A Universal Skolem Set of Positive Lower Density

Florian Luca
School of Mathematics, University of the Witwatersrand, Johannesburg, South Africa
Research Group in Algebraic Structures & Applications, King Abdulaziz University, Saudi Arabia
Centro de Ciencias Matemáticas UNAM, Morelia, Mexico
Max Planck Institute for Software Systems, Saarland Informatics Campus, Germany

Joël Ouaknine
Max Planck Institute for Software Systems, Saarland Informatics Campus, Germany

James Worrell
Department of Computer Science, University of Oxford, UK

Abstract

The Skolem Problem asks to decide whether a given integer linear recurrence sequence (LRS) has a zero term. Decidability of this problem has been open for many decades, with little progress since the 1980s. Recently, a new approach was initiated via the notion of a Skolem set – a set of positive integers relative to which the Skolem Problem is decidable. More precisely, $\mathcal{S}$ is a Skolem set for a class $\mathcal{L}$ of integer LRS if there is an effective procedure that, given an LRS in $\mathcal{L}$, decides whether the sequence has a zero in $\mathcal{S}$. A recent work exhibited a Skolem set for the class of all LRS that, while infinite, had density zero. In the present work we construct a Skolem set of positive lower density for the class of simple LRS.

2012 ACM Subject Classification Computing methodologies → Symbolic and algebraic algorithms; Computing methodologies → Number theory algorithms

Keywords and phrases Linear Recurrence Sequences, Skolem Problem, Exponential Diophantine Equations, Sieve Methods

Digital Object Identifier 10.4230/LIPIcs.MFCS.2022.73

Funding *Joël Ouaknine*: Also affiliated with Keble College, Oxford as `emmy.network` Fellow, and supported by DFG grant 389792660 as part of TRR 248 (see `https://perspicuous-computing.science`).

1 Introduction

An (integer) linear recurrence sequence (LRS) $\langle u_n \rangle_{n=0}^{\infty}$ is a sequence of integers satisfying a recurrence of the form

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

where the coefficients $a_1, \ldots, a_k$ are integers. The celebrated theorem of Skolem, Mahler, and Lech [23, 16, 14] describes the set of zero terms of such a recurrence:

▶ **Theorem 1.** *Given an integer linear recurrence sequence $\langle u_n \rangle_{n=0}^{\infty}$, the set $\{n \in \mathbb{N} : u_n = 0\}$ 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 the merge 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

47th International Symposium on Mathematical Foundations of Computer Science (MFCS 2022).
Editors: Stefan Szeider, Robert Ganian, and Alexandra Silva; Article No. 73; pp. 73:1–73:12
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

of zeros of a given non-degenerate linear recurrence sequence. It is readily seen that 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. We refer to [5, Chapter 6] and [24, Chapter X] for expository accounts of the Skolem-Mahler-Lech Theorem and discussion of the ineffectiveness of known proofs.

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 [18, 26]. However for LRS of order 5 it is not currently known how to decide Skolem's Problem. For a (highly restricted) subclass of LRS, the paper [1] obtains nearly matching complexity lower and upper bounds for the problem.

In computer science, Skolem's Problem lies at the heart of key decision problems in formal power series [21, 4], stochastic model checking [20], control theory [6, 10], and loop termination [19]. The problem is also closely related to membership problems on commutative matrix groups and semigroups, as considered in [7, 13]. We note that in several of the above-mentioned citations, the Skolem Problem is used as a reference to show hardness of other open decision problems.

A recent paper [15] initiated a new approach to the decidability of Skolem's Problem. Rather than place syntactic 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, [15] introduced the following definition.

▶ **Definition 2.** *An infinite set $\mathcal{S} \subseteq \mathbb{N}$ is a* Skolem set *for a class $\mathcal{L}$ of LRS if there is an effective procedure that, given an LRS $\langle u_n \rangle_{n=0}^{\infty}$ in $\mathcal{L}$, determines whether there exists $n \in \mathcal{S}$ with $u_n = 0$.*

The main technical contribution of [15] was to exhibit a Skolem set for the class of all LRS. Such sets are called *Universal Skolem sets.* Specifically, define $f : \mathbb{N} \setminus \{0\} \to \mathbb{N}$ by $f(n) = \lfloor \sqrt{\log n} \rfloor$, and inductively define the sequence $\langle s_n \rangle_{n=0}^{\infty}$ by $s_0 = 1$ and $s_n = n! + s_{f(n)}$ for $n > 0$. Then $\{s_n : n \in \mathbb{N}\}$ is a Universal Skolem set. This construction yields a very sparse set, which has density zero. This leads to the question of whether one can construct Skolem sets of positive lower density. Decidability of the Skolem Problem is, by definition, equivalent to the assertion that $\mathbb{N}$ is a Universal Skolem set (in fact, decidability follows already from the existence of Universal Skolem set that contains an infinite arithmetic progression). Hence seeking Skolem sets of increasingly higher density is a natural direction in which to make progress on the Skolem Problem.

