Conjectural Decidability of the Skolem Problem

Florian Luca[1,2]

Mathematics Division, Stellenbosch University, Stellenbosch, South Africa

Joël Ouaknine[2,3,4]

Max Planck Institute for Software Systems, Saarland Informatics Campus, Saarbrücken, 66123, Germany

James Worrell[5]

Department of Computer Science, Oxford University, Wolfson Building, Parks Road, Oxford, OX1 3QD, UK

Abstract

The Skolem Problem asks to determine whether a given integer linear recurrence sequence (LRS) has a zero term. This problem, whose decidability has been open for many decades, arises across a wide range of topics in computer science, including loop termination, formal languages, automata theory, and probabilistic model checking, amongst many others.

In the present paper, we introduce a notion of "large" zeros of (non-degenerate) linear recurrence sequences, i.e., zeros occurring at an index larger than a double exponential of the magnitude of the data defining the given LRS. We establish two main results. First, we define an infinite set of prime numbers, termed "good", having density one amongst all prime numbers, with the following property: for any large zero of a given LRS, there is an interval around the large zero together with an upper bound on the number of good primes possibly present in that interval. The bound in question is much lower than one would expect if good primes were distributed similarly as ordinary prime numbers, as per the Cramér model in number theory. We

[1]Also affiliated with the Max Planck Institute for Software Systems, Germany.
[2]Supported by ERC grant DynAMiCs (101167561).
[3]Supported by DFG grant 389792660 as part of TRR 248.
[4]Also affiliated with Keble College, Oxford as emmy.network Fellow.
[5]Supported by EPSRC grant EP/X033813/1.

therefore conclude, conditionally on a strengthening of the classical Cramér conjecture, that large zeros do not exist, which would entail decidability of the Skolem Problem. Second, we show unconditionally that large zeros are very sparse: the set of positive integers that can possibly arise as large zeros of some LRS has null density. This in turn immediately yields a Universal Skolem Set of density one, answering a question left open in the literature.

Keywords: Skolem Problem, linear recurrence sequences, decidability, Cramér conjecture

---

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

$$u_{n+k} = a_1 u_{n+k-1} + \cdots + a_k u_n \tag{1}$$

where the coefficients $a_1, \ldots, a_k$ are integers. The celebrated theorem of Skolem, Mahler, and Lech [1, 2, 3] describes the set of zero terms of such a recurrence:

Theorem 1.1. *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 Thm. 1.1 can be refined by considering the notion of *non-degeneracy* of an LRS. An LRS is non-degenerate if in its minimal recurrence no quotient of two distinct roots of the characteristic polynomial is a root of unity.[6] A given LRS can be effectively decomposed as the 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 existence of a procedure to do so is equivalent to the existence of a procedure to determine whether

[6]For basic definitions, facts, and properties concerning linear recurrence sequences, we refer the reader to standard texts such as [4, Chaps. 1 and 2], [5, Chap. 4], or [6, Chap. 4].

an arbitrary given LRS has a zero term; the latter is known as the Skolem Problem. We refer to [7, Chap. 6] and [8, Chap. X] for expository accounts of the Skolem-Mahler-Lech theorem and discussion of the ineffectiveness of known proofs.

In computer science, the Skolem Problem lies at the heart of key decision problems in formal power series [9, 10], stochastic model checking [11], control theory [12, 13], and loop termination [14]. The problem is also closely related to membership problems on commutative matrix groups and semigroups, as considered in [15, 16]. We note that in several of the above-mentioned citations, the Skolem Problem is used as a reference benchmark to establish hardness of other open decision problems.

Decidability of the Skolem 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 [17, 18]. However for LRS of order 5 it is not currently known how to decide the Skolem Problem. For a (highly restricted) subclass of LRS, the paper [19] obtains nearly matching complexity lower and upper bounds for the problem.

Some recent lines of research have succeeded in establishing conditional decidability of the Skolem Problem for simple LRS (i.e., LRS none of whose characteristic roots are repeated), assuming certain classical number-theoretic conjectures [20, 21]. Nevertheless, to the best of our knowledge, no putative algorithm has to date been proposed to solve the Skolem Problem in full generality.

A different approach was initiated in [22, 23, 24] via the notion of *Universal Skolem Sets*. An infinite, recursive set $\mathcal{S} \subseteq \mathbb{N}$ is a Universal Skolem Set if there is some algorithm which, given any LRS, determines whether or not the LRS has a zero in $\mathcal{S}$. Decidability of the Skolem Problem is then of course equivalent to the assertion that $\mathbb{N}$ is itself a Universal Skolem Set. The authors of [22] succeded in exhibiting a *sparse* Universal Skolem Set, i.e., a set having null density, and left open the question of whether Universal Skolem Sets of strictly positive density, or even density one, could be constructed (the interest in high-density Universal Skolem Sets being that they approximate $\mathbb{N}$ more closely). The question was partially answered in [23], which presented a positive-density Universal Skolem Set, albeit restricted to simple LRS, and in [24], which exhibited a Universal Skolem Set of strictly posi-

tive density, and even established density 1 subject to the Bateman-Horn conjecture in number theory.

