Logical Methods in Computer Science
Volume 19, Issue 4, 2023, pp. 16:1–16:34
https://lmcs.episciences.org/

Submitted Dec. 23, 2022
Published Nov. 28, 2023

COMPUTING THE DENSITY OF THE POSITIVITY SET FOR LINEAR RECURRENCE SEQUENCES

EDON KELMENDI

Queen Mary University of London
e-mail address: e.kelmendi@qmul.ac.uk

Abstract. The set of indices that correspond to the positive entries of a sequence of numbers is called its positivity set. In this paper, we study the density of the positivity set of a given linear recurrence sequence, that is the question of how much more frequent are the positive entries compared to the non-positive ones. We show that one can compute this density to arbitrary precision, as well as decide whether it is equal to zero (or one). If the sequence is diagonalisable, we prove that its positivity set is finite if and only if its density is zero. Further, arithmetic properties of densities are treated, in particular we prove that it is decidable whether the density is a rational number, given that the recurrence sequence has at most one pair of dominant complex roots.

Finally, we generalise all these results to symbolic orbits of linear dynamical systems, thereby showing that one can decide various properties of such systems, up to a set of density zero.

1. Introduction

Linear recurrence sequences (LRS) are infinite sequences of rational numbers $\langle u_n \rangle_{n \in \mathbb{N}}$, whose every entry is a linear combination of the k preceding entries. That is, a sequence that satisfies a recurrence relation:

$$u_n = a_1 u_{n-1} + \cdots + a_k u_{n-k}, \tag{1.1}$$

for all $n > k$, where $a_1, \ldots, a_k$ are rationals and $a_k \neq 0$. The constants $a_1, \ldots, a_k$, and $u_1, \ldots, u_k$ uniquely identify the sequence.

Firmly grounded as one of the fundamental families of finitely represented number sequences, they are ubiquitous in mathematics and computer science; their importance is evident. A basic object of study in modern number theory, they appear in the investigation of pseudo-random number generators, in cellular automata, as solutions of some Diophantine equations, as the number of $\mathbb{F}_q$-points on varieties, to name just a few examples. Furthermore, they are intrinsically related to linear dynamical systems, and the field of dynamical systems as a whole.

From another point of view, a linear recurrence sequence can be seen as a kind of restricted Turing machine, namely one that has a single loop inside which the variables are updated by a linear function. As such programs permeate any larger piece of software,

Key words and phrases: linear recurrence sequences, linear dynamical systems, density, positivity set.

DOI:10.46298/LMCS-19(4:16)2023

verifying their correctness has become increasingly important in recent years. This motivation has driven further interest in algorithmic questions regarding these sequences.

This field has been a rather active area of research in the past few decades — a considerable body of work has amassed. The wide-scoped monograph [EVDPS$^+$03] by Everest, van der Poorten, Shparlinski, and Ward is a place where one can find central results, their applications, as well as a taste of techniques that have proven useful. Here we recount only a brief summary of the theorems that are directly relevant to the present work.

We start with a basic question: What does the zero set of a linear recurrence sequence $\{n : u_n = 0\}$ look like? The wonderfully simple answer, provided in 1934 by Thoralf Skolem [Sko34] using p-adic analysis, is that the zero set of a linear recurrence sequence is a finite union of arithmetic progressions and a finite set. In other words, the zero set is ultimately periodic. This theorem was soon after generalised to sequences of algebraic numbers by Mahler [Mah35], and then later on by Lech, to sequences of members of any ring of characteristic zero [Lec53]. An elementary proof of Skolem's theorem can be found in [Han85], see also the discussion in Chapter 2.1 of [EVDPS$^+$03]. Unfortunately, even though we know the form of zero sets, we do not know how to decide if it is empty. Every known proof of this result uses, in some way or other, p-adic analysis, resulting in a non-constructive argument. The question of whether one can decide if there exists some n, such that $u_n = 0$, known as Skolem's problem, remains to this day, the central open problem for LRS.

However, there are some partial results for sequences of low order[1]: With the help of Baker's theorem for linear forms in logarithms of algebraic numbers, Mignotte, Shorey, and Tijdeman [TMS84, Theorem 2], and in parallel Vereshchagin [Ver85, Theorem 4], proved that for sequences of order at most four, one can decide whether their zero set is empty. In the direction of hardness, Skolem's problem is known to be NP-hard [BP02].

One can raise the same questions about the positivity set $\{n : u_n > 0\}$. This set, however, unlike the zero set, does not admit a clean description. In fact the positivity problem (is there some n such that $u_n > 0$) is more general than the Skolem problem. That is, there is a polynomial reduction from Skolem's problem to the positivity problem (with a quadratic increase in the order). The positivity problem is known to be decidable for LRS of order at most five [OW13], where Baker's theorem plays a crucial role again. In the direction of hardness, a decision procedure for the positivity problem for LRS of order six would allow one to compute the homogeneous Diophantine approximation type of a large class of transcendental numbers [OW13, Theorem 5.2]. Which suggests that such a procedure must come hand-in-hand with a deeper understanding — than hitherto exists — of Diophantine approximations of transcendental numbers.

Questions of asymptotic nature seem to be slightly more approachable. For example, one can decide if a sequence has infinitely many zeros [BM76, Theorem 2]. The corresponding problem for the positivity set, *i.e.* are there infinitely many n, for which $u_n > 0$ is not known to be decidable, however. This problem is called the ultimate positivity problem[2]. In fact, as for positivity, a similar link to Diophantine approximations exists [OW13, Theorem 5.1]. Nevertheless, there is an important positive result: namely that the ultimate positivity problem is decidable for *diagonalisable* LRS [OW14]. A sequence is diagonalisable if its

[1]The order of the sequence is the smallest k for which the sequence satisfies a recurrence like (1.1).

[2]Ultimate positivity is the question: "is it true that after some point every entry of the sequence is positive?". If we ignore the zeros, ultimate positivity does not hold if and only if the negativity set is infinite, or the positivity set of $\langle -u_n \rangle_{n \in \mathbb{N}}$ is infinite.

characteristic polynomial, which for a sequence that satisfies (1.1) is

$$x^k - a_1 x^{k-1} - \cdots - a_{k-1}x - a_k, \tag{1.2}$$

has no repeated roots. In fact, it is possible to go much further for diagonalisable sequences [AKK+21]: One can decide any asymptotic ω-regular property, even when the property itself is part of the input. For example, one can ask whether the sign pattern "-+-" occurs infinitely often in the sequence.

For the general case not much progress has been made however, it remains a long standing, difficult, open problem to decide anything about the positivity set of a general LRS, in particular whether this set is empty, or whether it is finite. In the present paper, we prove that it is possible to decide some things about another notion of size of a subset of naturals: its *density*.

Recall that the density of a set $S \subseteq \mathbb{N}$ is

$$\lim_{n\to\infty} \frac{|\{1, 2, \ldots, n\} \cap S|}{n},$$

where the vertical bars denote cardinality (note that the limit need not exist). The density is a notion used to measures how *large* an infinite subset of natural numbers is.

Example 1.1. Here is a trivial LRS: $u_1 = 1$ and $u_n = -u_{n-1}$. Clearly its positivity set are the odd numbers, and its density is equal to $1/2$.

Example 1.2. It is possible to construct linear recurrence sequences that are equal[3] to $\cos(n\theta)$, $n \in \mathbb{N}$. If θ is a rational multiple of π, the positivity set of these sequences will have some rational density, if however θ is not a rational multiple of π then, we will later see, that the density is equal to $1/2$.

The density of the positivity set of any linear recurrence sequence always exists. This fact was proved by Bell and Gerhold [BG07, Theorem 1], and is our principal starting point. With the exception of the paper above, to the best of our knowledge there is no other work that deals with the density of the positivity set. The paper [BM76] can however be interpreted as providing an algorithm to compute the density of the zero set.

We now describe the results of this paper. The first one is of a qualitative nature:

Theorem 1.3. *There is a procedure that inputs a* LRS *and decides whether the density of its positivity set is equal to 1.*

The same procedure can be used to decide whether the density is equal to 0, after a trivial pre-processing step.

Bell and Gerhold have observed, by using an equidistribution theorem due to Weyl, a version of which can be found in Cassels's book [Cas59], that the density is equal to the Lebesgue measure of a certain set. We proceed along the same path and go further by constructing this set, for which it is necessary to explicitly describe the multiplicative relations among the roots of the polynomial in (1.2). Afterwards, the problem is reduced to checking the emptiness of a semialgebraic set, which can be done using the decidability of the theory of real closed fields, *i.e.* Tarski's algorithm. These tools have been successfully

[3]We have defined LRS to be sequences of rational numbers, but one can define LRS over larger rings, as is done for this example, where the ring is $\mathbb{R}$. We chose the restriction to rationals for simplicity, although all the results of this paper can be proved for real algebraic numbers, at least.

employed by Ouaknine, Worrell, and others, on a number of related problems, it is not surprising that they prove useful to bear on the problems of this paper as well.

We will show that this problem is both NP and co-NP hard, while the procedure in Theorem 1.3 runs in PSPACE. When the order of the sequence is fixed, the complexity drops to PTIME.

Although we do not yet know how to decide whether the sequence has infinitely many positive entries, we can decide whether there are *many* of them, in the sense of having non-zero density, using Theorem 1.3. Another point of view is that the question "is the density 0?" over-approximates the question "is the positivity set finite?", because a positive answer to the latter implies the same for the former. However, for the family of diagonalisable sequences, the implication becomes an equivalence — the two questions are the same:

Theorem 1.4. *In a diagonalisable sequence the positivity set is finite if and only if its density is zero.*

Theorem 1.3 and Theorem 1.4 together imply the main theorem of [OW14], that ultimate positivity is decidable for diagonalisable LRS. However the proof has the same ingredients, in particular a result on the growth of LRS by Evertse, van der Poorten and Schlickewei, which is based on a lower bound for sums of S-units, itself based on the deep "subspace theorem" of Schmidt.

The main theorem of this paper says that we can compute densities to arbitrary precision:

Theorem 1.5. *There is a procedure that inputs a* LRS $\langle u_n \rangle_{n \in \mathbb{N}}$ *and a positive rational number* $\epsilon \in \mathbb{Q}$*, and computes some* δ'*, such that* $|\delta - \delta'| < \epsilon$*, where* δ *is the density of the positivity set of* $\langle u_n \rangle_{n \in \mathbb{N}}$*.*

The complexity is the same as for the density 1 problem; the problem is in PSPACE in the description of the LRS and $\lceil \epsilon^{-1} \rceil$, but it drops to polynomial time when the order of the sequence is fixed.

The idea of the proof of Theorem 1.5 is straightforward. We have to approximate the Lebesgue measure of a certain subset of the d-dimensional unit cube. To this end, we draw a grid of N^d points and count how many of these points fall in the set. It then remains to prove that we can decide whether a given rational point is a member of the set, and to upper bound the error term. For the latter we use a result of Koiran [Koi95]. We note that it is possible, instead of testing for every point whether it belongs to the set, to test it for fewer points that are picked randomly, resulting in a faster Monte-Carlo type algorithm.

Let us give a simple example that illustrates some of the ideas behind the theorems above.

Example 1.6. Consider the following simple program:

```
x=0; y=6; z=4;
while true do
    | { x := 4x + 3y
    | { y := 4y − 3x
    | { z := 5z
    | if y + z > 0 then
    |     | Region A
    | else
    |     | Region B
    | end
end
```

where the assignments to the local variables x, y, z are done in parallel.

It is not immediately evident from looking at this program that, for example, Region A is entered infinitely often. The algorithm from Theorem 1.5 can be used to conclude not only Region A is entered infinitely often, but that it is entered with frequency:

$$0.732279\ldots = \frac{\cos^{-1}(-2/3)}{\pi}.$$

At first sight, it might seem strange to see notions related to circles and triangles such as π and cos appearing in the answer of a simple question about a simple program, but the reality is that only through them can we understand the program above. Let us explain the answer in more detail. The value of $y + z$ in the n-th iteration of the loop is clearly equal to

$$\begin{pmatrix} 0 & 6 & 4 \end{pmatrix} \cdot \begin{pmatrix} 4 & -3 & 0 \\ 3 & 4 & 0 \\ 0 & 0 & 5 \end{pmatrix}^n \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

Multiplying this quantity with 5^{-n} will not change its sign and we see that after the multiplication, the update matrix is a rotation in the first two coordinates:

$$\begin{pmatrix} 4/5 & -3/5 & 0 \\ 3/5 & 4/5 & 0 \\ 0 & 0 & 1 \end{pmatrix}^n = \begin{pmatrix} \cos n\phi & -\sin n\phi & 0 \\ \sin n\phi & \cos n\phi & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where the angle ϕ is $\cos^{-1}(4/5)$. Multiplying this matrix with the row vector and the column vector above, we see that the variable $y + z$ in the n-th iteration of the loop has the same sign as $6\cos(n\phi) + 4$. So now the question: with what frequency does the loop enter Region A? has been reduced to the question: for how many n is $\cos(n\phi) > -2/3$? When the angle ϕ is not a rational multiple of π (which is the case here), by Weyl's equidistribution theorem, $n\phi$ is uniformly recurrent modulo π, meaning that for any interval I in $[0, \pi]$, the frequency with which $n\phi \mod \pi$ enters I is proportional to the size of I (that is length of the interval divided by π). As a consequence, since $\cos(n\phi) > -2/3$ if and only if $n\phi$ modulo π belongs to the interval $[0, \cos^{-1}(-2/3)]$, the answer follows.