The main result of the present paper exhibits a set $\mathcal{S}$ of positive lower density, i.e., having

$$\liminf_{n\to\infty} \frac{|\mathcal{S} \cap \{1, \ldots, n\}|}{n} > 0 \,,$$

such that $\mathcal{S}$ is a Skolem set for the class of simple LRS. For short we call $\mathcal{S}$ a *Simple Universal Skolem set.* Recall that a simple LRS is one for which the characteristic roots of its minimal-order recurrence are simple. The Skolem Problem for simple LRS is open already for LRS of order 5.

The construction of $\mathcal{S}$ is fundamentally different from the example in [15]. Roughly speaking, our set consists of positive integers n that admit many representations of the form $n = Pq + a$, with P, q prime and q, a logarithmic in n. The formal definition of the set is given in Section 3. In Section 4 we show that one can decide, given a simple LRS

$\langle u_n \rangle_{n=0}^{\infty}$, whether there exists $n \in \mathcal{S}$ with $u_n = 0$. The crucial ingredients here are results of Schlickewei and Schmidt [22] that give explicit upper bounds on the number of solutions of certain exponential Diophantine equations. Such results have previously been used to give effective upper bounds on the *number* of zeros of a given non-degenerate LRS (see [8, Theorem 2.7]), thereby strengthening the statement of the Skolem-Mahler-Lech Theorem (which, recall, asserts mere finiteness of the number of zeros). However such bounds do not obviously yield a solution to the Skolem Problem itself, which would instead require effective bounds on the *magnitude* of the zeros of an LRS. The essential novelty of our approach is, via the notion of representation, to leverage bounds on the number of solutions of exponential Diophantine equations to obtain bounds on the magnitude of the zeros of a simple LRS that lie in $\mathcal{S}$. In Section 5, we show that the set $\mathcal{S}$ has positive lower density. Here, classical number-theoretic techniques for upper bounding the number of pairs of primes in certain linear relations (such as twin primes) play the main role.

As discussed in the Conclusion, an extended version of this paper will show that for the set $\mathcal{S}$ introduced in this paper, the Skolem Problem is decidable relative to $\mathcal{S}$ for the class of all LRS, not just the simple ones. In the Conclusion we also briefly discuss the prospects of refining our analysis of the density of $\mathcal{S}$.

2 Background

In this section we briefly recall some relevant notation and definitions concerning number fields. We also recall a result from [9, 25] on unit equations that is derived from the Subspace Theorem and that will play a crucial role in our construction of a Simple Universal Skolem set.

Throughout we use the Vinogradov notation $f \ll g$ for $f \in O(g)$.

Recall that a *number field* $\mathbb{K}$ is a subfield of $\mathbb{C}$ that is finite dimensional as a vector space over $\mathbb{Q}$. The subring of algebraic integers in $\mathbb{K}$ is denoted $\mathcal{O}_{\mathbb{K}}$. For such a field $\mathbb{K}$, we denote by $\mathrm{Gal}(\mathbb{K}/\mathbb{Q})$ the group of automorphisms of $\mathbb{K}$. Given $\alpha \in \mathbb{K}$, the norm of α is defined by

$$N_{\mathbb{K}/\mathbb{Q}}(\alpha) = \prod_{\sigma \in \mathrm{Gal}(\mathbb{K}/\mathbb{Q})} \sigma(\alpha)\,.$$

The *norm* $N_{\mathbb{K}/\mathbb{Q}}(\alpha)$ is rational for all $\alpha \in \mathbb{K}$; moreover $N_{\mathbb{K}/\mathbb{Q}}(\alpha)$ is an integer if $\alpha \in \mathcal{O}_{\mathbb{K}}$. Clearly we have $|N(\alpha)| < M^{d_{\mathbb{K}}}$, where $d_{\mathbb{K}}$ is the degree of $\mathbb{K}$ and

$$M := \max_{\sigma \in \mathrm{Gal}(\mathbb{K}/\mathbb{Q})} |\sigma(\alpha)|$$

is the *house* of α.

We say that $\alpha, \beta \in \mathbb{K}$ are *multiplicatively dependent* if there exist integers k, ℓ, not both zero, such that $\alpha^k = \beta^\ell$. Observe that if $\alpha \in \mathbb{K}$ is not a root of unity then given $\sigma \in \mathrm{Gal}(\mathbb{K}/\mathbb{Q})$, every multiplicative relations $\alpha^k = \sigma(\alpha)^\ell$ is such that $k = \pm\ell$. Indeed, repeatedly applying σ to this relation we deduce that $\alpha^{k^d} = (\sigma^d(\alpha))^{\ell^d}$ for all $d \geq 1$. In particular, choosing d to be the order of σ we get that $\alpha^{k^d} = \alpha^{\ell^d}$ and hence $k = \pm\ell$.

We recall that every ideal in $\mathcal{O}_{\mathbb{K}}$ can be written uniquely as the product of prime ideals. Given a rational prime $p \in \mathbb{Z}$, we say that a prime ideal $\mathfrak{p}$ lies above p if $\mathfrak{p}$ is a factor of $p\mathcal{O}_{\mathbb{K}}$. In this case we have that $p \mid N_{\mathbb{K}/\mathbb{Q}}(\alpha)$ for all $\alpha \in \mathfrak{p}$.