In this paper we propose an explicit bound for the largest zero of a non-degenerate LRS in terms of the data describing the LRS. We call zeros that exceed this bound *large zeros* of the LRS. Evidently, decidability of the Skolem Problem would follow from a proof that large zeros do not exist. Using deep upper bounds on the cardinality of the set of zeros of non-degenerate algebraic LRS due to Amoroso and Viada, we show that the set of positive integers arising as large zeros of some non-degenerate LRS has null density, which in turn yields a Universal Skolem Set of unconditional density one. We present this result in Sec. 5.

While a proof that large zeros do not exist currently seems well out of reach, we give a heuristic argument as to why this should nevertheless be expected. This argument is based on an analogue of the well-known Cramér conjecture on gaps between consecutive primes. This conjecture, originally formulated by Cramér in 1936 [25] and subsequently refined by various number theorists into its present form, asserts that, for some constant $\kappa > 1$, for every prime p the distance to the next largest prime is at most $\kappa(\log p)^2$. The conjecture is based on the heuristic that the sequence of prime numbers behaves similarly to a Poisson-like random process in which the probability of a number x being prime is $1/\log x$. The largest observed prime gap is approximately $0.9206(\log p)^2$ (involving a search over all primes up to $4 \cdot 10^{18}$) [26], however the best known upper bound on prime gaps is $O(p^{0.525})$, due to Baker, Harman, and Pintz [27], which is far from Cramér's conjectured bound. Cramér himself proved that, under the Riemann hypothesis, prime gaps are bounded above by $O(p^{0.5} \log p)$ [25]. On the other hand, the best known lower bound on largest prime gaps is $\Omega\left(\frac{\log p \log\log p \log\log\log\log p}{\log\log\log p}\right)$, due to Ford, Green, Konyagin, Maynard, and Tao [28], which is some way from the conjectured upper bound. We refer to [29] for a discussion of Cramér's conjecture and its refinements.

Here we define a subset of so-called *good* primes based on divisibility properties of LRS. We show that the set of good primes has density one in the set of all primes, or in other words that, asymptotically speaking, almost all primes are good primes. We further show that if the Cramér conjecture applies also to gaps between consecutive good primes, then large zeros of LRS cannot exist. The proof of the latter result proceeds by establishing an upper bound on the number of good primes in the neighbourhood of a

large zero that violates the conjectured upper bound on gaps between good primes. In other words, if good primes are distributed according to Cramér's heuristic then large zeros cannot exist and the Skolem Problem is decidable.

2. Background

We will need some basic notions concerning algebraic numbers. All material can be found in [30]. Recall that a *number field* $\mathbb{K}$ is a subfield of $\mathbb{C}$ that is finite dimensional as a vector space over $\mathbb{Q}$. We assume that $\mathbb{K}$ is a Galois extension of $\mathbb{Q}$, that is, it arises as the splitting field of a polynomial with integer coefficients. All elements of $\mathbb{K}$ are algebraic over $\mathbb{Q}$, that is, they arise as roots of polynomials with integer coefficients. Those elements that arise more specifically as roots of monic polynomials with integer coefficients are called *algebraic integers*. The algebraic integers in $\mathbb{K}$ form a subring, denoted $\mathcal{O}_{\mathbb{K}}$.

For a number field $\mathbb{K}$, we denote by $\mathrm{Gal}(\mathbb{K}/\mathbb{Q})$ the group of field automorphisms of $\mathbb{K}$. Given $\alpha \in \mathbb{K}$, the *norm* of α is defined by

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

The norm $\mathcal{N}_{\mathbb{K}/\mathbb{Q}}(\alpha)$ is rational for all $\alpha \in \mathbb{K}$; moreover $\mathcal{N}_{\mathbb{K}/\mathbb{Q}}(\alpha) = 0$ iff $\alpha = 0$, and $\mathcal{N}_{\mathbb{K}/\mathbb{Q}}(\alpha)$ is an integer if $\alpha \in \mathcal{O}_{\mathbb{K}}$. Clearly we have $|\mathcal{N}_{\mathbb{K}/\mathbb{Q}}(\alpha)| \leq M^{d_{\mathbb{K}}}$, 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 recall that every ideal in $\mathcal{O}_{\mathbb{K}}$ can be written uniquely up to the order of its factors 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 $P \mid \mathcal{N}_{\mathbb{K}/\mathbb{Q}}(\alpha)$ for all $\alpha \in \mathfrak{p}$.

Let $\mathfrak{p}$ be a prime ideal of $\mathcal{O}_{\mathbb{K}}$ lying above $P \in \mathbb{Z}$. Recall that the Frobenius automorphism $\sigma \in \mathrm{Gal}(\mathbb{K}/\mathbb{Q})$ corresponding to $\mathfrak{p}$ is such that $\sigma(\alpha) \equiv \alpha^P \bmod \mathfrak{p}$ for all $\alpha \in \mathcal{O}_{\mathbb{K}}$; in fact it is the unique Galois automorphism with this property. Note however that for the Frobenius automorphism to be well-defined on $\mathbb{K}$, it is necessary for P to be unramified, for which it suffices to check that P does not divide the discriminant of $\mathbb{K}$.