Example 1.6 and Example 1.2 show that density can be both a rational and an irrational quantity. Therefore, the algorithm in Theorem 1.5 cannot *a priori* be used to decide

quantitative questions, such as whether the density is larger than some given rational. We give a partial result in this direction but leave the general case open:

Theorem 1.7. *There is a procedure that inputs a* LRS *that has at most one pair of dominant complex roots, and decides whether the density of its positivity set is rational, and if it is, computes it exactly.*

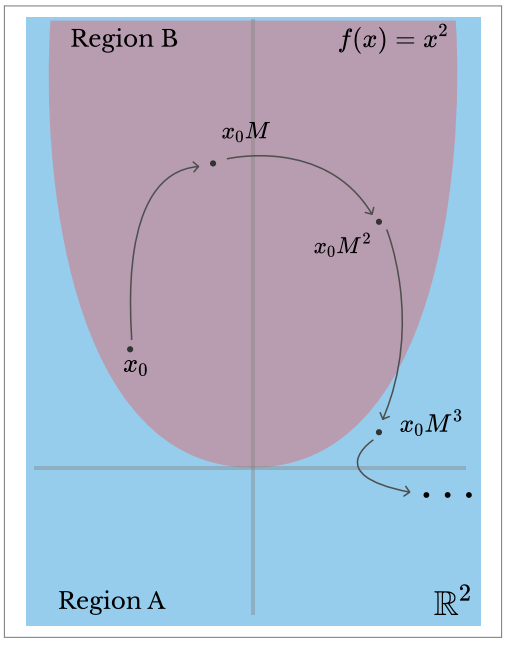
We also prove that when there are no (non-trivial) multiplicative relations among the dominant roots, the density is a *period*, as defined by Kontsevich and Zagier [KZ01]. We note that conjectures by Kontsevich and Zagier, and of Grothendieck predict the transcendence degree of field extensions of $\mathbb{Q}$ generated by a finite set of intervals, but we do not pursue this conjectural direction further.

Finally, we take a step back, and consider what makes the proofs of the theorems above work. One way of answering this question is to say that the *sign sequence* of a LRS is isomorphic (except in a set of density zero) to an ω-word that belongs to a family of words, which we call toric words. Such words, we prove, have some pleasant properties, one of which is that we can compute the frequencies with which any given pattern appears. Taking this point of view, allows us to generalise Theorems 1.3, 1.4, and 1.5, to linear dynamical systems.

A **linear dynamical system** is given via a square $k \times k$ matrix M with rational entries, and an initial point $x_0 \in \mathbb{Q}^k$. Its orbit is the sequence of points

$$x_0, x_0M, x_0M^2, x_0M^3, \dots. \tag{1.3}$$

In order to be able to illustrate an instance, suppose that $k = 2$ and consider the following question.



How frequently do members of the orbit in (1.3) enter Region B? More generally, suppose that we have a partition of the Euclidian space $\mathbb{R}^k$ into semialgebraic sets $S_1, \dots, S_\ell$. The latter are sets that one can define with polynomial inequalities, which we will define precisely later. The orbit (1.3) then defines an ω-word w over the alphabet $\{1, 2, \dots, \ell\}$, in the obvious way. This is a symbolic orbit of the dynamical system. One can then ask how frequently does a letter b appear in w? In other words, what is the density of the subset of indices n where $w_n = b$? This and slightly more general questions can be answered by studying toric words; in the sense that there are analogues of Theorems 1.3, 1.4, and 1.5. The last section, section 7, is devoted to these questions.

The rest of this paper is organised as follows. section 2 contains the principal definitions and generalities. section 3 is a technical section where we define a strong non-degeneracy

condition and split the sequence into subsequences that satisfy it, as a pre-processing step for the algorithms that follow. section 4 deals with the density 1 problem, as well as the analysis for diagonalisable sequences. In the section that follows we give the procedure to compute the density. In the end, in section 6, we give the proof of Theorem 1.7, deciding when the density is a rational number.

A preliminary version of this paper appeared in [Kel22].

Acknowledgment

I am grateful to James Worrell for many helpful discussions.

2. Sequences and Densities

A sequence $\langle u_n \rangle_{n\in\mathbb{N}}$ that satisfies a recurrence relation (1.1) for all $n > k$, but does not satisfy any linear recurrence with fewer terms, is called a LRS of **order** k. The **characteristic polynomial** of such sequence is the polynomial (1.2), whose roots are, say

$$\Lambda_1, \Lambda_2, \ldots, \Lambda_l,$$

assumed to be distinct, with respective multiplicities $m_1 \ldots, m_l$, where $1 \leq l \leq k$. The sequence $\langle u_n \rangle_{n\in\mathbb{N}}$ can be written as a **generalised power sum** (see [EVDPS$^+$03, Section 1.1.6]):

$$u_n = \sum_{i=1}^{l} f_i(n)\ \Lambda_i^n, \tag{2.1}$$

where the polynomials f_i have algebraic coefficients, $f_i \in \overline{\mathbb{Q}}[x]$, and the degree of f_i is $m_i - 1$. The converse also holds, any sequence $\langle u_n \rangle_{n\in\mathbb{N}}$ that can be written in the form (2.1) is a LRS over algebraic numbers. The sequences whose roots all have multiplicity 1, *i.e.* there are no repeated roots, are called **diagonalisable** (or simple) sequences.

A LRS is given by the numbers $a_1, \ldots, a_k$ and $u_1, \ldots, u_k$. From which, it is possible to compute descriptions of the constants in (2.1) in polynomial time in the bitlength of the input. By a **description** of an algebraic number we mean[4] a first-order formula that defines it, typically this is the number's minimal polynomial together with intervals specifying where its real and imaginary parts lie. To compute the descriptions of the roots, one runs a root isolation algorithm on the characteristic polynomial (to compute the approximating intervals), see for example [YS11] and [BPR06]. Afterwards, for the computation of polynomials f_i, one solves a system of linear equations of polynomial size in the input. All this can be done in polynomial time. As a consequence, we assume that we have computed the descriptions of every constant in (2.1), and that the roots are ordered by their modulus, *i.e.*

$$|\Lambda_i| \geq |\Lambda_{i+1}|.$$

Density (also referred to as natural density, or asymptotic density in the literature) is a notion that measures how large a subset $S \subseteq \mathbb{N}$ of natural numbers is. It is defined as:

$$\mathcal{D}(S) \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{|\{1, 2, \ldots, n\} \cap S|}{n}, \tag{2.2}$$

[4]There are other encodings of an algebraic number α. Mostly one uses the fact that a number field $\mathbb{Q}(\alpha)$ is a vector space of finite dimension. For our purposes however, it is more convenient to define algebraic numbers by first-order formulas over the reals (defined below).

where by the vertical bars we denote the cardinality of the set. Not every set has a density; the limit might not exist. However they do have lower and upper density, which are defined by replacing limit with lim inf and lim sup respectively.

Example 2.1. Here is the density of some simple subsets of natural numbers.

(1) An (infinite) arithmetic progression, with common differences d, has density $1/d$. If the set $S \subseteq \mathbb{N}$ is such that the difference between consecutive elements of S is at most d, then the lower density of S is larger than $1/d$.
(2) The squares $\{n^2 \ : \ n \in \mathbb{N}\}$ have density zero. To prove this, it suffices to observe that the cardinality of the squares in $\{1, 2, \ldots, n\}$ is in $O(\sqrt{n})$.
(3) The primes have density zero due to the prime number theorem.

The principal object of study in this paper is the density of the **positivity set**:

$$\mathcal{D}\left(\{n \ : \ u_n > 0\}\right)$$

of a given LRS $\langle u_n \rangle_{n \in \mathbb{N}}$. Bell and Gerhold proved that it always exists:

Theorem 2.2 [BG07, Theorem 1]**.** *The positivity set of any linear recurrence sequence has a density.*

The negativity set is just the positivity set of the sequence $\langle -u_n \rangle_{n \in \mathbb{N}}$ (which is plain, from (2.1) and the discussion above, that it can be computed). Therefore in the rest of this paper, we only deal with the density of the positivity set, which is simply referred to as the **density of the sequence**.

We will make ample use of procedures for deciding the first-order **theory of real closed fields**, proved by Tarski [Tar51]. In this logic the atomic formulas are

$$f(x_1, \ldots, x_n) \geq 0,$$

where $f \in \mathbb{Z}[x_1, \ldots, x_n]$ is a polynomial with integer coefficients. The atomic formulas can be connected with Boolean connectives, and one is allowed to quantify over real numbers. Subsets of $\mathbb{R}^n$ defined by such formulas are called **semialgebraic** sets. In the paper cited above, Tarski proved that there exists a procedure that inputs a first-order sentence and decides whether it is true when interpreted over the reals.

We can also interpret such formulas over the complex numbers instead of the reals, using the embedding of $\mathbb{C}^n$ to $\mathbb{R}^{2n}$, handling the real and imaginary parts individually.

Note that our definition of descriptions of algebraic numbers is a simple formula in Tarski's logic. Other formulas that we will construct will be equally simple in the following sense: they will belong to the existential fragment, *i.e.* formulas of the type

$$\exists x_1 \exists x_2 \cdots \exists x_n \qquad \Phi(x_1, \ldots, x_n),$$

where Φ is quantifier-free. The complexity of this fragment is relatively low:

Theorem 2.3 ([Can88, Theorem 3.3] and [Ren92, Theorem 1.1], respectively)**.** *The existential theory of reals is decidable in* PSPACE*. When the number of variables is fixed, the complexity drops to* PTIME[5]*.*

The theorems above expect the polynomials in the input to be written as a sequence of coefficients, each encoded in binary. Hence the exponents are assumed to be encoded in unary.

[5]The PTIME upper bound holds for the full logic, when the number of variables is fixed, not only the existential fragment.

3. Strongly Non-Degenerate Subsequences

Let $P \in \mathbb{N}$, and consider subsequences of the form:

$$\{\langle u_{nP+\ell}\rangle_{n\in\mathbb{N}} \ : \ 0 \leq \ell < P\}\,. \tag{3.1}$$

Each one is itself a LRS ([EVDPS+03, Theorem 1.3]). One can easily observe this fact from the equality (2.1): the roots of the subsequence are Λ_i^P and the polynomials $f_i(nP+\ell)$ are multiplied by the constant Λ_i^ℓ.

The purpose of this section is a crucial preprocessing step that splits the sequence into subsequences (3.1), for a particular $P \in \mathbb{N}$, which we will compute. The subsequences have a number of properties (enumerated in a lemma at the end of the section) that make them more amenable. Effectively dividing the initial problem into easier sub-problems, we can recombine answers of the sub-problems to get the answer for the initial problem. For example, if we know the densities of the P subsequences, then the density of the original sequence is equal to their sum divided by P. Or for the density 1 problem: the original sequence has density 1 if and only if all the subsequences have density 1.

In our case the period P is a product:

$$P \stackrel{\mathsf{def}}{=} P_1 \cdot P_2,$$

where P_1 comes from degeneracy, and P_2 from multiplicative relations among the roots. Let N be the bitlength of the input and k the order of the sequence, later in this section we will prove that P will have the upper bound:

$$P \in 2^{\mathcal{O}(k^5 \log\log N)}. \tag{3.2}$$

Before we give the definitions of the periods P_1 and P_2, let us first discuss the description of the roots Λ_i^P, as this is important for the complexity upper bounds when P is large. Let $r \in \mathbb{N}$, and let $z \in \overline{\mathbb{Q}}$ be an algebraic number with description $\phi(x)$ (*i.e.* the formula $\phi(x)$ holds if and only if $x = z$). There are two ways to describe the number z^r:

(1) The *trivial way*: saying that there exists some x such that $\phi(x)$ and

$$y = \underbrace{x \cdot x \cdots x}_{r \text{ times}}.$$

Resulting in a constant increase on the number of variables, and a linear increase in r on the size of the formula.

(2) The *repeated squaring way*: saying that there exist a roughly $s := \log r$ number of variables $x_1, \ldots, x_s$ such that

$$\phi(x_1) \text{ and } y = x_s \text{ and } x_{i+1} = x_i \cdot x_i,\ 1 \leq i \leq s.$$

Resulting in a $\log r$ increase in both the number of variables and the size of the formula.

We will use both methods, depending on which complexity bound we want to derive.

Proposition 3.1. *For a given constant P, bounded by (3.2), the description of any Λ_i^P can be computed in polynomial time. Furthermore, such a description grows both in the number of variables and in size by a term in $\mathcal{O}(k^5 \log\log N)$.*

When the order of the sequence is fixed, the size of the description grows by a term in $\mathcal{O}(\log N)$ while the number of variables by a constant.

Proof. Using the repeated squaring method results in a formula that grows both in size and in the number of variables by a $\log P$ term, hence the first statement of the proposition.

When k, the order of the sequence is fixed however, it makes more sense to use the trivial way of constructing the formula, because this will result in a constant increase in the number of variables, and a linear in P increase in the size of the formula. Since for fixed k, P is in $\mathcal{O}(\log N)$ the second statement of the proposition follows. □

Now we define P_1 and P_2, as well as show how to compute them. In the end of this section we summarise the properties that every subsequence $\langle u_{nP+\ell}\rangle_{n\in\mathbb{N}}$ has.

3.1. **Period P_1.** We begin with the standard notion of degeneracy. A LRS is said to be **degenerate** if it has two distinct roots Λ_i and Λ_j, whose ratio Λ_i/Λ_j is a root of unity. One can test in PTIME whether a given sequence is degenerate by checking whether any of its ratios of distinct roots satisfies a cyclotomic polynomial of appropriate degree. Consult Section 3 in [YLN95]. If the sequence is degenerate, taking the least common multiple of all the orders of roots of unity that can occur in this way, we get a quantity P_1, such that all the subsequences with period P_1 are either identically zero, or non-degenerate. The quantity P_1 is upper bounded only by a function in the order of the sequence:

Theorem 3.2 [EVDPS$^+$03, Theorem 1.2]**.** *Let $\langle u_n\rangle_{n\in\mathbb{N}}$ be a* LRS *of order k. Then there is a constant*

$$M_k \in 2^{\mathcal{O}(k\sqrt{\log k})},$$

such that for some $P_1 \leq M_k$, each subsequence

$$\langle u_{nP_1+\ell}\rangle_{n\in\mathbb{N}},$$

$0 \leq \ell < P_1$ is either identically zero, or is non-degenerate.