We will need a result of Schlickewei and Schmidt [22] that gives upper bounds on the number of integer solutions of certain exponential Diophantine equations. The result (which we have specialised to our setting) is as follows:

▶ **Theorem 3** ([22, Theorem 1]). *For $i = 1, \ldots, \ell$, let α_i, β_i, C_i be non-zero and lie in a number field of degree d over $\mathbb{Q}$. Suppose that the system of equations*

$$\alpha_i^{z_1}\beta_i^{z_2} = \alpha_j^{z_1}\beta_j^{z_2} \quad i, j \in \{1, \ldots, \ell\}$$

has no non-zero solution in integers z_1, z_2. Then the number of solutions of the equation

$$\sum_{i=1}^{\ell} C_i \alpha_i^{x_1}\beta_i^{x_2} = 0 \tag{2}$$

in integers x_1, x_2 for which no proper sub-sum of the left-hand side vanishes is at most $2^{35\ell^2} d^{6\ell^2}$.

3 The Definition of the Set $\mathcal{S}$

In this section we give the definition of our Simple Universal Skolem set $\mathcal{S}$.

For a positive real number $x > 0$, denote by $\log x$ the natural logarithm of x. For a positive integer $k \geq 1$, we inductively define the iterated logarithm function $\log_k x$ as follows: $\log_1 x := \log x$, and for $k \geq 2$ we set $\log_k x := \max\{1, \log_{k-1}(\log x)\}$. Thus, for x sufficiently large, $\log_k x$ is the k-fold iterate of log applied to x. We omit the subscript when $k = 1$.

Fix a positive integer parameter X. We define disjoint intervals

$$A(X) := \left[\log_2 X, \sqrt{\log X}\right] \quad \text{and} \quad B(X) := \left[\frac{\log X}{\sqrt{\log_3 X}}, \frac{2\log X}{\sqrt{\log_3 X}}\right].$$

We further define a *representation* of an integer $n \in [X, 2X]$ to be a triple (q, P, a) such that $q \in A(X)$, $a \in B(X)$, P and q are prime, and $n = Pq + a$. We say that two representations (q, P, a) and (q', P', a') of the same number *overlap* if either $a + q = a' + q'$ or $a - q = a' - q'$. It is clear that two overlapping representations $(q, P, a) \neq (q', P', a')$ must have both $a \neq a'$ and $q \neq q'$.

We denote by $r(n)$ the number of representations of n. Finally we put

$$\mathcal{S}(X) := \{n \in [X, 2X] \ : \ r(n) > \log_4 X \text{ and no two representations of } n \text{ overlap}\}$$

and we define

$$\mathcal{S} := \bigcup_{X \geq 1} \mathcal{S}(X).$$

This completes the definition of the set $\mathcal{S}$. The rest of the paper is devoted to showing that $\mathcal{S}$ is a Simple Universal Skolem set and that it has positive lower density.

4 Solving the Simple Skolem Problem Relative to $\mathcal{S}$

The following result is the first half of the argument that the set $\mathcal{S}$, defined in Section 3, is a Simple Universal Skolem set. We use the bounds on the solutions of exponential Diophantine equations stated in Section 2 to show that the Skolem Problem for simple LRS is decidable relative to $\mathcal{S}$.

▶ **Proposition 4.** *Given a non-degenerate simple linear recurrence sequence $\langle u_n \rangle_{n=0}^{\infty}$, there is an effectively computable upper bound on the set $\{n \in \mathcal{S} : u(n) = 0\}$.*

Proof. Fix a non-degenerate simple linear recurrence sequence $\langle u_n \rangle_{n=0}^{\infty}$ with distinct characteristic roots $\alpha_1, \dots, \alpha_\ell$. Recall that non-degeneracy is the condition that α_i/α_j is not a root of unity for all $i \neq j$. It is well-known that u_n admits an exponential-sum representation

$$u_n = \sum_{i=1}^{\ell} C_i \alpha_i^n ,$$

where the constants C_i lie in the number field $\mathbb{K} := \mathbb{Q}(\alpha_1, \dots, \alpha_\ell)$. Multiplying the sequence $\langle u_n \rangle_{n=0}^{\infty}$ by a suitable integer, we may assume without loss of generality that the C_i lie in the ring of integers $\mathcal{O}_\mathbb{K}$.

Suppose that $n \in \mathcal{S}$ is such that $u_n = 0$. By the definition of $\mathcal{S}$ we have that $n \in \mathcal{S}(X)$ for some positive integer X. We show that there is an upper bound on X that is effectively computable from the description of the sequence $\langle u_n \rangle_{n=0}^{\infty}$. Since $n \leq 2X$ this gives the desired effective upper bound on n.

Fix $n \in \mathcal{S}(X)$ such that $u_n = 0$ and consider a representation (q, P, a) of n. Then $n = qP + a \geq X$, and so

$$P \geq \frac{X - a}{q} \geq \frac{X - \log X}{\sqrt{\log X}} > \sqrt{X} \tag{3}$$

for X sufficiently large.

