymbol{u})$ for a highly restricted class of LRS; see Section III-A for more details. Due to this restriction [10] did not provide a Universal Skolem Set in the sense of Definition 2.

Definition 2 is inspired by the notion of a *Universal Hilbert Set*, which we now briefly recall. Let $P(X,Y) \in \mathbb{Q}[X,Y]$ be an irreducible polynomial in two variables in which X has degree at least two. Hilbert's Irreducibility Theorem asserts that the set

$$S_P = \{n \in \mathbb{Z} : P(X,n) \text{ is reducible in } \mathbb{Q}[X]\}$$

has density zero, i.e.,

$$\lim_{T\to\infty} \frac{1}{T}\#(S_P \cap [-T,T]) = 0\,.$$

In fact, S. D. Cohen [4] proved that $\#(S_P \cap [-T,T]) = O(T^{1/2}\log T)$. On the other hand, there are polynomials P for which $\#(S_P \cap [-T,T]) = \Omega(T^{1/2})$, for example $P(X,Y) = X^2 - Y$ for which $S_P = \{m^2 : m \in \mathbb{Z}\}$. Motivated by such a result, a Universal Hilbert Set is an infinite set S of integers such that $S \cap S_P$ is finite for all irreducible polynomials $P(X,Y) \in \mathbb{Q}[X,Y]$. Bilu [1] proved that

$$\{m^3 + \lfloor \log\log|m| \rfloor : m \in \mathbb{Z},\ |m| \ge 3\}$$

is a Universal Hilbert Set, while Filaseta and Wilcox [9] constructed a dense Universal Hilbert Set.

II. Preliminaries on Linear Recurrence Sequences

A. Exponential-Polynomial Representation

Consider an integer sequence $\boldsymbol{u} = (u(n))_{n\ge 0}$ satisfying the following recurrence with coefficients $a_1,\dots,a_k \in \mathbb{Z}$:

$$u(n+k) = a_1 u(n+k-1) + \cdots + a_k u(n) \qquad (n \in \mathbb{N})\,. \tag{2}$$

We call k the *order* of the recurrence. While $\boldsymbol{u}$ may satisfy many different linear recurrences, it satisfies a unique recurrence of minimal order and we henceforth assume that the recurrence (2) is minimal for $\boldsymbol{u}$.

We write

$$F(X) := X^k - a_1 X^{k-1} - \cdots - a_k$$

for the *characteristic polynomial* of the recurrence (2). Assume that

$$F(X) = \prod_{i=1}^{\ell} (X - \lambda_i)^{\sigma_i}$$

is the factorisation of $F(X)$ in $\mathbb{C}[X]$. Here, $\lambda_1,\dots,\lambda_\ell$ are the distinct roots of $F(X)$, of multiplicity $\sigma_1,\dots,\sigma_\ell$, respectively. Note that minimality of the recurrence (2) implies that $a_k \ne 0$ and hence that the characteristic roots are non-zero.

We put $\mathbb{K} := \mathbb{Q}(\lambda_1,\dots,\lambda_\ell)$ for the splitting field of $F(X)$. It is well-known that $\boldsymbol{u}$ admits an exponential-polynomial representation:

$$u(n) = \sum_{i=1}^{\ell} A_i(n)\lambda_i^n, \tag{3}$$

where $A_i(X) \in \mathbb{K}[X]$ is a polynomial of degree at most $\sigma_i - 1$ for $i = 1,\dots,\ell$. Writing, for all $i = 1,\dots,\ell$,

$$A_i(X) = a_{i,0} + a_{i,1}X + \cdots + a_{i,\sigma_i-1}X^{\sigma_i-1}\,,$$

the vector of coefficients

$$(a_{1,0},\dots,a_{1,\sigma_1-1},a_{2,0},\dots,a_{2,\sigma_2-1},\dots,a_{\ell,0},\dots,a_{\ell,\sigma_\ell-1}) \tag{4}$$

is a solution of a system of linear equations determined by the first k values of the sequence $\boldsymbol{u}$:

$$\sum_{i=1}^{\ell} A_i(n)\lambda_i^n = u(n) \quad \text{for} \quad n = 0,1,\dots,k-1\,. \tag{5}$$

This is a system of k linear equations in the k unknowns shown in (4). The coefficient matrix for this linear system is given in Fig. 1.