3.2. **Period P_2.** The definition of P_2 requires a little bit more work. We have assumed that the roots are ordered by their modulus: $|\Lambda_i| \geq |\Lambda_{i+1}|$, suppose that the first j ones are dominant, *i.e.*,

$$|\Lambda_1| = \cdots = |\Lambda_j| > |\Lambda_{j+1}|.$$

Let d be the maximal degree of the polynomials $f_1, \ldots, f_j$ from (2.1), and suppose, without loss of generality, that it is exactly the polynomials $f_1, \ldots, f_m$ that are of degree d, for some $m \leq j$. Define the normalised roots:

$$\lambda_i \stackrel{\mathsf{def}}{=} \frac{\Lambda_i^{P_1}}{\left|\Lambda_i^{P_1}\right|} \qquad 1 \leq i \leq m.$$

We are interested in the multiplicative relations:

$$\mathcal{M}(\lambda_1, \ldots, \lambda_m) \stackrel{\mathsf{def}}{=} \{\vec{r} \in \mathbb{Z}^m \ : \ \lambda_1^{r_1}\lambda_2^{r_2}\cdots\lambda_m^{r_m} = 1\}.$$

This set with addition forms a subgroup of $\mathbb{Z}^m$. Since the latter is a free abelian group with a basis of m elements, by [Lan02, Theorem 7.3, Chapter I] the subgroup $\mathcal{M}$ is a free abelian group with some basis

$$\vec{b}_1, \ldots, \vec{b}_v \in \mathbb{Z}^m, \tag{3.3}$$

where $v \leq m$. Define

$$P_2 \stackrel{\text{def}}{=} 2 \prod |b_{s,t}|, \tag{3.4}$$

where the product is taken over all $1 \leq s \leq v$, and $1 \leq t \leq m$, for which $b_{s,t}$ is nonzero.

Lemma 3.3. *The integer P_2 is effective. It can be computed in* PSPACE. *When the order of the sequence is fixed, the computation can be performed in* PTIME.

Proof. We argue that we can compute the basis (3.3) and hence also P_2.

It follows from [vdPL77, Theorem 1], that there is an effective upper bound on the absolute value of the coordinates of the basis (3.3) of size:

$$2^{\mathcal{O}(k^2)} \prod_{i=2}^{m} \log H(\lambda_i),$$

where H is the Mahler measure, defined as follows. For an algebraic number $z \in \overline{\mathbb{Q}}$, with minimal polynomial

$$a_0 x^d + a_1 x^{d-1} + \cdots + a_d = a_0 (x - z_1) \cdots (x - z_d),$$

we say that its Mahler measure is:

$$H(z) \stackrel{\text{def}}{=} |a_0| \prod_{i=1}^{d} \max\{1, |z_i|\} \leq \sqrt{d} \max_{0 \leq i \leq d} |a_i|,$$

where the upper bound comes from [vdPL77, Lemma 1]. Using the fact that for any algebraic number $z \in \overline{\mathbb{Q}}$ and $r \in \mathbb{N}$, $H(z^r) = H(z)^r$, whose proof can be found in [Wal00, Chapter 3], via a straightforward computation, we can derive the following upper bound:

$$\max_{\substack{1 \leq s \leq v \\ 1 \leq t \leq m}} |b_{s,t}| \in 2^{\mathcal{O}(k^3 \log\log N)}, \tag{3.5}$$

where k is the order of the sequence and N is the bitlength of the input. For any $\vec{b} \in \mathbb{Z}^m$ with the same upper bound, the assertion

$$\vec{b} \in \mathcal{M}(\lambda_1, \ldots, \lambda_m),$$

is an existential first-order formula of polynomial size in N, due to Proposition 3.1. Which means that by brute force, we can compute a basis (3.3) in PSPACE by using the algorithm from Theorem 2.3. When the order k is fixed, the number of variables is constant. As a consequence of the second statement of Theorem 2.3, in this scenario, the basis can be computed in PTIME. □

From the definition of P_2, (3.4), the estimate (3.5) and Theorem 3.2, one can derive the upper bound (3.2).

3.3. **Properties of the Subsequences.** Let $0 \leq \ell < P$, we list a number of properties of the subsequence

$$\langle u_{nP+\ell} \rangle_{n \in \mathbb{N}}, \tag{3.6}$$

which we assume is not identically zero. We start by replacing the dependent roots as follows.

The only case when the group $\mathcal{M}(\lambda_1, \ldots, \lambda_m)$ is trivial is when $m = 1$, which implies that $\lambda_1 = 1$, because complex roots come as conjugate pairs (of the same multiplicity), and being a conjugate pair is a multiplicative relation (for algebraic numbers on the unit circle). In this case, every problem that we treat becomes trivial. Therefore suppose that $m > 1$. Then there exists some member of the basis (3.3) — say $\vec{b}_1$ without loss of generality — that has at least two non-zero coordinates. By definition,

$$\lambda_1^{b_{1,1}} \cdots \lambda_m^{b_{1,m}} = 1.$$

Suppose that $b_{1,m} \neq 0$. By using Euler's formula we see that we can write:

$$\lambda_m = \varrho \lambda_1^{-b_{1,1}/b_{1,m}} \cdots \lambda_{m-1}^{-b_{1,m-1}/b_{1,m}}, \tag{3.7}$$

where ϱ is a $b_{1,m}$-th root of unity (and hence also a P_2-th root of unity, by definition of P_2), and at least one of the exponents $b_{1,1}, \ldots, b_{1,m-1}$ is nonzero. Replacing λ_m in the other equations, and continuing in this manner, making at most v replacements, one for every member of the basis, we conclude that the set of indices $\{1, \ldots, m\}$ can be partitioned into the indices corresponding to the independent roots and depended roots of the form (3.7), more precisely it can be partitioned into subsets:

- I - a non-empty subset, with independent λ_i, *i.e.* that do not have multiplicative relations among themselves,
- D - a subset with dependent λ_i, *i.e.* those that can be written in the form (3.7), where in the right-hand side only members of I appear, and there is a factor ϱ which is a P_2-th root of unity (perhaps not primitive)[6], and
- U - an empty set or a singleton containing some i for which $\lambda_i = 1$.

The reason why U has cardinality at most 1 is as follows. By the process described above, we cannot obtain more than one equation of the type $\lambda_i^r = 1$, because among $\lambda_1, \ldots, \lambda_m$, the only root of unity that can appear is the number 1. Indeed, if there were some complex λ_i that is r-th root of unity, then its complex conjugate $\overline{\lambda_i}$ will also appear among the dominant roots $\lambda_1, \ldots, \lambda_m$ (with the same multiplicity), and $(\lambda_i/\overline{\lambda_i})^r = \lambda_i^{2r} = 1$, meaning that the sequences $\langle u_{nP_1+\ell} \rangle_{n \in \mathbb{N}}$ are degenerate, a contradiction of Theorem 3.2.

Rearrange the the roots λ_i such that for some η

$$I = \{1, \ldots, \eta\}, \qquad D = \{\eta + 1, \ldots, m - 1\}, \qquad U = \{m\}.$$

The case when D or U is empty is omitted, as it can be treated in essentially the same way. It is convenient to define for all i, $1 \leq i \leq m$:

$$\alpha_i \stackrel{\text{def}}{=} \lambda_i^{P_2} = \frac{\Lambda_i^P}{|\Lambda_i^P|},$$

[6]Here we see the reason behind the definition of P_2: In subsequences with the period P_2 we can directly write the dependent roots as a function of the independent ones; the factor ϱ disappears because it is a P_2-th root of unity.

and the rationals $q_{i,j} \in \mathbb{Q}$, $i \in D$, $j \in I$, such that:

$$\alpha_i = \prod_{j \in I} \alpha_j^{q_{i,j}}.$$

The generalised power sum form of the sequence (3.6) is:

$$u_{nP+\ell} = \sum_{i=1}^{l} \Lambda_i^{\ell} f_i(nP+\ell)(\Lambda_i^P)^n.$$

Dividing by $n^d|\Lambda_1^P|^n$ does not change the sign, where d is the largest degree of polynomials multiplying the dominant roots. We get the sequence:

$$\begin{aligned} v_n &\stackrel{\mathsf{def}}{=} \sum_{i=1}^{m} c_i \alpha_i^n + R(n) \\ &= \sum_{i \in I} c_i \alpha_i^n + \sum_{i \in D} c_i \prod_{j \in I} \alpha_j^{q_{i,j}} + c_m + R(n), \end{aligned} \tag{3.8}$$

where $c_i \in \overline{\mathbb{Q}}$, and $R(n)$ is some residue that tends to zero polynomially, *i.e.*

$$|R(n)| \in \mathcal{O}(n^{-\xi}), \text{ for some } \xi > 0. \tag{3.9}$$

Furthermore there are no multiplicative relations among the roots α_i, for $i \in I$, that is:

$$\mathcal{M}(\alpha_1, \ldots, \alpha_\eta) = \left\{\vec{0}\right\}. \tag{3.10}$$

A non-degenerate LRS whose signs are the same as some sequence that can be written like v_n above is what we call **strongly non-degenerate**. We summarise the properties of subsequences $\langle u_{nP+\ell}\rangle_{n\in\mathbb{N}}$.

Lemma 3.4. *For any ℓ, $0 \leq \ell < P$, the following statements are true for the sequence $\langle u_{nP+\ell}\rangle_{n\in\mathbb{N}}$ that is not identically zero:*

1. *is non-degenerate,*
2. *has finitely many zeros,*
3. *its entries have the same sign as the entries of $\langle v_n\rangle_{n\in\mathbb{N}}$ defined in* (3.8)*,*
4. *the description of the algebraic numbers c_i, α_i, and $q_{i,j}$ are of polynomial size, have polynomial many variables, and can be computed in* PSPACE*,*
5. *when the order of the sequence is fixed, the descriptions of the numbers above are of polynomial size, with a constant number of variables, and can be computed in* PTIME*.*

Proof. Property 1 comes from the fact that P_1 divides P and Theorem 3.2. Any non-degenerate sequence that is not identically zero has finitely many zeros [EVDPS$^+$03, Section 2.1], hence Property 2. The third property holds because we have obtained the sequence $\langle v_n\rangle_{n\in\mathbb{N}}$ by dividing with positive numbers.

To see that Property 4 holds for the roots α_i, first observe that P can be computed in PSPACE, as a consequence of Lemma 3.3, and the discussion in subsection 3.1. Then applying Proposition 3.1 gives the wanted conclusion. One makes a similar argument for the constants c_i. As for the rationals $q_{i,j}$, a combination of two facts is used. First, note that in the proof of Lemma 3.3 the basis (3.3) is being computed in PSPACE. Second, in the procedure that computes these rationals, described above, we do at most v^2 many replacements where $v \leq m$ is the size of the basis. Each such replacement can be done in PTIME.

For the last property, when k, the order of the sequence is fixed, the constant P is in $\mathcal{O}(\log N)$, and it can be computed in PTIME, due to Lemma 3.3. Note that in this case P_1 is constant. The property then follows by the same argument as for Property 4. □

Example 1.6 is not very interesting with respect to this section because after division by 5^n it is already in a strongly non-degenerate form. Here is a more suitable example.

Example 3.5. Let α be an algebraic number in the unit circle, for example:

$$\alpha \stackrel{\text{def}}{=} \frac{3}{5} + i\frac{4}{5},$$

and define:

$$\lambda_1 \stackrel{\text{def}}{=} \alpha^5, \qquad \lambda_2 \stackrel{\text{def}}{=} \alpha^3 \overbrace{\left(\frac{\sqrt{5}-1}{4} + i\frac{\sqrt{10+2\sqrt{5}}}{4}\right)}^{\varrho}.$$

Then one can come up with a LRS $\langle u_n \rangle_{n\in\mathbb{N}}$ over the real algebraic numbers[7], whose characteristic polynomial, split into linear factors, is:

$$f(x) \stackrel{\text{def}}{=} (x-\lambda_1)^2(x-\overline{\lambda}_1)^2(x-\lambda_2)^2(x-\overline{\lambda}_2)^2(x-1/2).$$

The sequence $\langle u_n \rangle_{n\in\mathbb{N}}$ in power sum form will look like:

$$u_n = (a_1 + na_2)\ \lambda_1^n + (\overline{a}_1 + n\overline{a}_2)\ \overline{\lambda}_1^n + (b_1 + nb_2)\ \lambda_2^n + (\overline{b}_1 + n\overline{b}_2)\ \overline{\lambda}_2^n + c\ 2^{-n},$$

for some algebraic numbers a_1, a_2, b_1, b_2, c. First let us isolate the dominant terms, to this end, since $|\lambda_i| = 1$, just divide the equality above by n, to get

$$a_2\lambda_1^n + \overline{a}_2\overline{\lambda}_1^n + b_2\lambda_2^n + \overline{b}_2\overline{\lambda}_2^n + R(n),$$

where the remainder $R(n)$ tends to zero as $n \to \infty$. Since the ratio of λ_i is a nonzero power of α (times ϱ or ϱ^{-1}) it cannot be a root of unity. Hence the sequence is non-degenerate, *i.e.* $P_1 = 1$. However, it is not strongly non-degenerate. Indeed, since ϱ is a fifth root of unity, there is a multiplicative relationship between λ_1 and λ_2, which is

$$\lambda_1^3 = \lambda_2^5.$$

So $(3, 0, -5, 0)$ and $(0, 3, 0, -5)$ form a basis of the subgroup of multiplicative relationships, hence P_2 in this case is equal to 450. So we look at the strongly non-degenerate subsequences $u_{450n+\ell}$, where $\ell \in \{0, \ldots, 449\}$. Define $a'_2 := a_2\lambda_1^\ell$, $b'_2 := b_2\lambda_2^\ell$, and $\gamma_i = \lambda_i^{450}$. Then we have:

$$u_{450n+\ell} = a'_2\gamma_1^n + \overline{a'}_2\overline{\gamma}_1^n + b'_2\gamma_2^n + \overline{b'}_2\overline{\gamma}_2^n + R(450n+\ell).$$

Finally, since ϱ is a primitive fifth root of unity, and therefore also a 450th root of unity we may write

$$\gamma_2 = \gamma_1^{3/5},$$

and with this replacement the equation above becomes:

$$u_{450n+\ell} = a'_2\gamma_1^n + \overline{a'}_2\overline{\gamma}_1^n + b'_2\gamma_1^{3n/5} + \overline{b'}_2\overline{\gamma}_1^{3n/5} + R(450n+\ell).$$

[7]This LRS is deliberately defined over the ring $\overline{\mathbb{Q}} \cap \mathbb{R}$ in order to keep the example small.