Let $\mathfrak{p}$ be a prime ideal of $\mathcal{O}_\mathbb{K}$ lying above P and let $\sigma \in \mathrm{Gal}(\mathbb{K}/\mathbb{Q})$ be a Frobenius automorphism corresponding to $\mathfrak{p}$, that is, such that $\sigma(\alpha) \equiv \alpha^P \bmod \mathfrak{p}$ for all $\alpha \in \mathcal{O}_\mathbb{K}$. From $u_n = 0$ and $n = qP + a$ we have

$$\sum_{i=1}^{\ell} C_i \alpha_i^{qP+a} = 0 .$$

Since $\sigma(\alpha_i) \equiv \alpha_i^P \bmod \mathfrak{p}$ for all $i \in \{1, \dots, \ell\}$, we can write

$$\sum_{i=1}^{\ell} C_i \sigma(\alpha_i)^q \alpha_i^a \equiv 0 \bmod \mathfrak{p} . \tag{4}$$

Taking norms, we see that P divides

$$N := N_{\mathbb{K}/\mathbb{Q}} \left(\sum_{i=1}^{\ell} C_i \sigma(\alpha_i)^q \alpha_i^a \right) .$$

Moreover, we have the inequality

$$|N| \leq M^{(1+q+a)d_\mathbb{K}} , \tag{5}$$

where $M = \max\{|\alpha_i|, |\beta_i|, |C_i| : 1 \leq i \leq \ell\}$ and $d_\mathbb{K}$ is the degree of $\mathbb{K}$ over $\mathbb{Q}$.

From the fact that $q \in A(X)$ and $a \in B(X)$, we see that for X sufficiently large we have $a, q < \frac{\log X}{8 d_\mathbb{K} \log M}$. In this case, Equation (5) yields $N \ll X^{1/4}$. Hence for X large enough, using Equation (3), we have $N < X^{1/2} < P$. But N is an integer that is divisible by P, so it must be zero. We conclude that, for sufficiently large X, the left-hand side of (4) is zero; that is,

$$\sum_{i=1}^{\ell} C_i \sigma(\alpha_i)^q \alpha_i^a = 0 . \tag{6}$$

Carrying over the terminology for representations, let us say that two different solutions q, a and q', a' of Equation (6) *overlap* if either $a + q = a' + q'$ or $a - q = a' - q'$. Our goal is to give an explicit upper bound on the size of any collection of pairwise non-overlapping solutions of (6). For this, it is enough to give an upper bound under the additional assumption that no vanishing proper sub-sum of the left-hand side of (6) vanishes. Multiplying the latter quantity by the number (at most 2^ℓ) of possible sub-sums gives the desired upper bound on the total number of non-overlapping solutions.

Consider an instance of Equation (6) for which no proper sub-sum vanishes. Let $\mathcal{G}$ be the additive group of vectors $(z_1, z_2) \in \mathbb{Z}^2$ such that

$$\sigma(\alpha_i)^{z_1}\alpha_i^{z_2} = \sigma(\alpha_j)^{z_1}\alpha_j^{z_2}$$

for all $1 \le i < j \le \ell$. Note that there exist effective upper bounds on the magnitude of the entries a basis of $\mathcal{G}$ [17].

For $(z_1, z_2) \in \mathcal{G}$ we have $\sigma(\alpha_i/\alpha_j)^{z_1} = (\alpha_i/\alpha_j)^{-z_2}$. As shown in Section 2, since α_i/α_j is not a root of unity we must have either $z_1 = z_2$ or $z_1 = -z_2$. We thus have the following three possibilities for the form of the group $\mathcal{G}$ – either $\mathcal{G} = \{\mathbf{0}\}$, $\mathcal{G} = \{(z, z) : z \in m\mathbb{Z}\}$, or $\mathcal{G} = \{(z, -z) : z \in m\mathbb{Z}\}$, where $m \in \mathbb{Z}$. We consider three cases according to these three eventualities.

Case (i): $\mathcal{G} = \{\mathbf{0}\}$. Applying Theorem 3, the total number of solutions (overlapping or not) of Equation (6) in this case is at most $2^{35\ell^2}(d_{\mathbb{K}})^{6\ell^2}$.

Case (ii): $\mathcal{G} = \{(z, z) : z \in m\mathbb{Z}\}$ for some $m \in \mathbb{Z}$. In this case we have that $\sigma(\alpha_i)^m\alpha_i^m$ takes the same value for all $i \in \{1, \ldots, \ell\}$. Dividing Equation (6) by the $\lfloor q/m \rfloor$-th power of this common value we get

$$\sum_{i=1}^{\ell} \widetilde{C}_i \alpha_i^{a-q} = 0\,,$$

where each constant $\widetilde{C}_i$ is uniquely determined by C_i and the residue of q modulo m.

In other words, for every solution (q, a) of Equation (6), we have that $q - a$ is a zero of one of a finite number (at most m^ℓ) of non-degenerate LRS, each of which takes values in $\mathbb{K}$ and has order at most k. Applying Theorem 3, the number of different values of $q - a$ over all solutions (q, a) is at most $(2k)^{35k^2}(d_{\mathbb{K}})^{6k^2}m^\ell$. But the latter quantity is then a bound on the cardinality of a set of pairwise non-overlapping solutions of Equation (6).