As shown in [7], the determinant of this matrix is

$$\det(\mathrm{Coef}) = \prod_{i=1}^{\ell}\prod_{j=1}^{\sigma_i-1} j! \prod_{i=1}^{\ell} \lambda_i^{\binom{\sigma_i}{2}} \prod_{1\le i<j\le \ell} (\lambda_j - \lambda_i)^{\sigma_i\sigma_j}\,. \tag{6}$$

Equation (6) shows that $\det(\mathrm{Coef})^2$ is non-zero and, being a symmetric expression in the algebraic integers $\lambda_1,\dots,\lambda_\ell$, is a rational integer.

Considering the recurrence (2), write

$$L := 1 + |a_1| + \ldots + |a_k|\,. \tag{7}$$

For each characteristic root λ_i such that $|\lambda_i| > 1$ we have

$$|\lambda_i| = \left| a_1 + \frac{a_2}{\lambda_i} + \cdots + \frac{a_k}{\lambda_i^{k-1}} \right| \le |a_1| + \cdots |a_k| \le L.$$

Hence $|\lambda_i| \le L$ for all $i \in \{1,\dots,\ell\}$. In combination with the determinant formula (6), it follows that

$$(\det(\mathrm{Coef}))^2 < (k!)^{2k} \cdot 2^{k^2} \cdot L^{2k^2}\,. \tag{8}$$

Write $\mathcal{O}$ for the ring of integers of the splitting field $\mathbb{K}$. The characteristic roots, $\lambda_1,\dots,\lambda_\ell$ all lie in $\mathcal{O}$ since they are roots of a monic polynomial with integer coefficients. Moreover, from the inequality (8), we can apply Cramer's rule to solve the system (5) and deduce that each coefficient $a_{i,j}$ in (4) is such that $\det(\mathrm{Coef})^2 a_{i,j} \in \mathcal{O}$. We immediately deduce the following proposition:

Proposition 4: There is a positive integer $C_0 < (k!)^{2k}2^{k^2}L^{2k^2}$ such that in the exponential-polynomial closed form

$$u(n) = \sum_{i=1}^{\ell} A_i(n)\lambda_i^n,$$

the coefficients of polynomial $A_i(X)$ lie in $\frac{1}{C_0}\mathcal{O}$ for $i = 1,\dots,\ell$.

$$\mathrm{Coef} := \begin{pmatrix} 1 & \ldots & 0 & 1 & \cdots & 0 & 1 & \ldots \\ \lambda_1 & \ldots & \lambda_1 & \lambda_2 & \cdots & \lambda_{\ell-1} & \lambda_\ell & \ldots \\ \lambda_1^2 & \ldots & 2^{\sigma_1-1}\lambda_1^2 & \lambda_2^2 & \cdots & 2^{\sigma_{\ell-1}-1}\lambda_{\ell-1}^2 & \lambda_\ell^2 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \lambda_1^{k-1} & \cdots & (k-1)^{\sigma_1-1}\lambda_1^{k-1} & \lambda_2^{k-1} & \cdots & (k-1)^{\sigma_{\ell-1}}\lambda_{\ell-1}^{k-1} & \lambda_\ell^{k-1} & \ldots \end{pmatrix}.$$

Fig. 1. Coefficient matrix for the linear system (5).

B. Non-Degeneracy

An LRS is said to be *degenerate* if in its minimal recurrence there are two distinct characteristic roots whose quotient is a root of unity. A non-degenerate LRS over any field of characteristic zero is either identically zero or has only finitely many zeros. In this section we recall for later use some well-known constructions that allow one to reduce the general case of Skolem's Problem to the case in which the given LRS is non-degenerate.

