Deciding ω-Regular Properties on Linear Recurrence Sequences

ANONYMOUS AUTHOR(S)

We consider the problem of deciding ω-regular properties on infinite traces produced by linear loops. Here we think of a given loop as producing a single infinite trace that encodes information about the signs of program variables at each time step. Formally, our main result is a procedure that inputs a prefix-independent ω-regular property and a sequence of numbers satisfying a linear recurrence, and determines whether the sign description of the sequence (obtained by replacing each positive entry with "+", each negative entry with "−", and each zero entry with "0") satisfies the given property. Our procedure requires that the recurrence be simple, *i.e.*, that the update matrix of the underlying loop be diagonalisable. This assumption is instrumental in proving our key technical lemma: namely that the sign description of a simple linear recurrence sequence is almost periodic in the sense of Muchnik, Semënov, and Ushakov. To complement this lemma, we give an example of a linear recurrence sequence whose sign description fails to be almost periodic. Generalising from sign descriptions, we also consider the verification of properties involving semi-algebraic predicates on program variables.

Additional Key Words and Phrases: linear recurrence sequence, omega-regular properties, almost periodic words

1 INTRODUCTION

The decidability of monadic second-order logic (MSO) over the structure $\langle \mathbb{N}, < \rangle$ is a pillar of the theory of automated verification [Büchi 1962]. Shortly after Büchi established this result, Elgot and Rabin [Elgot and Rabin 1966] began to investigate unary predicates $\boldsymbol{P} \subseteq \mathbb{N}$ for which the MSO theory of the structure $\langle \mathbb{N}, <, \boldsymbol{P} \rangle$ remains decidable. For example, decidability is known in case $\boldsymbol{P}$ denotes, respectively, the set of factorial numbers $\{n! : n \in \mathbb{N}\}$, the set $\{p(n) : n \in \mathbb{N}\}$ for a fixed but arbitrary polynomial p with positive integer coefficients, and the set of k-powers $\{k^n : n \in \mathbb{N}\}$ for every fixed $k \in \mathbb{N}$. On the other hand, there are natural examples of predicates for which decidability is open and apparently difficult, e.g., for $\boldsymbol{P}$ the set of primes, see [Bateman et al. 1993].[1] In general, the decision problem for MSO over $\langle \mathbb{N}, <, \boldsymbol{P} \rangle$ reduces to the problem of checking membership of a fixed ω-word (namely the characteristic word of the predicate $\boldsymbol{P}$) in an ω-regular language $\mathcal{L}$ that represents the formula whose truth is to be determined.

Semënov [Semënov 1984] gave a characterisation of those predicates $\boldsymbol{P}$ for which the MSO theory of $\langle \mathbb{N}, <, \boldsymbol{P} \rangle$ is decidable. This line of work was continued in [Carton and Thomas 2002; Rabinovich 2007]. An important sufficient condition for decidability is that the characteristic sequence of $\boldsymbol{P}$ be *almost periodic*: a notion originating in symbolic dynamics [Morse and Hedlund 1938; Muchnik et al. 2003]. Roughly speaking, a sequence is almost periodic if for any pattern that occurs infinitely often, the gaps between successive occurrences are bounded. A classical example of an almost periodic sequence is the Thue-Morse sequence, i.e., the sequence whose n-th entry is the parity of the number 1's in the binary expansion of n. Another example is the characteristic sequence of the predicate $\boldsymbol{P} = \{n \in \mathbb{N} : \sin(n\theta) > 0\}$ for a fixed real number θ. The notion of almost periodicity will be instrumental for the results of this paper.

[1] Note that the twin primes conjecture from number theory (that there are infinitely many pairs of primes that differ by two) can be formulated in MSO with the primality predicate.

2018. 2475-1421/2018/1-ART1 $15.00
https://doi.org/

From the point of view of program analysis, it is natural to consider extensions of MSO with unary predicates that encode properties of program variables at each time step. For example, consider a linear loop

$$\textbf{while true do} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \leftarrow \begin{pmatrix} -2 & 4 & 0 \\ 4 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Given initial values of the program variables, suppose we want to determine whether the variable x is ultimately increasing. Noting that variable z stores the previous value of x, we equivalently want to determine ultimate positivity of the sequence $\langle u_n \rangle_{n \in \mathbb{N}}$ defined by $u_n := x_n - z_n$, where x_n and z_n are the respective values of variables x and z after n executions of the loop body. This property can be written in MSO as $\exists m \, \forall n \cdot (n \geq m \Rightarrow \boldsymbol{P}(n))$ where $\boldsymbol{P} = \{n \in \mathbb{N} : u_n > 0\}$. Clearly we can use second-order quantification in MSO to express more complex properties, e.g., that variable x only increases on even steps of the execution.