3. Large Zeros and Good Primes

For an LRS $\mathbf{u} = \langle u_n \rangle_{n=0}^{\infty}$ as in (1), define its *height*[7] to be

$$H_{\mathbf{u}} := \max\{k, |a_1|, \ldots, |a_k|, |u_0|, \ldots, |u_{k-1}|\}\,.$$

Note that there are only finitely many different LRS of any specified height bound. In the rest of the paper, we shall focus exclusively on LRS "of sufficient height"; by this we implicitly postulate the existence of some absolute constant C and further assume that all LRS under consideration have height in excess of C. In our subsequent development, it would be straightforward to provide explicit suitable numerical values for each of the lower bounds that we require,[8] however an explicit numerical value for C would still depend on the precise formulation of the strengthening of the Cramér-Granville conjecture that we introduce in Sec. 4 (see Conjecture 4.2).

We say that n is a zero of $\mathbf{u}$ if $u_n = 0$, and we say that it is a *large zero* if the inequality

$$n < \exp\exp(10 H_{\mathbf{u}} \log H_{\mathbf{u}}) \tag{2}$$

fails. As we argue later on, there are good reasons to expect that (2) holds for *all* zeros of all sufficiently high non-degenerate LRS, which in turn would establish decidability of the Skolem Problem.[9]

3.1. Bad Primes and Good Primes

In this section, let $\mathbf{u}$ be a fixed sufficiently high non-degenerate LRS satisfying (1), and let us write $H := H_{\mathbf{u}}$ (that is, we omit the explicit dependence on $\mathbf{u}$).

[7] Note that we consider here the *magnitude* of the numbers defining a given LRS (rather than their bit size as is more common in complexity theory). An alternative definition in terms of bit size would of course be possible, only requiring adjusting (2) appropriately.

[8] In fact, in all of the height-related asymptotic inequalities that we derive, positing a lower bound of 12 on the height would suffice.

[9] The expression in (2) has of course been chosen in order for our mathematical argument to go through. In actual fact, it is plausible to expect that an expression involving a *single* exponential would suffice: as far as we are aware, there is currently no known construction of a family of non-degenerate LRS having zeros at indices of magnitude merely singly exponential in the height of the LRS, as defined above.

We can express the general term u_t of $\mathbf{u}$ in exponential-polynomial form, i.e.,

$$u_t = \sum_{i=1}^{s} Q_i(t)\alpha_i^t\,, \tag{3}$$

where $s \leq H$ and $\alpha_1, \ldots, \alpha_s$ are the roots of the characteristic polynomial

$$x^k - a_1 x^{k-1} - \cdots - a_k$$

of $\mathbf{u}$ and $Q_1, \ldots, Q_s$ are univariate polynomials. Note that all characteristic roots are algebraic integers since the characteristic polynomial is monic and comprises exclusively coefficients in $\mathbb{Z}$. Recall that if α_i has multiplicity μ_i as a characteristic root then Q_i has degree at most $\mu_i - 1$. Let $\mathbb{K} := \mathbb{Q}(\alpha_1, \ldots, \alpha_s)$. The coefficients of each Q_i are in $\mathbb{K}$ and can straightforwardly be computed from the initial values $u_0, \ldots, u_{k-1}$ of the sequence by solving a system of k linear equations, thanks to (3). By Cramer's determinant rule,[10] each of the coefficients of Q_i is the quotient of an algebraic integer by the determinant $\Delta := \det(M)$ of the matrix M below:[11]

$$\begin{bmatrix} 1 & \ldots & 0 & 1 & \ldots & 0 & 1 & \cdots \\ \alpha_1 & \ldots & \alpha_1 & \alpha_2 & \ldots & \alpha_{s-1} & \alpha_s & \ldots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ \alpha_1^{k-1} & \ldots & (k-1)^{\mu_1-1}\alpha_1^{k-1} & \alpha_2^{k-1} & \ldots & (k-1)^{\mu_s-1}\alpha_{s-1}^{k-1} & \alpha_s^{k-1} & \ldots \end{bmatrix}.$$

By the Cauchy root bound we have $|\alpha_i| \leq 1 + H$ for $i \in \{1, \ldots, s\}$. It follows that the squared Euclidean norm of each column vector above is at most

$$k(k-1)^{2(k-1)}(1+H)^{2k} < k^{2k}(1+H)^{2k}\,.$$

Thus, by the Hadamard inequality,

$$|\Delta|^2 < (k^{2k}(1+H)^{2k})^k = (k(1+H))^{2k^2}\,.$$

[10] This rule is named after the 18th-century Genevan mathematician Gabriel Cramer, who is presumably unrelated to the 20th-century Swedish mathematician Harald Cramér, whose work plays an important role in motivating the present article.

[11] The matrix M has s blocks, one for each characteristic root. For $\ell \in \{1, \ldots, s\}$ the ℓ-th block has dimension $k \times \mu_\ell$ and has (i,j)-th element $(i-1)^{(j-1)}\alpha_\ell^{(i-1)}$ for $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, \mu_\ell\}$.

The determinant Δ is in general, of course, a complex number. Note however that any Galois automorphism $\sigma \in \mathrm{Gal}(\mathbb{K}/\mathbb{Q})$ will permute the characteristic roots, and thus when applied to M will have the effect of permuting its columns. As a result, for any such σ, $\sigma(\Delta) = \pm\Delta$, and therefore the quantity Δ^2 is stable under Galois automorphisms. We conclude that Δ^2 must be a rational number, and since it is also by construction an algebraic integer,[12] we must have $\Delta^2 \in \mathbb{Z}$.