4. The Density 1 Problem

In this section we prove that it is decidable whether the density of a given sequence is equal to 0. The procedure expects a strongly non-degenerate sequence as input, *i.e.* a sequence of the form in (3.8) with the properties that are listed in Lemma 3.4. Suppose that we are given such a sequence and let δ be its density.

Note that the density of the *negativity* set of the sequence (which is the same as the density of $\langle -v_n \rangle_{n\in\mathbb{N}}$) is equal to $1-\delta$, because the zeros $\langle v_n \rangle_{n\in\mathbb{N}}$ do not affect the density, being finitely many; a consequence of Property 2 in Lemma 3.4. Hence the density of the sequence $\langle v_n \rangle_{n\in\mathbb{N}}$ is 0 if and only if the density of $\langle -v_n \rangle_{n\in\mathbb{N}}$ is 1. Thus the two problems, "is the density 1?" and "is the density 0?" are inter-reducible.

The argument for decidability of the density 0 problem is as follows. We define two open and measurable sets $\mathcal{P}$ and $\mathcal{Q}$ such that

$$\mathcal{P} = \emptyset \qquad \Leftrightarrow \qquad \mathcal{Q} = \emptyset, \tag{4.1}$$

and furthermore

$$\mathcal{Q} \text{ is semialgebraic} \qquad \text{and} \qquad \delta = \mu(\mathcal{P}), \tag{4.2}$$

where μ denotes the Lebesgue measure. Being open sets, it follows that $\delta > 0$ if and only if the semialgebraic set $\mathcal{Q}$ is nonempty, which can be decided, in particular because of Theorem 2.3. In this way decidability of the density 1 problem, *i.e.* Theorem 1.3, will follow from (4.1) and (4.2), as well as the reduction from the density 1 to the density 0 problem.

We proceed with the definitions of the sets $\mathcal{P}$ and $\mathcal{Q}$. Let $\mathbb{T}$ be the unit circle, *i.e.* the set of complex numbers $z \in \mathbb{C}$, for which $|z| = 1$. Define the auxiliary functions F and G which are $v_n - R(n)$ but the roots α_i are replaced by variables; more precisely F is a map from $[0,1]^\eta$ to the reals, and G a map from $\mathbb{T}^\eta$ to the reals, defined as:

$$F(\vec{\varphi}) \stackrel{\text{def}}{=} \sum_{i=1}^{\eta} c_i \exp(2\pi\mathbf{i}\ \varphi_i) + \sum_{i=\eta+1}^{m-1} c_i \exp\left(2\pi\mathbf{i} \sum_{j=1}^{\eta} q_{i,j}\varphi_j\right) + c_m,$$

$$G(\vec{z}) \stackrel{\text{def}}{=} \sum_{i=1}^{\eta} c_i z_i \qquad\qquad + \sum_{i=\eta+1}^{m-1} c_i \prod_{j=1}^{\eta} z_i^{q_{i,j}} \qquad\qquad + c_m.$$

Now the sets $\mathcal{P}$ and $\mathcal{Q}$ are defined as:

$$\mathcal{P} \stackrel{\text{def}}{=} \{\vec{\varphi} \in [0,1]^\eta \ : \ F(\vec{\varphi}) > 0\},$$

$$\mathcal{Q} \stackrel{\text{def}}{=} \{\vec{z} \in \mathbb{T}^\eta \ : \ G(\vec{z}) > 0\}.$$

As one can obtain $\mathcal{P}$ by applying $\log z/2\pi\mathbf{i}$ component-wise to elements of $\mathcal{Q}$, it is plain that $\mathcal{P}$ is non-empty if and only if $\mathcal{Q}$ is non-empty. Since $\mathcal{P}$ is open, it has non-zero measure if and only if it is non-empty. Furthermore, $\mathcal{Q}$ is semialgebraic, thus it only remains to show that $\delta = \mu(\mathcal{P})$.

The proof follows closely the proof of the main theorem of [BG07], and is crucially based on the following theorem, originally due to Weyl [Wey16, Satz 4], though we give a more modern reference from the book of Cassels.

Theorem 4.1 [Cas59, Theorem 1, page 64]**.** *Let $\theta_1, \ldots, \theta_k, 1 \in \mathbb{R}$ be linearly independent over $\mathbb{Q}$, and $S \subseteq [0,1]^k$ a measurable set, then*

$$\mathcal{D}\big(\{n \ : \ (n\theta_1 \mod 1, \ldots, n\theta_k \mod 1) \in S\}\big) = \mu(S).$$

It says that the fractional parts of $n\vec{\theta}$ fall in the set S with frequency that is equal to the measure of the set S, in other words they are uniformly distributed in the k-dimensional cube.

For $i \in \{1, \ldots, \eta\}$, define the arguments of the roots:

$$\theta_i \stackrel{\text{def}}{=} \frac{\log \alpha_i}{2\pi\mathbf{i}} \in [0, 1].$$

Since there are no multiplicative relations among the $\alpha_1, \ldots, \alpha_\eta$, from (3.10), we have that $\theta_1, \ldots, \theta_\eta, 1$ are linearly independent over $\mathbb{Q}$. To see this, write $\alpha_i = \exp(2\pi\mathbf{i}\theta_i)$ and observe that there are no multiplicative relations among the α_i if and only if there is no linear combination over $\mathbb{Q}$ of θ_i that is equal to an integer. As a consequence, we note that Theorem 4.1 is applicable to the tuple $\theta_1, \ldots, \theta_\eta$.

The proof of $\delta = \mu(\mathcal{P})$ is preceded by two lemmas. The first one says that the set of points that F maps to 0 has measure 0.

Lemma 4.2. $\mu\big(\{\vec{\varphi} \ : \ F(\vec{\varphi}) = 0\}\big) = 0.$

Proof. Since any generalised power sum is a LRS over $\overline{\mathbb{Q}}$ [EVDPS$^+$03, Section 1.1.6], the sequence

$$\langle F(n\vec{\theta})\rangle_{n\in\mathbb{N}} = \langle v_n - R(n)\rangle_{n\in\mathbb{N}}$$

is a non-degenerate LRS. As a corollary of the Skolem-Mahler-Lech theorem [EVDPS$^+$03, Section 2.1], this sequence has finitely many zeros, so

$$\mathcal{D}\Big(\Big\{n \ : \ F(n\vec{\theta}) = 0\Big\}\Big) = 0.$$

As noted above, we can apply Theorem 4.1 to $\vec{\theta}$, which implies

$$\mathcal{D}\Big(\Big\{n \ : \ F(n\vec{\theta}) = 0\Big\}\Big) = \mu\big(\{\vec{\varphi} \ : \ F(\vec{\varphi}) = 0\}\big),$$

where the set on the right-hand side is clearly measurable. Combining these two equations yields the statement of the lemma. □

This lemma can also be proved without appealing to the Skolem-Mahler-Lech theorem, by directly showing that the set that is being measured has empty interior.

The second lemma says that the indices in which the residue $R(n)$ is larger in absolute value than the dominating terms of the sequence, have upper density 0. This means that it is only the dominant part that plays any role on the density δ. Denote by $\hat{\mathcal{D}}$ the upper density (same as density except that the limit is replaced by lim sup): for all $S \subset \mathbb{N}$,

$$\hat{\mathcal{D}}(S) \stackrel{\text{def}}{=} \limsup_{n\to\infty} \frac{|\{1, 2, \ldots, n\} \cap S|}{n}.$$

Lemma 4.3. $\hat{\mathcal{D}}\Big(\Big\{n \ : \ |F(n\vec{\theta})| < |R(n)|\Big\}\Big) = 0.$

Proof. For $\epsilon > 0$, define:

$$\mathcal{P}_\epsilon \stackrel{\text{def}}{=} \{\vec{\varphi} \in [0, 1]^\eta \ : \ |F(\vec{\varphi})| \le \epsilon\},$$
$$\mathcal{R}_\epsilon \stackrel{\text{def}}{=} \Big\{n \in \mathbb{N} \ : \ |F(n\vec{\theta})| \le \epsilon\Big\}.$$

The residue $|R(n)|$ tends to zero as n gets larger (3.9), hence for all $\epsilon > 0$,

$$\hat{\mathcal{D}}\big(\left\{n \ : \ |F(n\vec{\theta})| < |R(n)|\right\}\big) \leq \mathcal{D}(\mathcal{R}_\epsilon). \tag{4.3}$$

The set $\mathcal{R}_\epsilon$ has density as a consequence of Theorem 4.1, also

$$\mathcal{D}(\mathcal{R}_\epsilon) = \mu(\mathcal{P}_\epsilon) = \int_{[0,1]^\eta} \mathbb{1}_{\mathcal{P}_\epsilon} d\mu,$$

where by $\mathbb{1}_{\mathcal{P}_\epsilon}$ we have denoted the indicator function of the set $\mathcal{P}_\epsilon$. Almost everywhere the function $\mathbb{1}_{\mathcal{P}_\epsilon}$ tends to $\mathbb{1}_{\mathcal{P}_0}$ as $\epsilon \to 0$, hence by Lebesgue's dominated convergence theorem [Bil08, Theorem 16.4] we have

$$\int_{[0,1]^\eta} \mathbb{1}_{\mathcal{P}_\epsilon} d\mu \to \int_{[0,1]^\eta} \mathbb{1}_{\mathcal{P}_0} d\mu = 0,$$

where the equality to zero comes from Lemma 4.2. Since (4.3) holds for all $\epsilon > 0$, the statement of the lemma follows. □

One consequence of Lemma 4.3 is that,

$$\delta = \mathcal{D}\big(\{n \ : \ v_n > 0\}\big) = \mathcal{D}\big(\left\{n \ : \ F(n\vec{\theta}) > 0\right\}\big).$$

The density on the right-hand side is equal to $\mu(\mathcal{P})$ by again applying Theorem 4.1.

Thus we have proved Theorem 1.3, that it is possible to decide whether the density is equal to 0 (or to 1). The complexity of the procedure is in PSPACE: the formula for non-emptiness of $\mathcal{Q}$ is of polynomial size due to Property 4 of Lemma 3.4, and hence whether it is true can be decided in PSPACE, Theorem 2.3.

The procedure runs in PTIME if the order of the sequence is fixed. This follows from Property 5 of Lemma 3.4 and Theorem 2.3.

Note that this lemma, Lemma 4.3, summarises the reason why we are able to decide certain properties of LRS *up to* a set of indices that has density zero. For a general LRS it is rather difficult to understand for which indices $n \in \mathbb{N}$, the dominant part v_n is larger in absolute value than the absolute value of the remainder $|R(n)|$. Indeed this is the source of complications due to which we do not yet know whether the Skolem, positivity or ultimate positivity problems are decidable. It requires a deep understanding of certain arithmetic properties of the algebraic numbers α_i. However, Lemma 4.3 says that the indices n for which the dominant part is smaller, form a subset of $\mathbb{N}$ that has density zero, therefore in matters of density, these indices that are hard to understand have no effect.

4.1. **Complexity Lower Bounds.** It is possible to re-purpose the proofs of [BP02] and [OW14] to show that the density 1 problem is both NP and co-NP hard. This indicates that the problem lies somewhere above these two classes, and is possibly PSPACE-complete.

Theorem 4.4. *The density 1 problem is* NP*-hard.*

Proof. In essence, we will show that the proof of Blondel and Portier in [BP02], also implies the statement of the theorem. It works as follows.

An instance of 3-SAT is a Boolean formula in variables $x_1, \ldots, x_n$ of the form:

$$C_1 \wedge C_2 \wedge \cdots \wedge C_m, \tag{4.4}$$

where each C_i is the disjunction of exactly three terms, where a term is either x_i or $\neg x_i$, $i \in \{1, 2, \ldots, n\}$. The 3-SAT problem is NP-hard, and will be the problem we reduce from.

The first reduction is into another problem, one about regular languages, which we describe now.

Fix a unary alphabet $\Sigma := \{a\}$. A regular expression over this alphabet is built using the empty word ϵ, words a^n, $n \in \mathbb{N}$, union, and Kleene star. Here is an example of such a regular expression:

$$\epsilon \cup aaa \cup (\epsilon \cup aa)^*.$$

There is a polynomial reduction from 3-SAT to the problem that inputs such a regular expression and decides whether the language that it describes is different from the language a^*. The reduction is as follows.

Compute $p_1, \ldots, p_n$ the first n prime numbers, which can be done in polynomial time (and they are all smaller than n^2). Define the function $h : \mathbb{N} \to \mathbb{N}^n$ that maps

$$k \mapsto (k \bmod p_1, k \bmod p_2, \ldots, k \bmod p_n),$$

where by $k \bmod p_i$ we denote the residue after dividing k by p_i. Call a natural number k a **code**, if and only if $h(k) \in \{0,1\}^n$.

There is a regular expression E_0 such that the word a^k belongs to the language E_0 describes, $L(E_0)$, if and only if k is *not* a code. That is the expression:

$$E_0 \stackrel{\text{def}}{=} \bigcup_{i=1}^{n} \bigcup_{j=2}^{p_i - 1} a^j (a^{p_i})^*.$$

Now let C be one of the conjuncts in (4.4), and suppose that it involves the variables x_r, x_s, x_t. Consider a manner of setting bits x_r, x_s, x_t such that the conjunct C becomes false, *e.g.* respectively $(x_r, x_s, x_t) = (0, 1, 0)$ makes $C = 0$. Compute the smallest unique natural number l such that

$$(l \bmod p_r, l \bmod p_s, l \bmod p_t) = (0, 1, 0),$$

and the regular expression

$$a^l (a^{p_r p_s p_t})^*. \tag{4.5}$$

Let E be the union of all such regular expressions (at most 8 for each C_i) and of E_0. Denote by L the language of E. Now by construction we have that the two following statements are equivalent for all $k \in \mathbb{N}$:

- the word $a^k \notin L$,
- k is a code and the valuation $h(k)$ makes the formula (4.4) true.

Indeed, for the forward direction if a^k does not belong to the language then it does not belong to $L(E_0)$ either, which means that it is a code, and it does not belong to the languages of expressions (4.5) that encode valuations that falsify the conjuncts. The same argument can be used for the converse as well.