As in Subsection II-A we assume an integer sequence $\boldsymbol{u} = (u(n))_{\geq 0}$ that satisfies an order-k recurrence with characteristic polynomial

$$F(X) = X^k - a_1 X^{k-1} - \cdots a_k .$$

Let $\lambda_1, \ldots, \lambda_\ell$ be the roots of $F(X)$, with respective multiplicities $\sigma_1, \ldots, \sigma_\ell$.

We recall the following straightforward proposition.

Proposition 5: Let M be the least common multiple of the orders of the roots of unity appearing among the quotients of distinct characteristic roots of $\boldsymbol{u}$. Then for each $0 \leq j < M$, the subsequence $\boldsymbol{u}_{M,j} := (u(Mn+j))_{n\geq 0}$ is non-degenerate and satisfies a recurrence of order at most k (the order of the recurrence defining $\boldsymbol{u}$).

Proof It can be shown (see, e.g., [5, Lemma 2.1]) that for each $0 \leq j < M$, the subsequence $\boldsymbol{u}_{M,j}$ satisfies a recurrence whose characteristic polynomial is

$$G(X) := (-1)^{k(M+1)} \mathrm{Res}_Y (Y^M - X, F(Y)) .$$

Now $G(X)$ has degree at most k and has roots among $\{\lambda_1^M, \ldots, \lambda_\ell^M\}$. This implies that $\boldsymbol{u}_{M,j}$ is non-degenerate and satisfies an order-k recurrence. □

In the remainder of this section we explain how the constant M in Proposition 5 may be computed. Given distinct characteristic roots λ_i, λ_j, the quotient λ_i/λ_j is a number of degree at most $k(k-1)$, where k is the order of the recurrence (2). Hence if the quotient is a root of unity of order m, then $\varphi(m) \leq k(k-1)$, where φ is the Euler function. Now an elementary inequality states that $\varphi(m) \geq m/(2\log(2m))$ for all $m \geq 1$ (see page 279 in [11]). Define M_0 to be the largest positive integer m such that $m/(2\log(2m)) \leq k(k-1)$. Then M_0 is the maximum order of a root of unity appearing among the quotients λ_i/λ_j.

The first step in computing M is to find the polynomial

$$f(X) := \prod_{i=1}^{\ell} (X - \lambda_i) .$$

This is the *radical* polynomial of $F(X)$, which is monic and has all the roots of $F(X)$ as simple roots. To find it, one starts by computing the polynomial $D(X) := \gcd(F(X), F'(X))$ which can be done by using the Euclidean algorithm for polynomials. Note that $D(X) = \prod_{i=1}^{\ell} (X - \lambda_i)^{\sigma_i - 1}$. Thus, $D(X)$ is a divisor of $F(X)$ and $f(X) = F(X)/D(X)$. Having $f(X)$ we would like to test if $f(X)$ has two roots $\lambda_i \neq \lambda_j$ such that λ_i/λ_j is a root of unity of order a given positive integer $m \leq M_0$. To do this we consider

$$P_m(X) := \mathrm{Res}_Y \left(f(XY), \frac{Y^m - 1}{Y - 1} \right).$$

This is a polynomial in X whose leading term is ± 1 and whose roots are of the form λ_i/ζ, where $\zeta \neq 1$ satisfies $\zeta^m = 1$. Thus, in order to test whether $\lambda_i/\lambda_j = \zeta$ for some $i \neq j$ and such ζ, it suffices to compute

$$\mathrm{Res}(P_m(X), f(X))$$

and see whether this is the zero integer or not. In case it is the zero integer, then $(\lambda_i/\lambda_j)^m = 1$ for some distinct roots $\lambda_i \neq \lambda_j$ of $F(X)$.