Let us now consider the LRS $\mathbf{v} := \Delta^2\mathbf{u}$, noting that $\mathbf{u}$ and $\mathbf{v}$ share the same zeros. Writing

$$v_t = \sum_{i=1}^{s} P_i(t)\alpha_i^t \,,$$

we observe that all the coefficients of each of the P_i are algebraic integers. We therefore have, for each $1 \leq i \leq s$,

$$P_i(t) = \Delta^2 Q_i(t) = \sum_{j=0}^{\mu_i - 1} c_{i,j} t^j \,.$$

We wish to estimate the size of each $c_{i,j} \in \mathcal{O}_{\mathbb{K}}$. From our earlier calculation via Cramer's determinant rule, noting that $|u_0|, \ldots, |u_{k-1}|$ are all bounded above by $H \leq 1 + H$, and invoking the Hadamard inequality once more, we conclude that the house of each $c_{i,j}$ is bounded above by

$$|\Delta|(k^k(1+H)^k)^k < (k(1+H))^{2k^2} < (1+H)^{4H^2} < H^{H^3} \,. \tag{4}$$

Let $\sigma \in \Sigma_s$ be any permutation of the first s integers and let

$$\beta_i := \alpha_{\sigma(i)} \quad \text{for} \quad i = 1, \ldots, s \,.$$

For some nonnegative integer m consider the algebraic integer

$$v_{m,\sigma} = \sum_{i=1}^{s} P_i(m)\beta_i\alpha_i^m \,. \tag{5}$$

We need one last technical ingredient in order to define what it means for a prime number to be *bad*. Let $X > \exp\exp e$ be a power of 2. We say that $\mathbf{u}$ *is small at level* X provided that

$$H < \frac{\log\log X}{10\log\log\log X} \,.$$

[12] Note that every entry of M is an algebraic integer.

Definition 3.1. *We say that a prime $P > \exp\exp e$ is bad, if there exist X a power of 2 with $X/2 < P < X$, together with an LRS $\mathbf{u}$ which is small at level X, a permutation $\sigma \in \Sigma_s$, and an integer $m \in [0, X^{1/4}]$, such that*

- *The algebraic integer $v_{m,\sigma}$ defined in (5) above is non-zero, and*
- *P is a prime factor of $\mathcal{N}_{\mathbb{K}/\mathbb{Q}}(v_{m,\sigma})$.*

Let $\mathcal{P}_{\mathrm{bad}}(X)$ be the set of bad primes in $[X/2, X]$, and $\mathcal{P}_{\mathrm{bad}} := \bigcup \mathcal{P}_{\mathrm{bad}}(2^k)$, where the union is taken over all positive integers k such that $2^k > \exp\exp e$.

Proposition 3.2. *We have*

$$\#\mathcal{P}_{\mathrm{bad}}(X) < X^{2/3}$$

for all $X > X_0$, where X_0 is some effective absolute constant.

Proof. In order to estimate the size of $\mathcal{P}_{\mathrm{bad}}(X)$, we first need to find out:

1. How many such expressions (5) are there?
2. How large are they?

For (**1**), let us count the number of distinct possible LRS of size at most H. Such LRS have coefficients $a_1, \ldots, a_k$ and initial values $u_0, \ldots, u_{k-1}$ all in $[-H, H]$, an interval containing at most $2H + 1 < 3H$ integers. Altogether for fixed k there are at most $(3H)^{2k} \leq (3H)^{2H}$ $2k$-tuples, and summing up over k we derive an upper bound of $H(3H)^{2H} < H^{3H}$ distinct possible LRS of size at most H.

This in turn is an upper bound on the number of s-tuples $((Q_i, \alpha_i))_{i=1}^s$. We must then multiply this quantity with the number of possible permutations of the characteristic roots, which is at most $H! < H^H$. There are therefore at most H^{4H} linear recurrence sequences $\mathbf{w} = \langle w_m \rangle_{m=0}^{\infty}$ whose m-th term is given by

$$w_m = \sum_{i=1}^{s} P_i(m)\beta_i\alpha_i^m \quad \text{for all} \quad m \geq 0\,.$$

This answers (**1**). As for (**2**), recall that the coefficients of P_i are of absolute value at most H^{H^3}, as per (4). $P_i(m)$ comprises at most H monomials, the largest of which is at most $m^H < X^H$, and the largest root has magnitude

at most $1 + H < 2H$. Thus each individual term w_m is of absolute value at most

$$H^{H^3+1}(2H)X^H(2H)^{X^{1/4}} = \exp\left((H^3+1)\log H + \log(2H) + H\log X + X^{1/4}\log(2H)\right) < \exp(X^{0.26})$$

for $X > X_0$, since H is tiny in comparison to X. Hence the norm of the number shown in (5) is of size at most

$$\exp(H!X^{0.26}) < \exp(X^{0.27}) \quad \text{for} \quad X > X_0\,,$$

since the degree of $\mathbb{K}$ is at most $H!$ (as $\mathbb{K}$ is the splitting field of a polynomial of degree at most H). Moreover, as noted earlier, there are at most H^{4H} such expressions. Thus a bad prime P divides an integer which is a product of such numbers and is of size at most

$$\exp(H^{4H}X^{0.27}) < \exp(X^{0.28}) \quad \text{for} \quad X > X_0\,.$$

Therefore the number of possible choices for P is at most $X^{0.28}$. Since the number of choices for m is at most $X^{0.25}$, we conclude that, for $X > X_0$, the cardinality of $\mathcal{P}_{\text{bad}}(X)$ is at most

$$X^{0.25+0.28} < X^{2/3}\,,$$

as required. □

Finally, let us write $\mathcal{P} = \{p_1, p_2, \ldots\}$ to denote the set of prime numbers, enumerated in increasing order, and let $\mathcal{P}_{\text{good}} := \mathcal{P} \setminus \mathcal{P}_{\text{bad}} = \{g_1, g_2, \ldots\}$ denote the subset of *good* primes, again enumerated in increasing order. Note that, by Prop. 3.2 along with the prime number theorem, the set of bad primes has null density amongst the prime numbers. This in turn entails that good primes have density one amongst all prime numbers.