By the Chinese reminder theorem we see that for any $v \in \{0,1\}^n$ there exists some k such that $h(k) = v$. This then implies that the 3-SAT formula (4.4) is satisfiable if and only if there is some $k \in \mathbb{N}$ such that $a^k \notin L$. Thus we have made the first reduction from 3-SAT.

We observe one property of the language L which we have just constructed. Define

$$p = \prod_{i=1}^{n} p_i.$$

By construction of L we have that for all $k \in \mathbb{N}$

$$a^k \notin L \qquad \Leftrightarrow \qquad a^{k+lp} \notin L, \text{ for all } l \in \mathbb{N}. \tag{4.6}$$

Indeed, if k is not a code then trivially $k + lp$ is not a code either, and if k falsifies one of the conjuncts, then so does $k + lp$, by definition (4.5).

Now we continue with the final reduction, from the problem about languages to the density 1 problem.

From the regular expression E, construct in polynomial time a non-deterministic finite automaton $\mathcal{A}$ that recognises the language $L \setminus \{\epsilon\}$, and such that it has a unique initial and a unique final state. Suppose that its states are $\{1, 2, \ldots, t\}$, where 1 is the initial state and t the final one. Let M be the adjacency matrix of $\mathcal{A}$. Observe that the number $M^k_{i,j}$ is exactly the number of runs of length k from state i to state j. Then by construction, for all $k \in \mathbb{N}$,

$$M^k_{1,t} \neq 0 \qquad \Leftrightarrow \qquad a^k \in L.$$

The sequence $\langle M^n_{1,t} \rangle_{n \in \mathbb{N}}$ is in fact a LRS whose every entry is non-negative. This LRS has a zero if and only if the 3-SAT instance is satisfiable. From (4.6), if this LRS has a zero then it has infinitely many of them, which fall on an infinite arithmetic progression with common differences at most p. In this case the density of the positivity set is < 1, otherwise, if the sequence has no zeros, the density is equal to 1. It follows that the density is not equal to one if and only if the 3-SAT instance (4.4) is satisfiable. □

Theorem 4.5. *The density 1 problem is co-NP-hard.*

Proof Sketch. This lower bound follows immediately from [OW14, Section 5], so we give only a sketch.

Consider the following problem. Given a polynomial $f \in \mathbb{Q}[x_1, \ldots, x_n]$ of degree at most 4, decide whether there are real numbers $x_1, \ldots, x_n$ such that

$$f(x_1, \ldots, x_n) = 0.$$

This problem, known as 4-FEAS (for feasibility) is NP-hard, see for example [BCSS98, Page 104, Theorem 1]. The complement decision problem, *i.e.* where one inputs a polynomial f as above and one has to decide whether *for all* real numbers $x_1, \ldots, x_n$ we have

$$f(x_1, \ldots, x_n) \geq 0, \tag{4.7}$$

is then co-NP-hard. By dividing the non-constant terms of f with a certain integer that can be computed in polynomial time from f, we construct a different polynomial f' such that (4.7) holds if and only if for all real $x_1, \ldots, x_n$ in the closed unit interval $[0, 1]$, we have

$$f'(x_1, \ldots, x_n) \geq 0. \tag{4.8}$$

This problem is reduced in polynomial time to the ultimate positivity for LRS in [OW14]. The idea is to construct algebraic numbers $\lambda_1, \ldots, \lambda_n$ that lie on the unit circle, such that for all $i \in \{1, \ldots, n\}$ we have

$$\left\{ \left(\lambda_i^k + \overline{\lambda}_i^k \right)^2 \ : \ k \in \mathbb{N} \right\} \text{ is dense in } [0, 1],$$

and furthermore there are no multiplicative relations among the λ_i, and the expression in parenthesis is a linear recurrence sequence, with rational entries. Then we consider the

sequence

$$f\left((\lambda_1^k+\overline{\lambda}_1^k)^2,\ldots,(\lambda_n^k+\overline{\lambda}_n^k)^2\right), k\in\mathbb{N},$$

which is a LRS; denote it by $\langle u_n\rangle_{n\in\mathbb{N}}$. Since the set of $x_1,\ldots,x_n$ in the n-cube $[0,1]^n$ for which $f(x_1,\ldots,x_n)<0$ is open (denote it by X), it follows that (4.8) does not hold if and only if u_n has infinitely many negative entries.

It is possible, via the methods described in the beginning of the section, to conclude that by construction of $\langle u_n\rangle_{n\in\mathbb{N}}$, and Theorem 4.1, the density of negative entries is equal to $\mu(X)$, the Lebesgue measure of X. Hence the density of the positive entries is equal to 1 if and only if (4.8) holds. The theorem follows. □

Since in the beginning of this section we saw that an upper bound for the density 1 problem is PSPACE, and the indications from the two theorems above are that a matching lower bound might exist, a search in this direction is interesting for the future.

4.2. **The Case of Diagonalisable Sequences.** If the given LRS has only finitely many positive entries then the density of the sequence is 0. The converse, however, does not always hold, as it can be seen from the following example:

Example 4.6. One can construct an LRS $\langle w_n\rangle_{n\in\mathbb{N}}$ that is equal to

$$w_n \stackrel{\mathsf{def}}{=} \frac{n}{2}\lambda^n+\frac{n}{2}\overline{\lambda}^n+(1-n),$$

where $\lambda\in\mathbb{T}$ is some algebraic number in the unit circle, that is not a root of unity. Let $\theta=\log\lambda/2\pi\mathbf{i}$. Then, by writing $\lambda^n=\cos(2\pi n\theta)+\mathbf{i}\sin(2\pi n\theta)$, we see that

$$w_n>0 \qquad \Leftrightarrow \qquad \cos(2\pi n\theta)>1-\frac{1}{n}.$$

The sequence $\langle w_n\rangle_{n\in\mathbb{N}}$ has infinitely many positive entries [AKK+21, Proposition 4.1]. (This can be shown by appealing to Dirichlet's theorem [Lan95, Chapter 2, Theorem 1], and considering the Taylor's expansion of cosine.)

However the density of the positive entries is 0. Indeed if it had density $\delta>0$, then we could have chosen some n large enough such that the interval of φ for which $\cos(2\pi\mathbf{i}\,\varphi)>1-1/n$, is smaller than δ, at which point, by applying Theorem 4.1 one can derive a contradiction. The latter theorem is applicable because λ is not a root of unity, which means that θ is irrational, by definition.

The direction "density 0" implies "positivity set is finite", does however hold for an important class of LRS, namely the diagonalisable sequences. These are sequences $\langle t_n\rangle_{n\in\mathbb{N}}$ whose characteristic polynomial has no repeated roots, as a consequence of which, its generalised power sum is of the following form:

$$t_n \stackrel{\mathsf{def}}{=} \sum_{i=1}^{k} a_i\Lambda_i^n,$$

where a_i are some algebraic constants and Λ_i are the roots.

Theorem 4.7. *In a diagonalisable sequence the positivity set is finite if and only if its density is zero.*

Proof. We prove the contrapositive, *i.e.* we show that if $\langle t_n \rangle_{n\in\mathbb{N}}$ has infinitely many positive entries then it also has positive density. Assume that the roots are ordered by modulus, *i.e.* $|\Lambda_i| \geq |\Lambda_{i+1}|$, and assume that the first j roots have maximal modulus. Write

$$t_n = \underbrace{\sum_{i=1}^{j} a_i \Lambda_i^n}_{D(n)} + \underbrace{\sum_{i=j+1}^{k} a_i \Lambda_i^n}_{r(n)}.$$

Suppose that $|\Lambda_1| > 1$, indeed if it is not, we can always multiply the sequence with $\langle K^n \rangle_{n\in\mathbb{N}}$ for $K \in \mathbb{N}$ large enough, without changing the sign. Without loss of generality, we can also assume that the sequence is non-degenerate.

The proof hinges on a lower bound on the growth of LRS that was proved Evertse, and in parallel by van der Poorten and Schlickewei, using the subspace theorem. See the discussion in [EVDPS+03, Section 2.4] as well as the appendix of [FH20]. Applying this theorem to our case, we have that for all $\epsilon > 0$ there exists some threshold $n_0 \in \mathbb{N}$ such that:

$$|D(n)| \geq |\Lambda_1|^{(1-\epsilon)n} \text{ for all } n \geq n_0.$$

Since $|r(n)|$ can be upper bounded by some $c|\Lambda|^n$, with $c \in \mathbb{R}$ a constant, and $|\Lambda| < |\Lambda_1|$, it follows that we can pick some $\epsilon > 0$ for which we know that there exists some $n_0 \in \mathbb{N}$ such that:

$$|D(n)| > |r(n)| \text{ for all } n \geq n_0.$$

This is a stronger version of Lemma 4.3, signifying that asymptotically the sign depends only on that of the dominant terms[8]. As a consequence of the inequality above, since the sequence $\langle t_n \rangle_{n\in\mathbb{N}}$ has infinitely many positive terms, so does the sequence $\langle D(n) \rangle_{n\in\mathbb{N}}$.

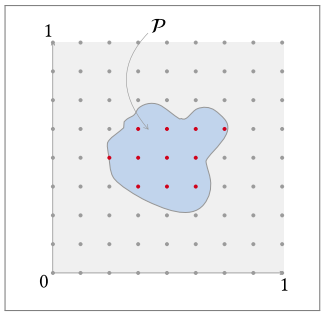
We sketch the rest of the proof. As in section 3 we can define the multiplicative relations among $\Lambda_1, \ldots, \Lambda_j$, and define a set $\mathcal{P}'$ analogous to the set $\mathcal{P}$, defined in the previous page. One can then prove that the set $\mathcal{P}'$ is open and furthermore it is non-empty as a consequence of the fact that $\langle D(n) \rangle_{n\in\mathbb{N}}$ has infinitely many positive entries. Non-emptiness implies that $\mathcal{P}'$ has non-zero measure, and finally, by applying Theorem 4.1, one concludes that the density of the sequence is positive. □

The algorithm that we have presented in this section is not the same, but it is quite similar to the algorithm of [OW14] for deciding ultimate positivity for diagonalisable sequences. We have shown that this algorithm can be used for deciding a different problem, namely whether the density of the sequence is zero, and that when the sequence is diagonalisable, the density 0 question is equivalent to the question of whether the sequence has only finitely many positive entries. The complexity lower bound of [OW14, Section 5] applies to our case as well.

[8]This inequality holds for general LRS. The difference is that for diagonalisable LRS, the dominant part $D(n)$ is easier to analyse.

5. Computing the Density

One method of approximating the density δ, which is the same as approximating the volume $\mu(\mathcal{P})$ of the set $\mathcal{P}$ is conceptually simple: draw a grid and count the points that belong to $\mathcal{P}$. We summarise this in the picture below.



From the grid of M^η points (in the example 9^2 points), we count how many are in $\mathcal{P}$, and denote this number by $C(M)$ (in the example this is equal to 11 red points). Since $\mathcal{P}$ is a measurable subset of the unit cube,

$$\frac{C(M)}{M^\eta} \to \mu(\mathcal{P}),$$

as M tends to infinity.

For this scheme to work, we need to be able to do two things. First, for any rational $\vec{q} \in [0,1]^\eta$, to be able to decide whether $\vec{q} \in \mathcal{P}$. And second, to be able to upper bound the quantity

$$\left| \frac{C(M)}{M^\eta} - \mu(\mathcal{P}) \right|, \tag{5.1}$$

by a function in M. We prove that both are feasible.

Lemma 5.1. *Given any rational $\vec{q} \in [0,1]^\eta$, it is decidable whether $\vec{q} \in \mathcal{P}$.*

Proof. Let $0 \leq k/n \leq 1$ be a rational number. The complex number $\exp(2\pi\mathbf{i}/n)$ is a primitive n-th root of unity, which we can isolate as a root of $x^n - 1$. It follows that $\exp(2\pi\mathbf{i}k/n) = \exp(2\pi\mathbf{i}/n)^k$ is an algebraic number that we can easily define. Consequently the assertion $\vec{q} \in \mathcal{P}$, which is equivalent to $F(\vec{q}) > 0$, is a first-order formula whose truth can be decided by Tarski's algorithm, Theorem 2.3. □

For an upper bound on the error (5.1), we use the work of Koiran [Koi95, Theorem 3]. To introduce his theorem we need to define the parameter $\kappa(\mathcal{P})$ first, and estimate it.

Let $S \subseteq [0,1]^\eta$ be a measurable set, define $\kappa(S)$ to be the maximal number of connected components of the intersection $L \cap S$ where L is an axis-parallel line. In other words, draw a line parallel to any one of the axes, and count how many times it goes in and out of the set. To estimate $\kappa(\mathcal{P})$, in our case, this translates to fixing all but one parameter of function F and counting how many times it will change its sign. More precisely, consider the function that we get by fixing all but one parameter of F, it will be of the form:

$$H(\varphi) \stackrel{\text{def}}{=} z_0 \exp(2\pi\mathbf{i}\ \varphi) + \sum_{i=1}^{\ell} z_i \exp(2\pi\mathbf{i}\ r_i\varphi) + c,$$

defined for $\varphi \in [0,1]$, where $\ell \leq m$, c,z_i are some algebraic numbers, and r_i are taken among the $q_{i,j}$, $\eta < i < m$, $1 \leq j \leq \eta$. The nature of the constants is such that H is a real-valued function. How many times does H change its sign in its domain $[0,1]$? By continuity, the answer is upper bounded by the number of zeros of H, which we will estimate. To this end, let $r_i = a_i/b_i$, for co-prime integers a_i, b_i, and define

$$b \stackrel{\text{def}}{=} \operatorname{lcm}\{b_1, \ldots, b_\ell\}.$$

Set $a'_i \in \mathbb{N}$ to be such that $r_i = a'_i/b$.

Lemma 5.2. *The function H has at most*

$$\hat{q} \stackrel{\text{def}}{=} \max\left\{b, a'_1, \ldots, a'_\ell\right\}$$

zeros in the unit interval $[0, 1]$.