Case (iii): $\mathcal{G} = \{(z, -z) : z \in m\mathbb{Z}\}$ for some $m \in \mathbb{Z}$. The argumentation is almost exactly as in Case (ii). We have that $\sigma(\alpha_i)^m\alpha_i^{-m}$ takes the same value for all $i \in \{1, \ldots, \ell\}$. Dividing Equation (6) by the $\lfloor q/m \rfloor$-th power of this common value we get

$$\sum_{i=1}^{\ell} \widetilde{C}_i \alpha_i^{a+q} = 0$$

where each constant $\widetilde{C}_i$ is uniquely determined by C_i and the residue of q modulo m. The argument now follows exactly as in Case (ii). In particular, we get the same upper bound on the number of solutions under Case (iii) as under Case (ii).

We can now summarise and wrap up. If $n \in \mathcal{S}(X)$ is a zero of u_n then for every representation $n = qP + a$ it holds that (q, a) is a solution of Equation (6). Moreover, since $n \in \mathcal{S}(X)$, no two representations of n are overlapping. Since we have an effective upper bound on the cardinality of a set of pairwise non-overlapping solutions of (6), we get an effective upper bound (that does not depend on n) for the number of representations of n. Finally, since by the definition of $\mathcal{S}(X)$ the number of representations is at least $\log_4 X$, we obtain the desired upper bound on X. ◀

5 The Set $\mathcal{S}$ has Positive Lower Density

Our goal in this section is to show that the set $\mathcal{S}$ has positive lower density. The key tool is the following result, (see [11, Chapter 2.6, Theorem 2.3]), derived using Selburg's upper-bound sieve, that bounds from above the number of times that two linear forms simultaneously take prime values.

▶ **Theorem 5.** *Let a_1, a_2, b_1, b_2 be integers such that*

$$E := |a_1a_2(a_1b_2 - a_2b_1)|$$

is non-zero. Then

$$|\{t \leq X : a_1t + b_1, a_2t + b_2 \text{ both prime}\}| \ll \frac{X}{(\log X)^2}\frac{E}{\varphi(E)},$$

where the implied constant is independent of a_1, a_2, b_1, b_2, and φ denotes Euler's totient function.

To help illustrate this result, observe that in case $a_1 = a_2 = 1$, $b_1 = 0$, and $b_2 = 2$, Theorem 5 gives an upper bound on the number of twin primes less than X.

We will also need the following straightforward proposition. (Note that here and in the rest of this section, variables p and q always range over prime numbers.)

▶ **Proposition 6.** $\sum_{q\in A(X)} \frac{1}{q} \sim \log_3 X$

Proof. We use the fact (see [3, Theorem 13.6]) that there exists an absolute constant c such that

$$\sum_{p\leq t} \frac{1}{p} = \log\log t + c + O\left(\frac{1}{\log t}\right)$$

for all $t \geq 2$. Using this fact, we have

$$\begin{aligned}\sum_{q\in A(X)} \frac{1}{q} &= \log\log\sqrt{\log X} - \log\log\log_2 X + o(1)\\ &= \log_3 X - \log_4 X + o(1)\\ &\sim \log_3 X\end{aligned}$$

for large enough X. ◀

We further note the following useful fact concerning the Euler function (see [3, Exercise 2.10(xii)]):

$$\sum_{k=1}^{n} \frac{k}{\varphi(k)} \ll n\,. \tag{7}$$

The rest of this section is devoted to a proof of the following result:

▶ **Theorem 7.** *The set $\mathcal{S}$, defined in Section 3, has strictly positive lower density.*

Recall from Section 3 that $\mathcal{S} := \bigcup_{X\geq 1} \mathcal{S}(X)$, where $\mathcal{S}(X) \subseteq [X, 2X]$. Hence, to prove that $\mathcal{S}$ has positive lower density, it suffices to show that $|\mathcal{S}(X)| \gg X$ for X sufficiently large. We first argue that for sufficiently large X we have

$$|\{n \in [X, 2X] : r(n) > \log_4 X\}| \gg X\,. \tag{8}$$

After that, we deal with those n that have overlapping representations (cf. the definition of $\mathcal{S}(X)$ in Section 3).

To prove (8) we use the moment method. To set this up, let

$$S_i := \sum_{\substack{n\in[X,2X]\\ r(n)>\log_4 X}} r(n)^i \qquad \text{for } i \in \{0,1,2\}\,.$$

The inequality (8) is equivalent to the assertion that $S_0 \gg X$. Thinking of S_1 as the inner product of the vector

$$\langle r(n) : n \in [X,2X], r(n) > \log_4 X\rangle$$

and the constant all 1's vector, applying the Cauchy-Schwartz inequality we get that

$$S_2 S_0 \geq S_1^2.$$

To show that $S_0 \gg X$ we will use a lower bound on S_1 and an upper bound on S_2.