4. The Cramér Argument

In this section we present a heuristic argument supporting the assertion that large zeros of sufficiently high LRS do not exist. The strategy is as follows. Assuming that good primes are distributed similarly as ordinary

primes, then according to the Cramér model in number theory, one would expect that Cramér's conjecture on gaps between primes applies also to good primes. More precisely, this conjecture postulates the existence of precise upper bounds on the largest possible gap between consecutive primes, and is predicated on the heuristic that the primes behave as a set of randomly distributed integers with asymptotic density conforming to the prime number theorem. However we show that around any large zero of an LRS there is an interval and an upper bound on the number of good primes in the interval that together contradict the above Cramér-type conjecture on gaps between good primes. We therefore surmise that large zeros do not exist.

Recall that $\mathcal{P} = \{p_1, p_2, \ldots\}$ denotes the set of prime numbers enumerated in increasing order.

Conjecture 4.1 (Cramér-Granville)**.** *For some $\kappa > 1$,*

$$\limsup_{j\to\infty} \frac{p_{j+1} - p_j}{(\log p_j)^2} = \kappa\,.$$

Cramér initially suggested that the constant κ in Conjecture 4.1 might be 1 [25], but several decades later, building on substantial developments in the field, Granville produced evidence that $\kappa \geq 2e^{-\gamma} \approx 1.1229\ldots$, where γ is the Euler–Mascheroni constant [29]. There is in any event considerable computational evidence in support of the Cramér-Granville conjecture [31, 32].

As noted earlier, thanks to Prop. 3.2 and the prime number theorem, good primes have density one amongst all prime numbers:

$$\lim_{X\to\infty} \frac{\#\left(\mathcal{P}_{\text{good}} \cap [0, X]\right)}{\#\left(\mathcal{P} \cap [0, X]\right)} = 1\,.$$

In other words, asymptotically speaking, almost all primes are good primes. Accordingly, it seems reasonable to suppose that good primes should behave similarly to ordinary primes, or at least should exhibit similar "statistical" properties. We therefore formulate the following strengthening of the Cramér-Granville conjecture:

Conjecture 4.2. *For some $\eta > 1$,*

$$\limsup_{j\to\infty} \frac{g_{j+1} - g_j}{(\log g_j)^2} = \eta\,.$$

We now have the following result.

Theorem 4.3. *Conjecture 4.2 implies that large zeros of sufficiently high LRS do not exist. More precisely, assuming Conjecture 4.2, there exists an absolute constant $C > 0$ such that, for all non-degenerate LRS $\mathbf{u}$ of height $H_{\mathbf{u}}$ at least C, whenever $u_n = 0$ then $n < \exp\exp(10H_{\mathbf{u}} \log H_{\mathbf{u}})$.*

Proof. Conjecture 4.2 can be reformulated as follows: there exist $\eta > 1$ and $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$, the interval

$$[n - \eta(\log n)^2, n]$$

always contains some good prime. In turn, this implies that the interval $[n - 0.5n^{1/4}, n]$ must contain at least $n^{1/4}/(2\eta(\log n)^2)$ distinct good primes for n sufficiently large.

Suppose now that there is some sufficiently high LRS $\mathbf{u}$ having large zero $u_n = 0$. By definition, this means that

$$n \geq \exp\exp(10H_{\mathbf{u}} \log H_{\mathbf{u}})\,. \tag{6}$$

Let X be the power of 2 such that $X/2 \leq n < X$. We first claim that $\mathbf{u}$ is small at level $Y := X/2$, i.e., that the inequality

$$H_{\mathbf{u}} < \frac{\log\log Y}{10\log\log\log Y} \tag{7}$$

holds.

Suppose for a contradiction that Eq. (7) fails. Then

$$\log H_{\mathbf{u}} \geq \log\log\log Y - \log(10\log\log\log Y) \geq \frac{3}{4}\log\log\log Y\,,$$

where the second inequality follows from the assumption that Y is sufficiently large (since $\mathbf{u}$ is assumed to be sufficienctly high). Combining with the negation of (7), we get

$$10H_{\mathbf{u}} \log H_{\mathbf{u}} \geq 10\left(\frac{\log\log Y}{10\log\log\log Y}\right)\frac{3}{4}\log\log\log Y = \frac{3}{2}\log\log Y\,,$$

i.e., $\exp\exp(10H_{\mathbf{u}} \log H_{\mathbf{u}}) \geq Y^{(\log Y)^{0.5}}$, whence (from (6)) we conclude that $n \geq Y^{(\log Y)^{0.5}}$, contradicting (for sufficiently large Y) the fact that $n < X = 2Y$ and thereby establishing the claim.