Proof. We can write H as

$$z_0 \left(\exp(2\pi \mathbf{i}\ \varphi/b)\right)^b + \sum_{i=1}^{\ell} z_i \left(\exp(2\pi \mathbf{i}\ \varphi/b)\right)^{a'_i} + c,$$

which is a polynomial of degree at most $\hat{q}$, and hence can have at most that many zeros. □

Having estimated thus the parameter $\kappa(\mathcal{P})$, we have the following upper bound on the error:

Theorem 5.3 [Koi95, Theorem 3]**.** *For all $M \in \mathbb{N}$,*

$$\left|\frac{C(M)}{M^\eta} - \mu(\mathcal{P})\right| \leq \frac{\eta\kappa(\mathcal{P})}{M} \leq \frac{\eta\hat{q}}{M}.$$

Now Theorem 1.5 follows from Lemma 5.1 and Theorem 5.3. Indeed if we want to compute the density δ up to precision ϵ, it suffices to choose $M \geq \eta\hat{q}/\epsilon$, then for every member of

$$\left\{\left(\frac{k_1}{M}, \ldots, \frac{k_\eta}{M}\right)\ :\ 0 \leq k_i \leq M, 1 \leq i \leq \eta\right\}, \tag{5.2}$$

test whether it is in $\mathcal{P}$, and in this way compute the quantity $C(M)/M^\eta$ which by the proposition above is guaranteed to differ from the density by no more than ϵ.

Even though M is exponential in the input, by using the repeated squaring way of expressing the exponents in the formulas, as in section 3, it is possible to construct formulas of polynomial size for testing whether points of the grid (5.2) belong to $\mathcal{P}$. In particular to define $\exp(2\pi\mathbf{i}/M)$, the formula says that it is a root of $x^M - 1$ (which is of polynomial size), and that both the real and imaginary parts are positive and minimal. It follows that the algorithm for approximating the density is making exponentially many calls to a PSPACE algorithm (due to Theorem 2.3), each of which is used to decide whether to increment a counter that is upper bounded by M^η. Hence this algorithm is running in PSPACE on $\lceil\epsilon^{-1}\rceil$ and N, the bitlength of the description of the sequence. A similar analysis yields a PTIME upper bound in N and $\lceil\epsilon^{-1}\rceil$ when the order of the sequence is fixed.

Instead of testing whether *every* point in the grid belongs to $\mathcal{P}$, intuitively, we could test it for a smaller number $M' < M$, but choose the points uniformly at random. This is the Monte-Carlo integration method [Koi95]. It results in a number of points in the set $C'(M')$ for which it is known that for all $\epsilon > 0$,

$$\frac{1}{M^\eta}\left|C'(M') - C(M)\right| \leq \epsilon$$

holds with probability at least $1 - 2e^{-2M'\epsilon^2}$. This can be demonstrated using Hoeffding's inequality.

6. When is the Density a Rational Number?

We have proved that it is possible to decide whether the density of a given sequence is equal to 0, or to 1. Can we also decide whether the density is larger than some $q \in \mathbb{Q}$? The approximating scheme of the previous section is *a priori* of no help: since it might be the case that it outputs the estimates $\delta_1, \delta_2, \ldots$, for error bounds $\epsilon_1 > \epsilon_2 > \cdots$ such that q belongs to all intervals $(\delta_i - \epsilon_i, \delta_i + \epsilon_i)$. A natural approach to tackling this difficulty is to ask whether the density itself is an irrational number. If the density is irrational then it has some ϵ-neighbourhood which does not contain q, which means that for a sufficiently small ϵ_i, q does not belong in the interval $(\delta_i - \epsilon_i, \delta_i + \epsilon_i)$.

In this section we report some progress of this direction. We begin by showing that when there are no non-trivial multiplicative relations among the roots, density is a *period* as defined by Kontsevich and Zagier [KZ01], *i.e.* an integral of an algebraic function over a semialgebraic set. Afterwards, we prove that when there is at most one pair of dominant complex roots, it is decidable whether the density is rational, in which case we can compute it exactly.

6.1. Density as a Period.

The complex roots of a sequence $u_n = \sum f_i(n)\Lambda_i^n$ come in conjugate pairs. Furthermore if $\Lambda_j = \overline{\Lambda_i}$ then also $f_j(n) = \overline{f_i(n)}$. See [HHHK05, Proposition 2.13] for a proof. The multiplicative relations due to complex conjugacy, *i.e.* $\lambda_j\lambda_i = 1$, where $\lambda_i = \Lambda_i/|\Lambda_i|$ is the normalised root, we call **trivial relations**. Here we study sequences that do not have any non-trivial multiplicative relations among the roots. Under this restriction, the function F, defined in section 4, has the following form:

$$F(\vec{\varphi}) = \sum_{i=1}^{\eta} c_i \exp(2\pi\mathbf{i}\ \varphi_i) + \sum_{i=1}^{\eta} \overline{c_i} \exp(-2\pi\mathbf{i}\ \varphi_i) + c_m,$$

since the only dependent roots are the complex conjugates of the independent ones. Using Euler's formula, and a trigonometric identity, we see that this function can also be written as:

$$F(\vec{\varphi}) = c + \sum_{i=1}^{\eta} r_i \cos\left(2\pi(\varphi_i + \tau_i)\right),$$

where $c = c_m \in \mathbb{R}$, $r_i = |c_i|$, and τ_i is the argument of c_i.

We proceed by getting rid of the translation by τ_i. Define:

$$F'(\vec{\varphi}) \stackrel{\mathsf{def}}{=} F(\vec{\varphi} - \vec{\tau}).$$

Recall that $\mathcal{P}$ was defined as the set of $\vec{\varphi}$ for which $F(\vec{\varphi}) > 0$, and observe that

$$\mathcal{P}' \stackrel{\mathsf{def}}{=} \left\{\vec{\varphi}\ :\ F'(\vec{\varphi}) > 0\right\} = \mathcal{P} + \vec{\tau}.$$

Since $\mathcal{P}'$ is obtained from $\mathcal{P}$ by a translation, they have the same measure. Furthermore, as a consequence of symmetry of cosine we have:

$$\mu(\mathcal{P}') = 2^{\eta}\mu\big(\underbrace{\mathcal{P}' \cap [0, 1/2]^{\eta}}_{\hat{\mathcal{P}}}\big).$$

So the density of the sequence can be derived from the volume of $\hat{\mathcal{P}}$. We write the latter as a certain integral. To this end, define the set $\mathcal{L}$ as,

$$\mathcal{L} \stackrel{\text{def}}{=} \left\{ \vec{x} \in [-1,1]^\eta \ : \ c + \sum_{i=1}^{\eta} r_i x_i > 0 \right\}.$$

Observe that the function $\cos^{-1}(\vec{x})/2\pi$, denoted $g(\vec{x})$, is a continuously differentiable bijection from $[-1,1]^\eta$ to $[0,1/2]^\eta$, and that furthermore:

$$g(\mathcal{L}) = \hat{\mathcal{P}}.$$

Denote by g' the Jacobian of g, then a variable change (see [Spi18, Theorem 3-13]) leads to:

$$\mu(\hat{\mathcal{P}}) = \int_{g(\mathcal{L})} d\vec{\varphi} = \int_{\mathcal{L}} |\det g'| d\vec{x} = \frac{1}{(2\pi)^\eta} \int_{\mathcal{L}} \prod_{i=1}^{\eta} \frac{1}{\sqrt{1-x_i^2}} d\vec{x}.$$

From here it follows that $\mu(\mathcal{P})$ is rational if and only if

$$\int_{\mathcal{L}} \prod_{i=1}^{\eta} \frac{1}{\sqrt{1-x_i^2}} d\vec{x} \quad \in \quad \mathbb{Q}\ \pi^\eta. \tag{6.1}$$

The class of numbers that can be expressed as integrals of algebraic functions over semialgebraic sets are known as **periods** [KZ01]. They contain all algebraic numbers, as well as their logarithms, and some transcendental numbers like π; they are exceedingly commonplace however not well understood.

We do not know how to decide (6.1), but we point out to some work that might prove to be helpful. One is Conjecture 1 in [KZ01], that says that if one period has two different representations as integrals, one can obtain one from the other through three simple operations: additivity, change of variables and Stokes's formula. It is not clear however, even if the conjecture were to be true, how one can calculate a sequence of such operations. A more direct conjecture is one made by Grothendieck that predicts the transcendence degree of field extension of $\mathbb{Q}$ that are generated by a finite set of periods. See [Ayo14] for definitions and a discussion about these two conjectures. More seems to be known about the special case of curves [HW18], but in this case, for our purposes, we can give a more satisfactory answer by simpler means.

6.2. **One Pair of Dominant Complex Roots.** When there is at most one pair of dominant complex roots, we have $\eta = 1$ and the function F can be written as:

$$F(\varphi) = c + r\cos(2\pi(\varphi + \tau)).$$

Clearly when $|c| \geq |r|$ the density is either 1 or 0 depending on the sign of c, so assume that $|c| < |r|$. As we explained above, we can do away with the translation by τ when solely interested in density, and furthermore we can restrict φ to the interval $[0,1/2]$.

Since the sequence is non-degenerate, the ratio $\lambda/\overline{\lambda}$ is not a root of unity, which implies that φ is not a rational number. In this case, the equidistribution theorem, (Theorem 4.1), is applicable. As a consequence of that theorem, to calculate the density, it suffices to calculate the length of the interval in $[0,1/2]$ which includes all φ for which:

$$\cos(2\pi\varphi) > \frac{-c}{r}.$$

Depending on the sign of $-c/r$, the length of this interval is

$$\text{either} \quad \frac{\cos^{-1}(-c/r)}{2\pi} \quad \text{or} \quad 1 - \frac{\cos^{-1}(-c/r)}{2\pi},$$

in both cases it is rational if and only if $cos^{-1}(-c/r)$ is a rational multiple of π. In the remainder of this section we prove that we can decide whether the inverse cosine of a real algebraic number is a rational multiple of π.

Proposition 6.1. *Given a real algebraic number $\alpha \in [-1, 1]$ of degree d, it is decidable whether*

$$\cos^{-1}(\alpha) \in \mathbb{Q}\pi.$$

Proof. Clearly $\cos^{-1}(\alpha)$ is a rational multiple of π if and only if there is a rational $a/b \in \mathbb{Q}$, $b > 0$, such that $a \cos^{-1}(\alpha) = b\pi$. Which, in turn, holds if and only if there exists $a/b \in \mathbb{Q}$ (possibly different), $b > 0$, such that:

$$\cos\left(a \cos^{-1}(\alpha)\right) = (-1)^b.$$

To proceed we need the definition of the **Chebyshev polynomials of the first kind** of order n. These are univariate polynomials T_n, for $n \in \mathbb{N}$ that are characterised by the equation:

$$T_n(\cos\theta) \stackrel{\mathsf{def}}{=} \cos(n\theta).$$

One can also define them via a recurrence relation. We see that:

$$T_a(\alpha) = T_a(\cos\cos^{-1}(\alpha)) = \cos(a\cos^{-1}(\alpha)) = (-1)^b.$$

As a consequence $\cos^{-1}(\alpha)$ is a rational multiple of π if and only if there is some $n \in \mathbb{N}$, such that α is a root of

$$T_n(x) - 1 \qquad \text{or} \qquad T_n(x) + 1.$$

The roots of these polynomials are straightforward to describe:

Observation 6.2. Let $n \in \mathbb{N}$. All the roots of $T_n(x) + 1$ and of $T_n(x) - 1$ come from the set

$$\big\{ \pm\cos(k\pi/n) \ : \ 0 \le k \le n \big\}.$$

The proof of this observation follows plainly from the fact that for all $x \in \mathbb{R}$ such that $|x| \le 1$, we have

$$T_n(x) = T_n(\cos\cos^{-1} x) = \cos(n\cos^{-1} x)$$

and the fact that we can write $-\cos(k\pi/n)$ as $\cos(k\pi/n + \pi)$.

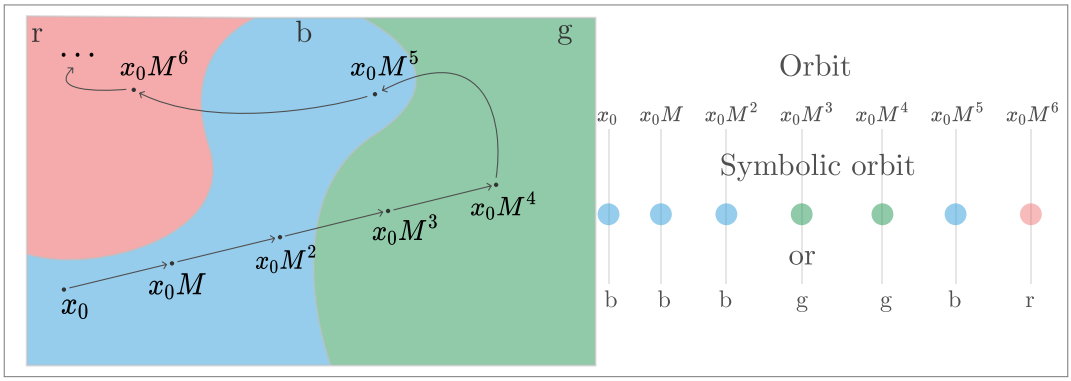
From Observation 6.2 and the discussion preceding it we conclude that $\cos^{-1}(\alpha)$ is a rational multiple of π if and only if it is equal to $\pm\cos(k\pi/n)$ for some $k, n \in \mathbb{N}$, $k \le n$. The numbers $\pm\cos(k\pi/n)$ are algebraic, indeed they satisfy the Chebyshev polynomial of order n, furthermore if $\gcd(k, n) = 1$ then $\cos(2k\pi/n)$ is an algebraic integer of degree $\Phi(n)/2$ [Leh33, Theorem 1], where Φ is the Euler's totient function.

Now, since α has degree d, we take some $N \in \mathbb{N}$ such that $\Phi(N) \ge 2d$. By testing (with the algorithms from Theorem 2.3 say) whether α is a root of any $T_n(x) \pm 1$, for $n \le N$ we can decide whether $\cos^{-1}(\alpha)$ is a rational multiple of π. □

7. Toric Words and Linear Dynamical Systems

As was described in the introduction, the decidability results of the preceding sections can be generalised to orbits of linear dynamical systems (LDS), where the positivity set is replaced by the set of indices corresponding to the members of the orbit that belong to a given semialgebraic set. In this section we explain how to achieve this generalisation.

Consider the orbit of a given LDS on the Euclidian plane.