5.1 Lower bound on S_1

In this section we show that $S_1 \gg X\sqrt{\log_3 X}$.

Fix $q \in A(X)$. Then for $P \in \left[\frac{X}{q}, \frac{1.5X}{q}\right]$ we have that $qP \in [X, 1.5X]$. Furthermore, if $a \in B(X)$, then for sufficiently large X we have $a < \log X < 0.5X$, so that $qP + a \in [X, 2X]$. Thus, for large X and each fixed prime $q \in A(X)$, the number of representations (q, P, a) with $qP + a \in [X, 2X]$ is at least the product of the number of primes $P \in \left[\frac{X}{q}, \frac{1.5X}{q}\right]$ and the cardinality of the set $B(X)$. But, for large enough X, the number of primes $P \in \left[\frac{X}{q}, \frac{1.5X}{q}\right]$ is

$$\pi\left(\frac{1.5X}{q}\right) - \pi\left(\frac{X}{q}\right) > \frac{0.3X}{q\log X}\,.$$

Thus, for fixed q, the number of representations (q, P, a) with $qP + a \in [X, 2X]$ is at least

$$\frac{0.3X}{q\log X}\,\frac{\log X}{\sqrt{\log_3 X}} \gg \frac{X}{q\sqrt{\log_3 X}}\,, \tag{9}$$

where the implied constant is independent of q.

Summing the lower bound (9) over $q \in A(X)$, we have

$$\begin{aligned}\sum_{n\in[X,2X]} r(n) \quad &\gg \quad \frac{X}{\sqrt{\log_3 X}}\left(\sum_{q\in A(X)} \frac{1}{q}\right)\\ &\gg \quad X\sqrt{\log_3 X} \quad \text{(by Proposition 6).}\end{aligned}$$

Finally, in order to get a lower bound on S_1 we must remove from the left-hand side above, the summands $r(n)$ for which $r(n) \leq \log_4 X$. But the contribution of these to the total sum is $O(X\log_4 X) = o(X\sqrt{\log_3 X})$, and so we conclude that $S_1 \gg X\sqrt{\log_3 X}$.

For future reference, we observe that essentially the same argument shows that $S_1 \ll X\sqrt{\log_3 X}$. Indeed, for each fixed $q \in A(X)$, the number of primes P such that $qP \in [X, 2X]$ is $\ll \frac{X}{q\log X}$. Since the number of $a \in B(X)$ is at most $\frac{\log X}{\sqrt{\log_3 X}}$, the number of representations (q, P, a) such that $qP + a \in [X, 2X]$ is $\ll \frac{X}{q\sqrt{\log_3 X}}$. Now, summing over $q \in A(X)$ and using Proposition 6, we obtain the bound $S_1 \ll X\sqrt{\log_3 X}$.

5.2 Upper bound on S_2

Our goal is to show that $S_2 \ll X \log_3 X$. To this end we consider S_2 as the number of pairs of representations $(q, P, a), (q', P', a')$ such that $qP + a = q'P' + a'$, with the common sum lying in the interval $[X, 2X]$. We break the count down into three cases.

Case (i). Let us first consider the number of such pairs with $a = a'$. In this case we have $qP = q'P'$. But then, since q and q' are small and P and P' are large, we get that $q = q'$ and $P = P'$, and hence $(q, P, a) = (q', P', a')$. Thus the number of such pairs is equal to S_1. But, as noted at the conclusion of Section 5.1, we have $S_1 \ll X\sqrt{\log_3 X}$.

Case (ii). We next consider the number of pairs of representations with $a \neq a'$ and $q = q'$. In this case we have

$$P - P' = m, \qquad \text{where} \qquad m := \frac{a - a'}{q}.$$

By Theorem 5, for each fixed m the number of primes $P \leq X/q$ such that $P + m$ is also prime is

$$\ll \frac{X}{q(\log X)^2} \frac{m}{\varphi(m)}.$$

Furthermore, for fixed q and m, the number of choices of a, a' with $m = \frac{a-a'}{q}$ is at most the number of choices of a, that is, at most $\frac{2 \log X}{\sqrt{\log_3 X}}$. Thus, for fixed q and m, the total number of pairs of representations $(q, P, a), (q', P', a')$ with $qP + a = q'P' + a' \in [X, 2X]$ and $m = \frac{a-a'}{q}$ is

$$\begin{aligned} &\ll \frac{X}{q(\log X)^2} \frac{m}{\varphi(m)} \frac{2 \log X}{\sqrt{\log_3 X}} \\ &= \frac{2X}{q(\log X)\sqrt{\log_3 X}} \frac{m}{\varphi(m)}. \end{aligned}$$

From the fact that $a, a' \in A(X)$ we have $m \leq \frac{2 \log X}{q\sqrt{\log_3 X}}$. Summing up over all values of m, we get that the total number of solutions for fixed q is

$$\begin{aligned} &\ll \frac{2X}{q(\log X)\sqrt{\log_3 X}} \left(\sum_{m \leq \frac{2 \log X}{q\sqrt{\log_3 X}}} \frac{m}{\varphi(m)} \right) \\ &\ll \frac{X}{q^2 \log_3 X}. \end{aligned}$$

In the above, we used the fact (see Equation (7)) that $\sum_{m \leq t} m/\varphi(m) \ll t$.