We note that the smallness of $\mathbf{u}$ at level $X/2$ immediately also entails its smallness at level X.

Next, write $n = P + m$, where $P \in [n - 0.5n^{1/4}, n]$ is a good prime and $0 \leq m < n^{1/4}$. As in the previous section, let $\alpha_1, \ldots, \alpha_s$ be the characteristic roots of $\mathbf{u}$, put $\mathbb{K} := \mathbb{Q}(\alpha_1, \ldots, \alpha_s)$, and let Δ^2 be the smallest positive integer such that, writing $\mathbf{v} := \Delta^2\mathbf{u}$, every term of v_t of $\mathbf{v}$ has a representation as an exponential polynomial

$$v_t = \sum_{i=1}^{s} P_i(t)\alpha_i^t$$

in which all polynomials P_i have algebraic-integer coefficients. We let k stand for the order of $\mathbf{u}$ and abbreviate $H_{\mathbf{u}}$ as H for the remainder of the proof.

Since $u_n = v_n = 0$, we get

$$0 = \sum_{i=1}^{s} P_i(P + m)\alpha_i^{P+m}\,.$$

We now reduce the above equation modulo $\mathfrak{p}$, where $\mathfrak{p}$ is some prime ideal of $\mathcal{O}_{\mathbb{K}}$ dividing P, from which we deduce that P divides

$$\mathcal{N}_{\mathbb{K}/\mathbb{Q}}\left(\sum_{i=1}^{s} P_i(m)\beta_i\alpha_i^m\right)\,, \tag{8}$$

where each $\beta_i = \sigma(\alpha_i)$ is obtained from applying the Frobenius automorphism induced by $\mathfrak{p}$ in $\mathbb{K}$ to α_i. Recall that it is necessary for this automorphism to be well defined that P not divide the discriminant of $\mathbb{K}$. Note however that the discriminant of $\mathbb{K}$ is a positive integer bounded above by the absolute value of the discriminant of the characteristic polynomial of $\mathbf{u}$, namely $\prod_{i\neq j}|\alpha_i - \alpha_j|$, and this last quantity is itself at most $(2 + H)^{s(s-1)} \leq H^{H^3}$ thanks to the Cauchy root bound $|\alpha_i| \leq 1 + H$. Since $\mathbf{u}$ is small at level $X/2$ and $X/2 \leq n$, we have $H < (\log\log n)/(10\log\log\log n)$, and hence

$$H^{H^3} < \exp\left(\left(\frac{\log\log n}{10\log\log\log n}\right)^3 \log\log\log n\right) < \exp((\log\log n)^3) < n/2\,.$$

Since P is larger than $n/2$ for n sufficiently large, we conclude that P cannot divide the discriminant of $\mathbb{K}$, as required.

Observe that P must lie in one of two intervals, namely $[X/4, X/2]$ or $[X/2, X]$. Since we know that $\mathbf{u}$ is small at both levels $X/2$ and X, it follows

that expression (8) cannot be non-zero, otherwise P, by definition, would be a bad prime. Expression (8) is therefore equal to zero.

Let us count how many expressions of the form (8) can vanish. More precisely, consider the (complex-valued) LRS $\mathbf{w} = \langle w_j \rangle_{j=0}^{\infty}$ whose j-th term is given by

$$w_j = \sum_{i=1}^{s} P_i(j)\beta_i\alpha_i^j \quad \text{for all} \quad j \geq 0\,,$$

and whose order is at most k. Amoroso and Viada [33] prove that the number of distinct positive integers m such that $w_m = 0$ is at most

$$(8k^k)^{8(k^k)^6} = \exp(8k^{6k}\log(8k^k)) \leq \exp(8H^{6H+1}\log(8H)) < \\ \exp(16H^{6H+1}\log H) < \exp(H^{7H}) = \exp\exp(7H\log H)\,.$$

Of course, given $\mathbf{u}$, the s-tuple $(\beta_1, \ldots, \beta_s)$ can be chosen in at most $s! < H^H$ ways. Thus the total number of possible zeros for expressions of the form (8) is at most

$$H^H \exp\exp(7H\log H) < \exp\exp(7H\log H + \log H + \log\log H) \\ < \exp\exp(8H\log H)\,.$$

Since distinct choices of P give rise to distinct such zeros,[13] and (as noted earlier) there are at least $n^{1/4}/(2\eta(\log n)^2)$ possible choices for P, we conclude that

$$\frac{n^{1/4}}{2\eta(\log n)^2} < \exp\exp(8H\log H)\,,$$

or equivalently

$$\frac{n^{1/4}}{(\log n^{1/4})^2} < 32\eta\exp\exp(8H\log H)\,.$$

Observe that, for any fixed $D > 0$, the inequality $x/(\log x)^2 < y$ implies that $x < y^2/D$, provided that x is sufficiently large. Applying this to the above inequality with $x = n^{1/4}$ and $D = 32\eta$, we derive the upper bound

$$n^{1/4} < \exp(2\exp(8H\log H))\,,$$

[13] Recall that $n = P + m$, and thus distinct choices of P entail distinct values of m.

whence

$$n < \exp(8\exp(8H\log H)) < \exp\exp(9H\log H)\,,$$

contradicting Eq. (6) to the effect that n is a large zero of $\mathbf{u}$. We conclude that large zeros do not exist, as required. □

5. Large Zeros Are Unconditionally Sparse

In this section, we prove unconditionally that large zeros are sparse, i.e., have null density amongst the positive integers.