III. Construction of An Infinite Universal Skolem Set

A. A Motivating Example

To give some intuition we start by briefly recalling from [10] a class of recurrence sequences $\boldsymbol{u} = (u(n))_{n\geq 0}$ for which one can decide, for each fixed constant c, whether there exists $n \in \{p^k : p \text{ prime}, k \leq c\}$ with $u(n) = 0$. The condition that we place on $\boldsymbol{u}$ is that the coefficients A_i in the exponential-polynomial representation (3) be rational numbers. Then, by rescaling, we can assume that $u(n) = \sum_{i=1}^{\ell} A_i \lambda_i^n$, where $A_1, \ldots, A_\ell \in \mathbb{Z}$ and $\lambda_1, \ldots, \lambda_\ell$ lie in the ring $\mathcal{O}$ of algebraic integers in some number field.[1] Under this condition, we have the following calculation:

$$\begin{aligned} u(1)^{p^k} &= (A_1\lambda_1 + \cdots + A_\ell \lambda_\ell)^{p^k} \\ &= A_1^{p^k}\lambda_1^{p^k} + \cdots + A_\ell^{p^k}\lambda_\ell^{p^k} + \\ &\qquad \sum_{\substack{0 \leq \alpha_1,\ldots,\alpha_\ell \leq p^k - 1 \\ \alpha_1 + \cdots + \alpha_\ell = p^k}} \binom{p^k}{\alpha_1, \ldots, \alpha_\ell} (A_1\lambda_1)^{\alpha_1} \cdots (A_\ell \lambda_\ell)^{\alpha_\ell} \\ &\equiv A_1 \lambda_1^{p^k} + \cdots + A_\ell \lambda_\ell^{p^k} \pmod{p\mathcal{O}} \\ &= u(p^k), \end{aligned}$$

[1] If $A_1, \ldots, A_\ell \in \mathbb{Z}$ and $\lambda_1, \ldots, \lambda_\ell$ is a complete set of roots of a polynomial with rational coefficients, then the expression $\sum_{i=1}^{\ell} A_i \lambda_i^n$ assumes rational values for all n if and only if $A_i = A_j$ for all pairs i and j such that λ_i and λ_j are Galois conjugates.

where the penultimate line follows by Fermat's Little Theorem and the fact that each multinomial coefficient is a multiple of p. But now, since $u(1)^{p^k}$ and $u(p^k)$ are both integers, we have $u(1)^{p^k} \equiv u(p^k) \bmod p$. We conclude that for p prime, $u(p^k) = 0$ only if p divides $u(1)$. Hence if $u(1)$ is not already zero then the set of primes p such that $u(p^k) = 0$ for some $k \le c$ is finite and computable.

Unfortunately, the assumption that $A_1, \dots, A_\ell$ be rational is very restrictive. In the following section we describe our construction of a Universal Skolem Set $\mathscr{S}$, for which we can decide whether $Z(\boldsymbol{u})$ meets $\mathscr{S}$ for an arbitrary integer LRS $\boldsymbol{u}$.

B. An Infinite Universal Skolem Set

Define $f : \mathbb{N} \setminus \{0\} \to \mathbb{N}$ by $f(n) = \lfloor \sqrt{\log n} \rfloor$ and consider the sequence $(s_n)_{n \ge 0}$, defined inductively by $s_0 = 1$ and $s_n = n! + s_{f(n)}$ for $n > 0$. Our goal is to show that for any integer linear recurrence sequence $\boldsymbol{u} = (u(n))_{n \ge 0}$, one can decide whether $u(s_n) = 0$ for some $n \in \mathbb{N}$.

Note that $s_0 < s_1 < s_2 < \cdots$. We will need the following simple proposition, giving a growth bound on $s_{f(n)}$.

Proposition 6: Given $d \in \mathbb{N}$, write $N_0 := e^{16d^4}$. Then $s_{f(n)} \le n^{\frac{1}{2d}}$ for all $n \ge N_0$.