Now we sum up over $q > \log_2 X$, getting that the number of such solutions is at most

$$\frac{X}{\log_3 X} \sum_{q > \log_2 X} \frac{1}{q^2} \ll \frac{X}{(\log_2 X)\log_3 X} = o(X).$$

Case (iii). Finally, let us count the rest of the solutions; namely the ones for which $a \neq a'$ and $q \neq q'$. Fixing a, a', q, q' we have

$$qP - q'P' = a' - a.$$

The general solution of the above equation in integers P and P' can be written in the form $P = p_0 + q't$ and $P' = p_0' + qt$, where t is an integer parameter and p_0, p_0' is a particular solution, chosen to be minimal among positive integer solutions (simultaneously, in both

coordinates). The condition that $Pq + a \le 2X$ implies that $P \le 2X/q$ and hence that $t \le \frac{2X}{qq'}$.
Using the assumption $a \neq a'$, we can apply Theorem 5 to deduce that the number of $t \le \frac{2X}{qq'}$ such that both $p_0 + q't$ and $p_0' + qt$ are prime is

$$\ll \frac{X}{qq'(\log X)^2} \frac{|a-a'|}{\varphi(|a-a'|)} .$$

We keep a and q fixed and sum up over $q' \neq q$ and $a' \neq a$, getting a bound of

$$\begin{aligned} &\ll \frac{X}{q(\log X)^2} \left(\sum_{q' \in A(X)} \frac{1}{q'} \right) \sum_{\substack{a' \in B(X) \\ a \neq a}} \frac{|a-a'|}{\varphi(|a-a'|)} \\ &\ll \frac{X}{q(\log X)^2} \cdot \log_3 X \cdot \frac{2\log X}{\sqrt{\log_3 X}} \quad \text{(by Proposition 6 and Equation 7)} \\ &= \frac{X\sqrt{\log_3 X}}{q \log X}. \end{aligned}$$

To conclude the argument, we sum up over all a's and all q's, getting an upper bound

$$\begin{aligned} &\ll \frac{X\sqrt{\log_3 X}}{\log X} \frac{2\log X}{\sqrt{\log_3 X}} \left(\sum_{q \in A(X)} \frac{1}{q} \right) \\ &\ll X \log_3 X \quad \text{(by Proposition 6).} \end{aligned}$$

Combining the bounds in the above three cases, we conclude that $S_2 \ll X \log_3 X$.

5.3 Putting Things Together

From the Cauchy-Schwarz inequality $S_0 S_2 \ge S_1^2$ and the above-established bounds $S_1 \gg X\sqrt{\log_3 X}$ and $S_2 \ll X \log_3 X$, we get that $S_0 \gg X$.

To transform the lower bound on S_0 to one on $\mathcal{S}(X)$, it remains to estimate the number of $n \in [X, 2X]$ that admit two overlapping representations. We claim that the total number of such n is $o(X)$. From this we conclude that $|\mathcal{S}(X)| \gg X$, which was our ultimate goal.

We conclude by justifying the preceding claim. For this it suffices to show that the number of pairs of overlapping representations in the interval $[X, 2X]$ is $o(X)$. To this end, consider two representations $(q, P, a) \neq (q', P', a')$ of the same number $n \in [X, 2X]$. Assume that $q + a = q' + a'$ (the argument in case $q - a = q' - a'$ requires only minor changes).

Now $n - (q + a) = q(P - 1) = q'(P' - 1)$, and so $qP - q'P' = (q' - q)$. Furthermore, as noted in Section 3, for two such overlapping representations, we have $q \neq q'$. Thus the general solution of the equation $qP - q'P' = (q' - q)$ in positive integers P and P' has the form $P = 1 + tq'$ and $P' = 1 + qt$ for a nonnegative integer parameter $t \ll X/qq'$. By Theorem 5, the number of t such that both $1 + tq$ and $1 + tq'$ are prime is

$$\begin{aligned} &\ll \frac{X}{qq'(\log X)^2} \frac{|q-q'|}{\varphi(|q-q'|)} \\ &\ll \frac{X \log_3 X}{qq'(\log X)^2} . \end{aligned}$$

In the above we have used the inequality $m/\varphi(m) \ll \log_2 m$ [12, Theorem 328], with $m = |q - q'| < \log X$ for large X.

The argument above shows that for $q \neq q'$, the number of pairs of primes P, P' for which there exist pairs of overlapping representations $(q, P, a), (q', P, a')$ of some $n \in [X, 2X]$ is at most $\frac{X \log_3 X}{qq'(\log X)^2}$. Summing up over the at most $\log X$ possible values of $a \in B(X)$, we see that for fixed $q \neq q'$ the total number of such pairs of overlapping representations is

$$\ll \frac{X \log_3 X}{qq' \log X}.$$

Here we used the fact that, thanks to the equation $q + a = q' + a'$, the choice of q, q', a uniquely determines a'.