To this end, let

$$\mathcal{L} := \{n \in \mathbb{N} : \text{there exists a non-degenerate LRS } \mathbf{u} \text{ such that } u_n = 0 \text{ and } n \geq \exp\exp(10H_{\mathbf{u}}\log H_{\mathbf{u}})\}\,.$$

Thus $\mathcal{L}$ is the set of large zeros of *some* non-degenerate LRS, without any height restrictions.

Theorem 5.1. *The set $\mathcal{L}$ has null density. In fact, writing $\mathcal{L}(X) = \mathcal{L} \cap [1, X]$, the inequality*

$$\#\mathcal{L}(X) = O\left(\frac{X}{(\log X)^B}\right)$$

holds with any constant $B > 0$.

Proof. We let X be large, and aim to count the n in $\mathcal{L}$ that lie in $[X/2, X]$. Jia [34] proved that the set of $n \in [X/2, X]$ such that the interval $I_n := [n - n^{1/19}, n]$ contains fewer than $X^{1/19}/(\log X)^2$ primes is of counting function $O(X/(\log X)^B)$ with any $B > 0$. Let us therefore consider $n \in [X/2, X]$ such that I_n contains at least $X^{1/19}/(\log X)^2$ primes. Write $n = P + m$, where $P \in I_n$ and $m < X^{1/19}$. Let $\mathbf{u}$ be an LRS having n as a large zero and whose recurrence is given by (1). Using the same notation and reasoning as in the proof of Thm. 4.3, we have, for sufficiently large X, that P must divide

$$\mathcal{N}_{\mathbb{K}/\mathbb{Q}}\left(\sum_{i=1}^{s} P_i(m)\beta_i\alpha_i^m\right)\,. \tag{9}$$

If the above expression is non-zero, then P is a bad prime, which in turn means that n is within $O(X^{1/19})$ of a bad prime. However, as shown in Prop. 3.2, the number of bad primes below X is $O(X^{2/3})$. Hence the total

number of such $n \in [X/2, X]$ is $O(X^{2/3+1/19}) = O(X/(\log X)^B)$ with any $B > 0$.

Finally, assume that expression (9) is zero, in which case

$$\sum_{i=1}^{s} P_i(m)\beta_i\alpha_i^m = 0\,.$$

Once again, as detailed in the proof of Thm. 4.3, the Amoroso-Viada bounds imply that the corresponding number of possible zeros for expressions of the form above is at most $\exp\exp(8H_{\mathbf{u}}\log H_{\mathbf{u}})$. Since distinct choices of P give rise to distinct such zeros, and since I_n contains at least $X^{1/19}/(\log X)^2$ primes, we deduce that

$$\frac{X^{1/19}}{(\log X)^2} < \exp\exp(8H_{\mathbf{u}}\log H_{\mathbf{u}})\,.$$

Noting that $n \leq X$, we have, for sufficiently large n, the following string of inequalities:

$$n < \left(\frac{n^{1/19}}{(\log n)^2}\right)^{20} \leq \left(\frac{X^{1/19}}{(\log X)^2}\right)^{20} < (\exp\exp(8H_{\mathbf{u}}\log H_{\mathbf{u}}))^{20} < \exp\exp(9H_{\mathbf{u}}\log H_{\mathbf{u}})\,,$$

or in other words that $n \notin \mathcal{L}$.

Putting everything together, we conclude that, for any $B > 0$, if X is sufficiently large then the number of large zeros in $[X/2, X]$ is $O(X/(\log X)^B)$, from which one immediately deduces the statement of the theorem. □

Corollary 5.2. *The set $\mathcal{S} := \mathbb{N} \setminus \mathcal{L}$ is a Universal Skolem Set of density one.*

Proof. It is clear that the set $\mathcal{L}$ is recursive, and hence that $\mathcal{S}$ is recursive as well.

Density one follows from Thm. 5.1, and universality follows from the fact that $\mathcal{S}$, by definition, doesn't contain any large zeros. Thus given any non-degenerate LRS $\mathbf{u}$ of size $H_{\mathbf{u}}$, its only possible zeros in $\mathcal{S}$ can only lie in the interval $[0, \exp\exp(10H_{\mathbf{u}}\log H_{\mathbf{u}})]$, which can readily be checked. □

6. Concluding Remarks

Thanks to Thm. 4.3, Conjecture 4.2 implies the *existence* of an algorithm to solve the Skolem Problem, as follows. Recall that Thm. 4.3 asserts the existence of an absolute constant C such that any non-degenerate LRS $\mathbf{u}$ of height $H_{\mathbf{u}} \geq C$ has no zeros at index larger than $\exp\exp(10H_{\mathbf{u}} \log H_{\mathbf{u}})$. On the other hand, there are only finitely many LRS of height at most C; therefore there exists *some* algorithm (call it an oracle) that correctly determines, for each such C-bounded LRS, whether or not it has a zero.