Proof We have $f(n) \le \lfloor n/2 \rfloor$ for all $n > 0$. It follows that for all $n > 0$ we have

$$\begin{aligned} s_n &\le n! + \lfloor n/2 \rfloor ! + \lfloor n/4 \rfloor ! + \cdots \\ &\le 2n! \end{aligned} \tag{9}$$

Assume now that $n \ge N_0$. Then

$$\begin{aligned} \log(s_{f(n)}) &\le \log(2 f(n)!) \quad \text{(by (9))} \\ &\le 2 f(n) \log(f(n)) \\ &\le (\log n)^{\frac{1}{2}} \log(\log n) \quad (\text{since } f(n) \le (\log n)^{\frac{1}{2}}) \\ &\le (\log n)^{\frac{3}{4}} \quad (\text{since } \log x \le x^{\frac{1}{4}} \text{ for all } x \ge 1) \\ &\le \tfrac{1}{2d} \log n \quad (\text{the assumption } n \ge e^{16d^4} \text{ yields} \\ &\qquad (\log n)^{1/4} \ge 2d). \end{aligned}$$

Exponentiating, it follows that $s_{f(n)} \le n^{\frac{1}{2d}}$ for all $n \ge N_0$. □

In the rest of this section we consider an integer sequence $\boldsymbol{u} = (u(n))_{n \ge 0}$ given by a recurrence of the form (2). From Proposition 4 we see that by scaling the sequence by a suitable integer constant, we may assume that $u(n)$ can be written in the form

$$u(n) = \sum_{i=1}^{\ell} A_i(n) \lambda_i^n , \tag{10}$$

where $\lambda_1, \dots, \lambda_\ell$ lie in the ring of integers $\mathcal{O}$ of the number field $\mathbb{K} = \mathbb{Q}(\lambda_1, \dots, \lambda_\ell)$ and $A_i(X) \in \mathcal{O}[X]$ for $i = 1, \dots, \ell$. Write $d = [\mathbb{K} : \mathbb{Q}]$ for the degree of $\mathbb{K}$ over $\mathbb{Q}$ and Δ for the discriminant of $\mathbb{K}$.

Proposition 7: For all $m, n, p \in \mathbb{N}$ such that p is a prime that does not divide Δ and $p^d \le m$, we have

$$u(n + m!) \equiv u(n) \bmod p .$$

Proof Since p does not divide Δ, the ideal $p\mathcal{O}$ splits as a product of distinct prime factors $p\mathcal{O} = \mathfrak{p}_1 \cdots \mathfrak{p}_g$. Each quotient $\mathcal{O}/\mathfrak{p}_i$ is a finite field $\mathbb{F}_{p^{f_i}}$, where $f_i \le d$ is the inertial degree of $\mathfrak{p}_i$.

Now for all $\alpha \in \mathcal{O}$ and all $i \in \{1, \dots, g\}$, we have $\alpha^{p^{f_i} - 1} \equiv 1 \bmod \mathfrak{p}_i$ and hence $\alpha^{m!} \equiv 1 \bmod \mathfrak{p}_i$, since $p^{f_i} - 1 \mid m!$. It follows that $\alpha^{m!} \equiv 1 \bmod p\mathcal{O}$. Since also $A_1[X], \dots, A_\ell[X] \in \mathcal{O}[X]$, for all $n \in \mathbb{N}$ we have

$$\begin{aligned} u(n + m!) &= \sum_{i=1}^{\ell} A_i(n + m!) \lambda_i^{n + m!} \\ &\equiv \sum_{i=1}^{\ell} A_i(n) \lambda_i^n \pmod{p\mathcal{O}} \\ &= u(n) . \end{aligned}$$