Summing up over q, q', we get that the number of such possibilities is

$$\begin{aligned} &\ll \frac{X \log_3 X}{\log X} \left(\sum_{q \in A(X)} \frac{1}{q} \right)^2 \\ &\ll \frac{X (\log_3 X)^3}{\log X} \quad \text{(by Proposition 6)} \\ &= o(X). \end{aligned}$$

This concludes the proof of the claim.

6 Conclusion

We have defined a set $\mathcal{S} \subseteq \mathbb{N}$ of positive lower density relative to which the Skolem Problem for simple LRS is decidable. In an extended version of this paper we will show that we can solve the Skolem Problem relative to (a slight variant of) $\mathcal{S}$ for all LRS, not just the simple ones. For this we use methodology of Amoroso and Viada [2] to give effective upper bounds on the number of solutions of a class of bivariate polynomial-exponential Diophantine equations, generalising the analysis in Section 4. We are also continuing to study the lower density of $\mathcal{S}$. In particular, a future work will show that the set has lower density one, under certain heuristic assumptions on the distributions of primes, similar to the Cramér heuristic (as used to justify Cramér's conjecture on prime gaps).

References

1 S. Akshay, N. Balaji, A. Murhekar, R. Varma, and N. Vyas. Near-optimal complexity bounds for fragments of the Skolem problem. In *37th International Symposium on Theoretical Aspects of Computer Science, STACS*, volume 154 of *LIPIcs*, pages 37:1–37:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

2 F. Amoroso and E. Viada. Small points on subvarieties of a torus. *Duke Mathematical Journal*, 150(3), 2009.

3 A. Baker. *A Comprehensive Course in Number Theory.* Cambridge University Press, 2012.

4 D. Beauquier, A. M. Rabinovich, and A. Slissenko. A logic of probability with decidable model checking. *J. Log. Comput.*, 16(4), 2006.

5 J. Berstel and C. Reutenauer. *Noncommutative Rational Series with Applications.* Cambridge University Press, 2010.

6 V. Blondel and J. Tsitsiklis. A survey of computational complexity results in systems and control. *Automatica*, 36(9):1249–1274, 2000. doi:10.1016/S0005-1098(00)00050-9.

7 J.-Y. Cai, R. J. Lipton, and Y. Zalcstein. The complexity of the A B C problem. *SIAM J. Comput.*, 29(6), 2000.

8 G. Everest, A. van der Poorten, I. Shparlinski, and T. Ward. *Recurrence Sequences.* American Mathematical Society, 2003.

9 J.-H. Evertse. On sums of s-units and linear recurrences. *Compositio Mathematica*, 53(2):225–244, 1984.

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

11 H. Halberstam and H.-E. Richert. *Sieve methods.* LMS Monographs. Academic Press, 1974.

12 G. H. Hardy and E. M. Wright. *An introduction to the theory of numbers.* Oxford, at the Clarendon Press, 1954. 3rd ed.

13 R. Kannan and R. J. Lipton. Polynomial-time algorithm for the orbit problem. *JACM*, 33(4), 1986.

14 C. Lech. A note on recurring series. *Ark. Mat.*, 2, 1953.

15 F. Luca, J. Ouaknine, and J. Worrell. Universal Skolem Sets. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS*, pages 1–6. IEEE, 2021.

16 K. Mahler. Eine arithmetische Eigenschaft der Taylor Koeffizienten rationaler Funktionen. *Proc. Akad. Wet. Amsterdam*, 38, 1935.

17 D. W. Masser. Linear relations on algebraic groups. In *New Advances in Transcendence Theory.* Cambridge University Press, 1988.

18 M. Mignotte, T. Shorey, and R. Tijdeman. The distance between terms of an algebraic recurrence sequence. *J. für die reine und angewandte Math.*, 349, 1984.

19 J. Ouaknine and J. Worrell. On linear recurrence sequences and loop termination. *ACM SIGLOG News*, 2(2):4–13, 2015.

20 J. Piribauer and C. Baier. On Skolem-hardness and saturation points in markov decision processes. In *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020*, volume 168 of *LIPIcs*, pages 138:1–138:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

21 G. Rozenberg and A. Salomaa. *Cornerstones of Undecidability.* Prentice Hall, 1994.

22 H.P. Schlickewei and W.P. Schmidt. The number of solutions of polynomial-exponential equations. *Compositio Mathematica*, 120:193–225, January 2000.

23 T. Skolem. Ein Verfahren zur Behandlung gewisser exponentialer Gleichungen. In *Comptes rendus du congrès des mathématiciens scandinaves*, 1934.

24 T. Tao. *Structure and randomness: pages from year one of a mathematical blog.* American Mathematical Society, 2008.

25 A. J. Van Der Poorten and H. P. Schlickewei. Additive relations in fields. *Journal of the Australian Mathematical Society. Series A. Pure Mathematics and Statistics*, 51(1):154–170, 1991.

26 N. K. Vereshchagin. The problem of appearance of a zero in a linear recurrence sequence (in Russian). *Mat. Zametki*, 38(2), 1985.