Now given an LRS $\mathbf{u}$, first decompose $\mathbf{u}$ into finitely many non-degenerate LRS, and check that none of these is identically zero. Next, if any of these LRS has height below C, invoke the aforementioned oracle to determine whether it has a zero. Finally, assuming that no zeros have yet been found, for each remaining LRS $\mathbf{v}$, check whether $\mathbf{v}$ has a zero in the interval $[0, \exp\exp(10H_{\mathbf{v}} \log H_{\mathbf{v}})]$.

Of course, even if Conjecture 4.2 were to be established and C explicitly known, and setting aside the question of how to obtain the oracle, the above algorithm unfortunately remains impractical, since the magnitudes involved are much too large to envisage any reasonable implementation.

References

[1] T. Skolem, Ein Verfahren zur Behandlung gewisser exponentialer Gleichungen, in: Comptes rendus du congrès des mathématiciens scandinaves, 1934.

[2] K. Mahler, Eine arithmetische Eigenschaft der Taylor Koeffizienten rationaler Funktionen, Proc. Akad. Wet. Amsterdam 38 (1935).

[3] C. Lech, A note on recurring series, Ark. Mat. 2 (1953).

[4] G. Everest, A. van der Poorten, I. Shparlinski, T. Ward, Recurrence Sequences, American Mathematical Society, 2003.

[5] M. Kauers, P. Paule, The Concrete Tetrahedron — Symbolic Sums, Recurrence Equations, Generating Functions, Asymptotic Estimates, Texts & Monographs in Symbolic Computation, Springer, 2011.

[6] R. P. Stanley, Enumerative combinatorics, Cambridge studies in advanced mathematics Volume 1, 2nd Edition (2011).

[7] J. Berstel, C. Reutenauer, Noncommutative Rational Series with Applications, Cambridge University Press, 2010.

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

[9] G. Rozenberg, A. Salomaa, Cornerstones of Undecidability, Prentice Hall, 1994.

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

[11] J. Piribauer, C. Baier, On Skolem-hardness and saturation points in Markov decision processes, in: ICALP, Vol. 168 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, pp. 138:1–138:17.

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

[13] N. Fijalkow, J. Ouaknine, A. Pouly, J. S. Pinto, J. Worrell, On the decidability of reachability in linear time-invariant systems, in: HSCC, ACM, 2019, pp. 77–86.

[14] J. Ouaknine, J. Worrell, On linear recurrence sequences and loop termination, ACM SIGLOG News 2 (2) (2015) 4–13.

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

[16] R. Kannan, R. J. Lipton, Polynomial-time algorithm for the orbit problem, JACM 33 (4) (1986).

[17] M. Mignotte, T. Shorey, R. Tijdeman, The distance between terms of an algebraic recurrence sequence, J. für die reine und angewandte Math. 349 (1984).

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

[19] S. Akshay, N. Balaji, A. Murhekar, R. Varma, N. Vyas, Near-optimal complexity bounds for fragments of the Skolem problem, in: STACS, Vol. 154 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, pp. 37:1–37:18.

[20] R. Lipton, F. Luca, J. Nieuwveld, J. Ouaknine, D. Purser, J. Worrell, On the Skolem problem and the Skolem conjecture, in: LICS, ACM, 2022, pp. 5:1–5:9.

[21] Y. Bilu, F. Luca, J. Nieuwveld, J. Ouaknine, D. Purser, J. Worrell, Skolem meets Schanuel, in: MFCS, Vol. 241 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022, pp. 20:1–20:15.

[22] F. Luca, J. Ouaknine, J. Worrell, Universal Skolem sets, in: LICS, IEEE, 2021, pp. 1–6.

[23] F. Luca, J. Ouaknine, J. Worrell, A universal Skolem set of positive lower density, in: MFCS, Vol. 241 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022, pp. 73:1–73:12.

[24] F. Luca, J. Maynard, A. Noubissie, J. Ouaknine, J. Worrell, Skolem meets bateman-horn, CoRR abs/2308.01152 (2023).

[25] H. Cramér, On the order of magnitude of the difference between consecutive prime numbers, Acta Arith. 2 (1936) 23–46.

[26] T. Oliveira e Silva, S. Herzog, S. Pardi, Empirical verification of the even Goldbach conjecture and computation of prime gaps up to $4 \cdot 10^{18}$, Math. Comp. 83 (288) (2014) 2033–2060.

[27] R. C. Baker, G. Harman, J. Pintz, The difference between consecutive primes, ii, Proceedings of the London Mathematical Society 83 (3) (2001) 532–562.

[28] K. Ford, B. Green, S. Konyagin, J. Maynard, T. Tao, Long gaps between primes, Journal of the American Mathematical Society 31 (1) (2018) 65–105.

[29] A. Granville, Harald Cramér and the distribution of prime numbers, Scandinavian Actuarial Journal 1995 (1) (1995) 12–28.

[30] A. Fröhlich, M. J. Taylor, Algebraic Number Theory, Vol. 27 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, 1993.

[31] A. Odlyzko, M. Rubinstein, M. Wolf, Jumping champions, Experimental Mathematics 8 (2) (1999) 107–118.

[32] T. R. Nicely, New maximal prime gaps and first occurrences, Math. Comput. 68 (227) (1999) 1311–1315.

[33] F. Amoroso, E. Viada, On the zeros of linear recurrence sequences, Acta Arith. 147 (4) (2011) 387–396.

[34] C. Jia, Almost all short intervals containing prime numbers, Acta Arith. 76 (1) (1996) 21–84.