Since $u(n)$ and $u(n + m!)$ both lie in $\mathbb{Z}$, the result follows. □

Referring to the recurrence (2), and writing

$$c_1 := \max\{|u(n)| : n = 0, \dots, k-1\} \quad \text{and} \quad c_2 := |a_1| + \cdots + |a_k| ,$$

we have $|u(n)| \le c_1 c_2^n$ for all n. Let N_0 be as defined in Proposition 6, and write

$$N_1 := \max\left(N_0, 563^d, (2 \log(c_1 c_2 |\Delta|))^{2d} \right) .$$

Proposition 8: For all $n \ge N_1$ we have

$$\left| u(s_{f(n)}) \right| < \prod_{\substack{p \text{ prime} \\ p^d \le n,\, p \nmid \Delta}} p . \tag{11}$$

Proof Consider the left-hand side of (11). From the growth bound on $|u(n)|$ and Proposition 6 (which is applicable since $n \ge N_0$), we have

$$|u(s_{f(n)})| \le c_1 c_2^{s_{f(n)}} \le c_1 c_2^{n^{\frac{1}{2d}}}$$

for all $n \ge N_0$. On the other hand, the right-hand side of (11) is at least $\frac{1}{|\Delta|} e^{\frac{1}{2} n^{\frac{1}{d}}}$ since, by Theorem 4 in [16], for $m \ge 563$ the product of all primes in the interval $(0, m)$ is at least $e^{\frac{1}{2} m}$. To complete the proof, we claim that $c_1 c_2^{n^{\frac{1}{2d}}} \le \frac{1}{|\Delta|} e^{\frac{1}{2} n^{\frac{1}{d}}}$.

Indeed, the assumption $n \ge N_1$ implies $n > (2 \log(c_1 c_2 |\Delta|))^{2d}$, and hence

$$\tfrac{1}{2} n^{\frac{1}{2d}} > \log c_2 + \log(|\Delta| c_1) .$$

It follows that

$$\tfrac{1}{2} n^{\frac{1}{d}} > n^{\frac{1}{2d}} \log c_2 + \log(|\Delta| c_1) .$$

Exponentiating, we have $e^{\frac{1}{2} n^{\frac{1}{d}}} > |\Delta| c_1 c_2^{n^{\frac{1}{2d}}}$, which establishes the claim. □

We now combine Propositions 7 and 8 to obtain the following key property:

Proposition 9: For all $n \ge N_1$, if $u(s_n) = 0$ then $u(s_{f(n)}) = 0$.

Proof By Proposition 7, for every prime p such that $p^d \le n$ and $p \nmid \Delta$ we have

$$\begin{aligned} u(s_n) &= u(n! + s_{f(n)}) \\ &\equiv u(s_{f(n)}) \pmod{p} . \end{aligned}$$

The above congruence and the assumption $u(s_n) = 0$ forces $u(s_{f(n)}) \equiv 0 \pmod p$ for all primes p as above. But then by Proposition 8, this entails that $u(s_{f(n)}) = 0$. □

We can now finish the proof of our main result.

Proof (Proof of Theorem 3). We need to prove that $\mathscr{S}$ is a universal Skolem set. But from Proposition 9 we see that given an LRS $\boldsymbol{u}$, there is a computable constant N_1 such that if $\boldsymbol{u}(s_n) = 0$ for some $n \in \mathbb{N}$ then already $u(s_n) = 0$ for some $n < N_1$. Hence we can decide whether $\boldsymbol{u}$ has a zero in $\mathscr{S}$. □

C. Characterising all zeros of an LRS in $\mathscr{S}$

In this section we observe that given an LRS $\boldsymbol{u}$, one can decide whether $\{n \in \mathbb{N} : u(s_n) = 0\}$ is finite. In case $\{n \in \mathbb{N} : u(s_n) = 0\}$ is finite, we furthermore exhibit an effectively computable constant N_2 such that $u(s_n) = 0$ only if $n \leq N_2$.