The latter is partitioned into some semialgebraic sets; in the example above in the red, blue, and green set. The orbit of the system is the sequence x_0M^n, $n \in \mathbb{N}$, whereas the symbolic orbit (for the red, blue, and green subsets of $\mathbb{R}^2$) retains only the information to which set the element of the orbit belongs. Hence it is an ω-word over the alphabet $\Sigma := \{r, b, g\}$. We want to compute how frequently some pattern occurs in this word. More precisely, let $s = bbbggbr\cdots$ be the symbolic orbit, and $w \in \Sigma^*$ a finite word (or pattern) of length $|w|$, and say that w **occurs** in s in position n if and only if

$$s(n)s(n+1)\cdots s(n+|w|) = w,$$

where by $s(n)$ we write the nth letter of s. We will show how to compute the density of the set:

$$\{n \ : \ w \text{ occurs in } s \text{ in position } n\},$$

as well as decide whether it is equal to 0 or 1. We will call it the **density of the pattern** w in s, and denote it by

$$\mathcal{D}(w, s).$$

This number gives rather precise (albeit asymptotic) information about the dynamics of the given LDS, namely it tells you in which set of the partition the system spends most its time.

How does this generalise the density of the positivity set of an LRS? Let $k \in \mathbb{N}$ and suppose that the LRS $\langle u_n \rangle_{n\in\mathbb{N}}$ is given with the recurrence relation:

$$u_n = a_1u_{n-1} + \cdots + a_ku_{n-k},$$

where $a_k \neq 0$, and the first k entries: $u_1, \ldots, u_k$. Denote by M its companion matrix:

$$M \stackrel{\text{def}}{=} \begin{pmatrix} 0 & \cdots & 0 & a_k \\ & & & a_{k-1} \\ & I & & \vdots \\ & & & a_1 \end{pmatrix},$$

where the block marked by I is the $(k-1)\times(k-1)$ identity matrix. Denote by $\mathbf{v}$ the row vector $(u_1,\ldots,u_k)$. Then clearly we have for all $n\in\mathbb{N}$,

$$k\text{th coordinate of } \mathbf{v}M^n \text{ is equal to } u_{n+k}.$$

The partition of $\mathbb{R}^k$ that we take is then the semialgebraic set

$$S \stackrel{\text{def}}{=} \left\{(x_1,\ldots,x_k)\in\mathbb{R}^k \;:\; x_k>0\right\},$$

and its complement $\tilde{S}$. The symbolic orbit will be an ω-word over a binary alphabet, where one letter (call it p) would imply that the corresponding entry is positive, while the other letter would imply that it is ≤ 0. Then the density of the positivity set of the LRS $\langle u_n\rangle_{n\in\mathbb{N}}$ is just the density of the pattern p in the symbolic orbit of M for $(S,\tilde{S})$.

We give now the precise definitions. Let $d\in\mathbb{N}$ and $\boldsymbol{\lambda}\in\mathbb{T}^d$. The **orbit** of $\boldsymbol{\lambda}$ is the sequence:

$$\boldsymbol{\lambda}(n)\stackrel{\text{def}}{=}(\lambda_1^n,\ldots,\lambda_d^n),\quad n\in\mathbb{N}.$$

Let $k\in\mathbb{N}$, and $S_1,\ldots,S_k\subset\mathbb{T}^d$. The **symbolic orbit** of $\boldsymbol{\lambda}$ for $S_1,\ldots,S_k$ is an infinite word s over the alphabet

$$2^{\{1,2,\ldots k\}},$$

defined as follows. For all $i\in\{1,\ldots,k\}$ and $n\in\mathbb{N}$,

$$i\in s(n) \qquad \Leftrightarrow \qquad \boldsymbol{\lambda}(n)\in S_i.$$

Same as in the example above (except that we have not assumed that the semialgebraic sets form a partition), the symbolic orbit is an abstraction of the orbit in which the only information we want to retain for a point $\boldsymbol{\lambda}(n)$ is to which sets $S_1,\ldots,S_k$ it belongs.

Let $w_1\in\Sigma_1^\omega$ and $w_2\in\Sigma_2^\omega$. We say that w_2 is a **coarsening** of w_1 if there exists a map

$$r \;:\; \Sigma_1\to\Sigma_2,$$

such that $w_2=r(w_1)$, applied letter-wise. In this case we also say that w_1 **refines** w_2. If both w_1 refines w_2 and vice-versa, we say that w_1 and w_2 are **isomorphic**; equivalently there exists an injective such map r, such that $w_2=r(w_1)$.

A **toric word** is then any word isomorphic to the symbolic orbit of some $\boldsymbol{\lambda}\in\mathbb{T}^d$ for some $S_1,\ldots,S_k\subset\mathbb{T}^d$, where the components of $\boldsymbol{\lambda}$ are algebraic numbers, and the sets $S_1,\ldots,S_k$ are semialgebraic.

Subfamilies of toric words have been studied in other contexts: For example, the special case $d=1$, $\lambda\in\mathbb{T}$ is not a root of unity, and there is one set $S_1\subset\mathbb{T}$ that is an interval, has a symbolic orbit that is a Sturmian word. Such symbolic orbits have been studied going back to Johann Bernoulli III (1744-1807), see the notes on Chapter 9 of [AS03]. Muchnik et al. consider the case where $S_1,\ldots,S_k$ are open and disjoint and prove that in that case the symbolic orbits are almost periodic [MSU03, Section 4.3]. These words also have connections to extensions of MSO logic over $(\mathbb{N},<)$; however for the purposes of this paper we will be content with only showing a couple of closure properties of these words.

It is convenient to assume that the semialgebraic sets $S_1,\ldots,S_k$ partition $\mathbb{T}^d$, we can do this without loss of generality:

Lemma 7.1. *Let $\boldsymbol{\lambda}$ and $S_1,\ldots S_k$ be as above. There exists a partition of $\mathbb{T}^d$ into semialgebraic sets $R_1,\ldots,R_h$ such that the symbolic orbit of $\boldsymbol{\lambda}$ for $S_1,\ldots,S_k$ is isomorphic to that of $\boldsymbol{\lambda}$ for $R_1,\ldots,R_h$.*

Proof. Denote by $S_\emptyset$ the relative complement in $\mathbb{T}^d$, of the union of $S_1, \ldots, S_k$. For all non-empty $J \subset \{1, \ldots, k\}$ define:

$$S_J \stackrel{\text{def}}{=} \bigcap_{i \in J} S_i - \bigcup_{i \notin J} S_i.$$

Then the sets S_J for $J \subset \{1, \ldots, k\}$ partition $\mathbb{T}^d$, and they are furthermore semialgebraic. Enumerate these sets as $R_1, \ldots, R_h$. The symbolic orbit of $\boldsymbol{\lambda}$ for $R_1, \ldots, R_h$ is an infinite word over the alphabet $2^{\{1,\ldots,h\}}$. However, since the sets $R_1, \ldots, R_h$ partition $\mathbb{T}^d$, the only letters that will appear in the symbolic orbit are the singletons $\{i\}$, $i \in \{1, \ldots, h\}$.

It is not difficult to see now that the symbolic orbit of $\boldsymbol{\lambda}$ for $S_1, \ldots, S_k$ and that of $\boldsymbol{\lambda}$ for $R_1, \ldots, R_h$ are isomorphic, where the isomorphism depends on the particular enumeration that we have chosen. □

We continue with the closure properties. Let

$$w_1 \in \Sigma_1^\omega, \qquad w_2 \in \Sigma_2^\omega,$$

be two infinite words over the respective alphabets Σ_1, Σ_2. The **product** of w_1 and w_2 is the word $w_1 \times w_2$ over the alphabet $\Sigma_1 \times \Sigma_2$ defined as

$$(w_1 \times w_2)(n) \stackrel{\text{def}}{=} (w_1(n), w_2(n)).$$

Proposition 7.2. *Toric words are closed under taking products.*

Proof. Let w_1, w_2 be two toric words, where w_1 is obtained from the symbolic orbit of $\boldsymbol{\lambda} \in \mathbb{T}^d$ for semialgebraic sets $S_1, \ldots S_k$, and w_2 from that of $\boldsymbol{\gamma} \in \mathbb{T}^e$ for $T_1, \ldots, T_e$. From Lemma 7.1, we may assume that $S_1, \ldots, S_k$ and $T_1, \ldots, T_e$ partition $\mathbb{T}^d$, respectively $\mathbb{T}^e$. Then $\{S_1, \ldots, S_k\} \times \{T_1, \ldots, T_e\}$ partitions $\mathbb{T}^{d+e}$ and furthermore those sets are semialgebraic. Now it is plain that the symbolic orbit of

$$(\lambda_1, \ldots, \lambda_d, \gamma_1, \ldots, \gamma_e),$$

for

$$\{S_1, \ldots, S_k\} \times \{T_1, \ldots, T_e\}$$

is isomorphic to the product of w_1 and w_2. □

Proposition 7.3. *Toric words are closed under coarsenings.*

Proof. Let w be isomorphic to the symbolic orbit of $\boldsymbol{\lambda} \in \mathbb{T}^d$ for $S_1, \ldots, S_k$. We can assume that $S_1, \ldots, S_k$ partitions $\mathbb{T}^d$ due to Lemma 7.1. Then any coarsening of w is isomorphic to the symbolic orbit of $\boldsymbol{\lambda}$ for $T_1, \ldots, T_l$, for some $t \leq k$, where T_i are made of unions of sets S_i, and are therefore semialgebraic. □

A pleasant property of toric words, among others, is that we can decide whether the density of any given pattern that occurs in it is 0, or 1, as well as compute it to arbitrary additive precision.

Theorem 7.4. *There is a procedure for the following problem. Given as input:*

- *semialgebraic sets* $S_1, \ldots, S_k \subset \mathbb{T}^d$,
- *algebraic numbers* $\boldsymbol{\lambda} \in \mathbb{T}^d$,
- *a pattern* $w \in 2^{\{1,\ldots,k\}^*}$,

decide whether the density of w in the symbolic orbit of $\boldsymbol{\lambda}$ for $S_1, \ldots, S_k$ is zero.

Proof. Due to Lemma 7.1, we can assume that $S_1, \ldots, S_k$ partition $\mathbb{T}^d$, and therefore that the pattern w is a finite word over the alphabet $\Sigma := \{1, \ldots, k\}$.

Given $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{T}^d$, we write $\boldsymbol{\alpha}\boldsymbol{\beta}$ for the vector:

$$(\alpha_1\beta_1, \ldots, \alpha_d\beta_d).$$

Denote by m the length of the word w, and define the following semialgebraic set:

$$T_0 \stackrel{\mathsf{def}}{=} \left\{ \boldsymbol{\alpha} \in \mathbb{T}^d \ : \ \boldsymbol{\alpha} \in S_{w(1)},\ \boldsymbol{\alpha}\boldsymbol{\lambda}(1) \in S_{w(2)},\ \ldots,\ \boldsymbol{\alpha}\boldsymbol{\lambda}(m-1) \in S_{w(m)} \right\}.$$

The set T_0 characterises all points, starting from which, the orbit of $\boldsymbol{\lambda}$ moves among $S_1, \ldots, S_k$ in the next m steps in pattern w. To rephrase this more precisely, denote by $s \in \{1, \ldots, k\}^\omega$ the symbolic orbit of $\boldsymbol{\lambda}$ for $S_1, \ldots, S_k$. And by t the symbolic orbit of $\boldsymbol{\lambda}$ for $T_0, \tilde{T}_0$, where the latter is the relative complement of T_0 in $\mathbb{T}^d$; then the infinite word t is over the alphabet $\{0, \tilde{0}\}$. By construction, the following statements are equivalent for all $n \in \mathbb{N}$:

(1) w occurs in s in position n,
(2) $t(n) = 0$,
(3) $\boldsymbol{\lambda}(n) \in T_0$.

This equivalence implies that

$$\mathcal{D}(w, s) = \mathcal{D}(0, t), \tag{7.1}$$

where by $\mathcal{D}(w, s)$ we have denoted the density of the pattern w in s.

The procedure computes the period P as in section 3, for the algebraic numbers $\boldsymbol{\lambda}$. This is a product of P_1 (the least common multiple of orders of roots of unity that one can obtain by taking ratios λ_i/λ_j, $i \neq j$), and P_2 that only depends on the multiplicative relations among the coordinates of $\boldsymbol{\lambda}$. We split the orbit of $\boldsymbol{\lambda}$ and the symbolic orbit t into subsequences (subwords) by taking indices $nP + \ell$, $n \in \mathbb{N}$, $0 \leq \ell < P$. Denote by t_ℓ such a subword, *i.e.*

$$t_\ell(n) \stackrel{\mathsf{def}}{=} t(nP + \ell).$$

Clearly the density of 0 in t is positive (and hence also $\mathcal{D}(w, s) > 0$, due to (7.1)) if and only if there exists some ℓ, $0 \leq \ell < P$, such that the density of 0 in t_ℓ is positive. As the procedure tries to find such an ℓ, we only need to show how to decide whether

$$\mathcal{D}(0, t_\ell) > 0,$$

for some fixed ℓ, $0 \leq \ell < P$.