Let $\boldsymbol{u} = (u(n))_{\geq 0}$ be a given integer LRS, satisfying an order-k recurrence. As shown in Section II-B, one can compute M such that for $0 \leq j < M$ each subsequence $\boldsymbol{u}_{M,j} := (u(Mn + j))_{n \geq 0}$ is non-degenerate and satisfies an order-k recurrence.

Now $\{n \in \mathbb{N} : u(s_n) = 0\}$ can only be infinite if one of the subsequences $\boldsymbol{u}_{M,j}$, for some $0 \leq j < M$, is identically zero and $\{n \in \mathbb{N} : s_n \equiv j \bmod M\}$ is infinite. But it is easy to test zeroness of $\boldsymbol{u}_{M,j}$—just determine whether the first k terms are all zero. It is also straightforward to determine whether $\{n \in \mathbb{N} : s_n \equiv j \bmod M\}$ is infinite. Indeed for all $n \geq M$ we have $s_n \equiv s_{f(n)} \bmod M$ and so, since the map f is surjective, $\{n \in \mathbb{N} : s_n \equiv j \bmod M\}$ is infinite if and only if there exists $n \leq f(M)$ with $s_n \equiv j \bmod M$. We conclude that we can decide infiniteness of $\{n \in \mathbb{N} : u(s_n) = 0\}$.

Suppose now that $\{n \in \mathbb{N} : u(s_n) = 0\}$ is finite; we seek an upper bound N_2 on this set. The discussion in the preceding paragraph shows that s_n belongs to some infinite arithmetic progression of zeros of $\boldsymbol{u}$ only if $n \leq f(M)$. Furthermore, Schlickewei and Schmidt [17] show that a non-zero integer sequence satisfying a non-degenerate order-k recurrence has at most $(2k)^{35k^3}$ zeros. Thus $\#\{n \in \mathbb{N} : u(s_n) = 0\} \leq T$, where

$$T := M(2k)^{35k^3} + f(M) .$$

Now define N_2 to be the least positive integer such that $f^{(T)}(N_2) > N_1$. Suppose for a contradiction that $u(s_n) = 0$ for some $n \geq N_2$. Then, repeatedly applying Proposition 9, we have that

$$u(s_{f(n)}) = 0, u(s_{f^{(2)}(n)}) = 0, \ldots, u(s_{f^{(T)}(n)}) = 0 ,$$

contradicting the fact that $\#\{n \in \mathbb{N} : u(s_n) = 0\} \leq T$. We conclude that N_2 is an upper bound for $\{n \in \mathbb{N} : u(s_n) = 0\}$.

IV. Conclusion

We have introduced the notion of a Universal Skolem Set and given an example of such a set. The latter was enumerated by a sequence $(s_n)_{n \geq 0}$ such that $n! \leq s_n \leq 2n!$. From these bounds it follows, e.g., using Stirling's approximation, that

$$\#\{n : s_n \leq T\} = \Theta\left(\frac{\log T}{\log\log T}\right) .$$

It is natural to ask whether one can construct a Universal Skolem Set of greater asymptotic density. Since the decidability of the Skolem Problem is equivalent to the assertion that $\mathbb{N}$ is itself a Universal Skolem Set, one may view the present paper as offering an alternative line of attack on the Skolem Problem.

The notion of Universal Skolem Set can be generalised to the notion of a *Skolem Set $\mathscr{S}$ for a class $\mathscr{L}$ of linear recurrence sequences*: a subset $\mathscr{S}$ of the positive integers such that there is an effective procedure that inputs a linear recurrence sequence $(u(n))_{n \geq 0}$ in $\mathscr{L}$ and decides whether $u(n) = 0$ for some $n \in \mathscr{S}$. A Universal Skolem Set is thus a Skolem Set for the class of all linear recurrence sequences, the set of prime numbers is a Skolem Set for the class of linear recurrences considered in Section III-A, and, by the results of [12], [18], the set of all positive integers is a Skolem Set for recurrence sequences of order at most 4. This more general notion of Skolem Set provides a setting in which one can study the tradeoffs between the density of a Skolem Set and the generality of the class of linear recurrence sequences to which it applies.