Define

$$T_\ell \stackrel{\mathsf{def}}{=} \left\{ \boldsymbol{\alpha} \in \mathbb{T}^d \ : \ \boldsymbol{\lambda}^\ell \boldsymbol{\alpha} \in T_0 \right\},$$

the translated T_0, so that $\boldsymbol{\lambda}(nP) \in T_\ell$ if and only if $\boldsymbol{\lambda}(nP+\ell) \in T_0$. To ease the notation write $\gamma_i = \lambda_i^P$. Taking subsequences, following section 3, we divide the γ_i into the independent ones and dependent ones. So there exists a partition of $\{1, \ldots, d\}$ into subsets I, D, which by rearranging assume that $I := \{1, \ldots, \nu\}$ and $D := \{\nu + 1, \ldots, d\}$, and rationals $q_{i,j} \in \mathbb{Q}$, $i \in D, j \in I$, such that we can write

$$(\gamma_1, \ldots, \gamma_d) = \left(\gamma_1, \ldots, \gamma_\nu, \prod_{j \in I} \gamma_j^{q_{\nu+1,j}}, \ldots, \prod_{j \in I} \gamma_j^{q_{d,j}} \right).$$

And furthermore there are no multiplicative relations among the $\gamma_1, \ldots, \gamma_\nu$. Imposing these dependencies for the coordinates $\nu+1, \ldots, d$ on the set T_ℓ, *i.e.* by requiring that

$$\gamma_i = \prod_{j \in I} \gamma_j^{q_{i,j}},$$

for all $i \in D$, we get a new semialgebraic set, denoted $\hat{T}_\ell$ which is a subset of $\mathbb{T}^\nu$. By definition we have that for all $n \in \mathbb{N}$, the following equivalences hold

$$(\gamma_1^n, \ldots, \gamma_\nu^n) \in \hat{T}_\ell \quad \Leftrightarrow \quad (\gamma_1^n, \ldots, \gamma_d^n) \in T_\ell \quad \Leftrightarrow \quad t_\ell(n) = 0.$$

Since $\gamma_1, \ldots, \gamma_\nu$ have no multiplicative relations, applying Theorem 4.1, and the equivalence just above, it follows that the density of 0 in the word t_ℓ is equal to the Lebesgue measure of $\hat{T}_\ell$. Since the latter is a semialgebraic set, its Lebesgue measure is nonzero if and only if $\hat{T}_\ell$ has nonempty interior. Indeed, a nonempty semialgebraic set that has empty interior must be a finite union of hyper-surfaces which have zero volume; for the converse it holds generally that any set with nonempty interior has positive volume.

Finally to decide whether $\hat{T}_\ell$ has nonempty interior we write a sentence in the first order logic of reals by saying that there exists some $r > 0$ and r-ball that is a subset of $\hat{T}_\ell$ and decide whether it is true by appealing to Theorem 2.3. □

Theorem 7.5. *There is a procedure for the following problem. Given as input:*

- *semialgebraic sets* $S_1, \ldots, S_k \subset \mathbb{T}^d$,
- *algebraic numbers* $\boldsymbol{\lambda} \in \mathbb{T}^d$,
- *a pattern* $w \in 2^{\{1,\ldots,k\}^*}$,
- *a rational constant* $\epsilon > 0$

compute the density of w *in the symbolic orbit of* $\boldsymbol{\lambda}$ *for* $S_1, \ldots, S_k$ *up to* ϵ *additive precision.*

Proof. From the proof of the preceding theorem, it suffices to only estimate the volume of the semialgebraic set $\hat{T}_\ell$. For this we proceed as in section 5. □

Having shown that we can compute the density of patterns in a toric word, it remains to show that the symbolic orbit of a LDS is similar to some toric word, which we can effectively construct. Indeed we will now prove that the symbolic orbit differs from a toric word in only a subset of indices that have zero density. Intuitively this is because the orbit of the LDS depends primarily on the dominant eigenvalues.

Theorem 7.6. *Let* $x_0 \in \mathbb{Q}^d$, $M \in \mathbb{Q}^{d \times d}$, *be a given* LDS *and* $S_1, \ldots, S_k$ *semialgebraic subsets of* $\mathbb{R}^d$. *Denote by* s *the symbolic orbit of* (x_0, M) *for* $S_1, \ldots, S_k$. *Then there exists a toric word* t *such that the set of* $n \in \mathbb{N}$ *for which*

$$s(n) \neq t(n),$$

has density zero.

Proof. Let $\langle u_n \rangle_{n \in \mathbb{N}}$ be a LRS and denote by s_1 the infinite word over the alphabet $\{p, \tilde{p}\}$, where we put the letter p in position n (*i.e.* $s_1(n) = p$) if and only if $u_n \geq 0$. We claim that:

Claim 7.7. There exists a toric word t_1 over the same alphabet that differs from s_1 only in a set of density zero.

Proof of Claim 7.7. Let $\alpha_1, \ldots, \alpha_r$ be the dominant characteristic roots of the sequence, *i.e.* those of maximal modulus, assumed distinct. Divide the sequence $\langle u_n \rangle_{n\in\mathbb{N}}$ by $|\alpha_1|^n n^{m-1}$ where m is the maximal multiplicity of the characteristic roots that have maximal modulus; same as in (3.8), to get a new sequence

$$v_n \stackrel{\mathsf{def}}{=} \underbrace{\sum_{i=1}^{r} c_i \alpha_i^n}_{D(n)} + R(n),$$

where $R(n)$ is some remainder that tends to zero as n grows larger. This new sequence has the exact same signs as the sequence $\langle u_n \rangle_{n\in\mathbb{N}}$ and therefore the same ω-word s_1. We let t_1 be the toric word that is the symbolic orbit of $\boldsymbol{\alpha}$ for $S, \tilde{S}$, where $S \subset \mathbb{T}^r$ is the semialgebraic set

$$\left\{ \boldsymbol{z} \in \mathbb{T}^r \ : \ \sum_{i=1}^{r} c_i z_i^n \geq 0 \right\},$$

and $\tilde{S}$ its relative complement in $\mathbb{T}^r$. So t_1 corresponds to the signs of $D(n)$, while s_1 corresponds to the signs of $D(n) + R(n)$. By splitting into sub-words $nP + \ell$, $0 \leq \ell < P$, for P defined in section 3 and applying Lemma 4.3, we get that sub-words differ only on a set of density zero (because the sign of the remainder matters only rarely). Since the union of a finite number of subsets of $\mathbb{N}$ that have density zero, also has density zero, the claim follows. □

Let f be a polynomial in $\mathbb{Z}[x_1, \ldots, x_d]$. The sequence

$$f(x_0 M^n), n \in \mathbb{N},$$

is an LRS. This is because every component of $x_0 M^n, n \in \mathbb{N}$ is itself an LRS, and these sequences are closed under point-wise addition and product. It follows that if we denote by S the semialgebraic set

$$\left\{ \boldsymbol{x} \in \mathbb{R}^d \ : \ f(\boldsymbol{x}) \geq 0 \right\},$$

and by $\tilde{S}$ its complement, the symbolic orbit of the LDS (x_0, M) for $S, \tilde{S}$ differs from a toric word only on a set of density zero; due to the claim above, Claim 7.7.

Suppose that we have two such symbolic orbits s_1, s_2, (ω-words over the alphabet $\{p, \tilde{p}\}$ with the semantics above), one for a polynomial f_1 and another for another polynomial f_2. Let t_1 respectively t_2 be the toric words to which they are similar. We can take the product $t_1 \times t_2$ which is also a toric word (thanks to Proposition 7.2) and then coarsen it by mapping (p, p), $(p, \tilde{p})$, $(\tilde{p}, p)$ to the same letter, say a, and $(\tilde{p}, \tilde{p})$ to the other letter, say b. The resulting word is toric (thanks to Proposition 7.3) and it differs in only a set of density zero from the symbolic orbit of (x_0, M) for the union of semialgebraic sets corresponding to $f_1 \geq 0$ and $f_2 \geq 0$. Similarly we proceed for intersection. Since semialgebraic sets are just unions and intersections of sets of $\boldsymbol{x}$ for which $f(\boldsymbol{x}) \geq 0$, the theorem follows. □

References

[AKK+21] Shaull Almagor, Toghrul Karimov, Edon Kelmendi, Joël Ouaknine, and James Worrell. Deciding omega-regular properties on linear recurrence sequences. *Proceedings of the ACM on Programming Languages*, 5(POPL):1–24, 2021.

[AS03] Jean-Paul Allouche and Jeffrey Shallit. *Automatic sequences: theory, applications, generalizations*. Cambridge university press, 2003.

[Ayo14] Joseph Ayoub. Periods and the conjectures of grothendieck and kontsevich-zagier. *European Mathematical Society. Newsletter*, (91):12–18, 2014.

[BCSS98] Lenore Blum, Felipe Cucker, Michael Shub, and Steve Smale. *Complexity and real computation*. Springer Science & Business Media, 1998.

[BG07] Jason P Bell and Stefan Gerhold. On the positivity set of a linear recurrence sequence. *Israel Journal of Mathematics*, 157(1):333–345, 2007.

[Bil08] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

[BM76] Jean Berstel and Maurice Mignotte. Deux propriétés décidables des suites récurrentes linéaires. *Bulletin de la Societe mathematique de France*, 79:175–184, 1976. doi:10.24033/bsmf.1823.

[BP02] Vincent D. Blondel and Natacha Portier. The presence of a zero in an integer linear recurrent sequence is np-hard to decide. *Linear Algebra and its Applications*, 351-352:91–98, 2002. doi:10.1016/s0024-3795(01)00466-9.

[BPR06] Saugata Basu, Richard Pollack, and Marie-Francois Roy. *Real Roots*, pages 351–401. Algorithms in Real Algebraic Geometry. Springer Berlin Heidelberg, 2006. doi:10.1007/3-540-33099-2_11.

[Can88] John Canny. Some algebraic and geometric computations in pspace. *Proceedings of the twentieth annual ACM symposium on Theory of computing - STOC '88*, 1988. doi:10.1145/62212.62257.

[Cas59] J. W. S. Cassels. *An Introduction To Diophantine Approximation*. Cambridge University Press, 1959.

[EVDPS+03] Graham Everest, Alfred Jacobus Van Der Poorten, Igor Shparlinski, Thomas Ward, et al. *Recurrence sequences*, volume 104. American Mathematical Society Providence, RI, 2003.

[FH20] Clemens Fuchs and Sebastian Heintze. On the growth of linear recurrences in function fields. *CoRR*, 2020. URL: http://arxiv.org/abs/2006.11074v1, arXiv:2006.11074.

[Han85] Georges Hansel. A simple proof of the skolem-mahler-lech theorem. In *International Colloquium on Automata, Languages, and Programming*, pages 244–249. Springer, 1985.

[HHHK05] Vesa Halava, Tero Harju, Mika Hirvensalo, and Juhani Karhumäki. Skolem's problem–on the border between decidability and undecidability. Technical report, Citeseer, 2005.

[HW18] Annette Huber and Gisbert Wüstholz. Transcendence and linear relations of 1-periods. *arXiv preprint arXiv:1805.10104*, 2018.

[Kel22] Edon Kelmendi. Computing the density of the positivity set for linear recurrence sequences. In *Proceedings of the 37th Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '22, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3531130.3532399.

[Koi95] Pascal Koiran. Approximating the volume of definable sets. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 134–141. IEEE, 1995.

[KZ01] Maxim Kontsevich and Don Zagier. *Periods*, pages 771–808. Mathematics Unlimited - 2001 and Beyond. Springer Berlin Heidelberg, 2001. doi:10.1007/978-3-642-56478-9_39.

[Lan95] Serge Lang. *Introduction to Diophantine Approximations*. Springer New York, 1995. doi:10.1007/978-1-4612-4220-8.

[Lan02] Serge Lang. *Algebra*. Graduate Texts in Mathematics. Springer New York, 2002. doi:10.1007/978-1-4613-0041-0.

[Lec53] Christer Lech. A note on recurring series. *Arkiv för Matematik*, 2(5):417–421, 1953.

[Leh33] Derrick H Lehmer. A note on trigonometric algebraic numbers. *Amer. Math. Monthly*, 40(3):165–166, 1933.

[Mah35] Kurt Mahler. *Eine arithmetische Eigenschaft der Taylor-koeffizienten rationaler Funktionen*. Noord-Hollandsche Uitgevers Mij, 1935.

[MSU03] An. Muchnik, A. Semenov, and M. Ushakov. Almost periodic sequences. *Theoretical Computer Science*, 304(1-3):1–33, 2003. doi:10.1016/s0304-3975(02)00847-2.

[OW13] Joël Ouaknine and James Worrell. Positivity problems for low-order linear recurrence sequences. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, 12 2013. doi:10.1137/1.9781611973402.27.

[OW14] Joël Ouaknine and James Worrell. *Ultimate Positivity is Decidable for Simple Linear Recurrence Sequences*, pages 330–341. Automata, Languages, and Programming. Springer Science + Business Media, 2014. doi:10.1007/978-3-662-43951-7_28.

[Ren92] James Renegar. On the computational complexity and geometry of the first-order theory of the reals. part i: Introduction. preliminaries. the geometry of semi-algebraic sets. the decision problem for the existential theory of the reals. *Journal of symbolic computation*, 13(3):255–299, 1992.

[Sko34] Thoralf Skolem. Ein verfahren zur behandlung gewisser exponentialer gleichungen und diophantischer gleichungen. *C. r*, 8:163–188, 1934.

[Spi18] Michael Spivak. *Calculus On Manifolds*. CRC Press, 2018. doi:10.1201/9780429501906.

[Tar51] Alfred Tarski. *A decision method for elementary algebra and geometry*. University of California Press, 1951.

[TMS84] R. Tijdeman, M. Mignotte, and T.N. Shorey. The distance between terms of an algebraic recurrence sequence. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1984(349):63–76, 1984. doi:10.1515/crll.1984.349.63.

[vdPL77] A.J. van der Poorten and J.H. Loxton. Multiplicative relations in number fields. *Bulletin of the Australian Mathematical Society*, 16(1):83–98, 1977. doi:10.1017/s0004972700023042.

[Ver85] N. K. Vereshchagin. Occurrence of zero in a linear recursive sequence. *Mathematical notes of the Academy of Sciences of the USSR*, 38(2):609–615, 1985. doi:10.1007/BF01156238.

[Wal00] Michel Waldschmidt. *Diophantine Approximation on Linear Algebraic Groups*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2000. doi:10.1007/978-3-662-11569-5.

[Wey16] Hermann Weyl. Uber die gleichverteilung von zahlen mod. eins. *Mathematische Annalen*, 77(3):313–352, 1916. doi:10.1007/bf01475864.

[YLN95] Kazuhiro Yokoyama, Ziming Li, and István Nemes. Finding roots of unity among quotients of the roots of an integral polynomial. In *Proceedings of the 1995 international symposium on Symbolic and algebraic computation - ISSAC '95*, - 1995. doi:10.1145/220346.220357.

[YS11] Chee K. Yap and Michael Sagraloff. A simple but exact and efficient algorithm for complex root isolation. In *Proceedings of the 36th international symposium on Symbolic and algebraic computation - ISSAC '11*, - 2011. doi:10.1145/1993886.1993938.