Acknowledgements

This work was carried out during a visit of Florian Luca at the Max Planck Institute for Software Systems in Saarbrücken, Germany from September 2020 to March 2021. He thanks the institution for its hospitality and excellent working conditions.

Joël Ouaknine was supported by ERC grant AVS-ISS (648701) and DFG grant 389792660 as part of TRR 248 (see `https://perspicuous-computing.science`).

James Worrell was supported by EPSRC Fellowship EP/N008197/1.

References

[1] Yu. Bilu. "A note on universal Hilbert sets", *J. Reine Angew. Math.*, **479**, 195–203 (1996).

[2] J. Berstel and C. Reutenauer. "Noncommutative Rational Series with Applications" *Encyclopedia of Mathematics and its Applications*, **137**, Cambridge University Press (2011).

[3] V. Blondel and J. Tsitsiklis. "A survey of computational complexity results in systems and control", *Automatica* **36**(9), 1249–1274 (2000).

[4] S. D. Cohen. "The distribution of Galois groups and Hilbert's irreducibility theorem", *Proc. London Math. Soc.*, **43**(3), 227– 250 (1981).

[5] H. Derksen. "A Skolem-Mahler-Lech Theorem in Positive Characteristic and Finite Automata", *Inventiones Mathematicae*, **168**, 175–224 (2007).

[6] G. Everest, A. van der Poorten, Igor Shparlinksi, Thomas Ward. "Recurrence Sequences", Mathematical Surveys and Monographs, **104**, AMS (2003).

[7] R. P. Flowe and A. G. Harris. "A Note on Generalized Vandermonde Determinants", *SIAM J. Matrix Anal. Appl.*, **14**, 1146–1151 (1993).

[8] N. Fijalkow, J. Ouaknine, A. Pouly, J. Sousa Pinto and J. Worrell. "On the decidability of reachability in linear time-invariant systems", *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control (HSCC)*, 77–86, ACM (2019).

[9] M. Filaseta and R. Wilcox. "An explicit dense universal Hilbert set", *Math. Proc. Cambridge Philos. Soc.,* **167**, 531–547 (2019).

[10] G. Kenison, R. Lipton, J. Ouaknine and J. Worrell. "On the Skolem problem and prime powers", *Proceedings of the 45th International Symposium on Symbolic and Algebraic Computing (ISSAC)*, 289–296, ACM (2019).

[11] F. Luca. "Exponential Diophantine equations", Notes from the International Autumn School on Computational Number Theory, Tutor. Sch. Workshops Math. Sci., Birkhäuser/Springer, Cham (2019).

[12] M. Mignotte, T. N. Shorey, and R. Tijdeman. "The distance between terms of an algebraic recurrence seqeunce", *Journal für die Reine und Angewandte Mathematik,* **349**, 63—76 (1984).

[13] J. Ouaknine and J. Worrell. "On Linear Recurrence Sequences and Loop Termination", ACM Siglog News, **2**(2), 4–13 (2015).

[14] J. Piribauer and C. Baier. "On Skolem-Hardness and Saturation Points in Markov Decision Processes", *Proceedings of the 47th International Colloquium on Automata, Languages, and Programming, (ICALP),* LIPIcs **168**, 1–17 (2020).

[15] G. Rozenberg and A. Salomaa. "Cornerstones of Undecidability", Prentice Hall (1994).

[16] J. B. Rosser and L. Schoenfeld. "Approximate formulas for some functions of prime numbers", *Illinois J. Math.,* **6**, 64–94 (1962).

[17] H. P. Schlickewei and W. M. Schmidt. "The number of solutions of polynomial-exponential equation", *Compositio Math.,* **120**, 193–225 (2000).

[18] N. K. Vereshchagin. "Occurrence of a zero in a linear recurrence sequence", *Mat. Zametki,* **38**(2), 177–189 (1985).