In the above example, the sequence $\mathbf{u} = \langle u_n \rangle_{n \in \mathbb{N}}$ satisfies the recurrence $u_{n+2} = -2u_{n+1} + 16u_n$. In general, for any linear loop and any polynomial function on the program variables, the sequence of values assumed by the function along an infinite execution of the loop is a linear recurrence sequence (LRS). This observation motivates the central object of study in this paper: the decidability of the MSO theory of the structure $\mathcal{S}_{\mathbf{u}} := \langle \mathbb{N}, \leq, \boldsymbol{P}, \boldsymbol{Z}, \boldsymbol{N} \rangle$ associated with an LRS $\mathbf{u}$, where $\boldsymbol{P} := \{n \in \mathbb{N} : u_n > 0\}$, $\boldsymbol{Z} := \{n \in \mathbb{N} : u_n = 0\}$, and $\boldsymbol{N} := \{n \in \mathbb{N} : u_n < 0\}$. This structure can be represented by the *sign description* $\langle \mathrm{sgn}(u_n) \rangle_{n \in \mathbb{N}} \in \{+, 0, -\}^\omega$ of $\mathbf{u}$, which is defined in the obvious way.

Computational problems concerning sign descriptions of LRS are notoriously difficult. For example, decidability of the Skolem Problem *"does a given LRS have a zero term?"* has been open for many decades. Decidability of the Positivity Problem: *"are all terms of a given LRS positive?"* is likewise a longstanding open problem [Ouaknine and Worrell 2013; Salomaa and Soittola 1978]. In view of these difficulties, we restrict attention to the class of *simple* LRS, i.e., those such that the characteristic polynomial of the defining recurrence has simple roots. In terms of our motivating example of linear loops, the associated LRS are simple whenever the update matrix of the loop is diagonalisable. Our main technical lemma shows that the sign description of every simple LRS is (effectively) almost periodic. We moreover give an example showing that almost periodicity fails without the assumption of simplicity.

Using the fact that simple LRS have effectively almost periodic sign descriptions, we establish our first main result:

Theorem 1.1. *For every fixed simple* LRS $\mathbf{u}$ *of rational numbers, it is decidable whether the sign description of* $\mathbf{u}$ *lies in a given* ω*-regular language* $\mathcal{L}$.

This result yields a large new class of structures with a decidable MSO theory, namely each structure $\mathcal{S}_{\mathbf{u}}$ for $\mathbf{u}$ a simple LRS.

We emphasize that Theorem 1.1 states the existence of a decision procedure for every fixed LRS $\mathbf{u}$. For our application of model checking linear loops, it is more natural to consider the LRS $\mathbf{u}$ as part of the input to the decision procedure, since the definition of $\mathbf{u}$ depends on the loop. This leads to our second main result:

Theorem 1.2. *Given a prefix-independent* ω*-regular language* $\mathcal{L}$ *and a simple* LRS $\mathbf{u}$*, it is decidable whether the sign description of* $\mathbf{u}$ *belongs to* $\mathcal{L}$.

This result allows us to model check prefix-independent MSO properties that refer to the signs of variables in linear loops. Recall here that a prefix-independent ω-regular language is one such that any two words with a common (infinite) suffix are either both in the language or both not in the language. Equivalently, such a language is a finite union of languages of the form $\Sigma^* \mathcal{L}^\omega$ for regular $\mathcal{L} \subseteq \Sigma^*$. Intuitively, prefix-independent languages specify asymptotic properties of ω-words, such as that a certain pattern occurs infinitely often or that some property is eventually true. The restriction to prefix-independent properties in Theorem 1.2 is connected to a non-uniformity in the proof that the sign description of a simple LRS is effectively almost periodic (cf. Theorem 3.1 and Remark 3.1), which in turn is due to our use of ineffecive number-theoretic bounds. Note that the ability to handle arbitrary ω-regular languages in Theorem 1.2 would immediately entail decidability of both Skolem's Problem and the Positivity Problem for simple LRS.

The sign description is a coarse abstraction of a given sequence. However the same techniques that provide us with Theorem 1.2 can be applied to much finer abstractions. Suppose that we are given m linear recurrence sequences and k semi-algebraic sets in $\mathbb{R}^m$. The m sequences can be seen as a dynamical system that evolves in $\mathbb{R}^m$. At each time step this system can be in any of the k semi-algebraic sets. This abstraction can be represented by an infinite word, whose alphabet is the power set of $\{1, \dots, k\}$. Almost periodicity of this richer symbolic semantics can be proved in a similar manner to Theorem 1.2.

Related Work. There have been a number of previous works that introduce symbolic semantics for linear systems, including linear loops and Markov chains, and give model checking procedures for this semantics. But the current paper is the first that establishes and benefits from almost periodicity of a symbolic semantics.

The paper [Karimov et al. 2020] examines a version of the decision problem considered in this paper, but with LTL formulas rather than MSO formulas, and restricting to recurrences of order at most 3 (corresponding to linear loops with at most 3 variables). In addition to the restriction on order, a major difference with the present paper is that [Karimov et al. 2020] does not use the notion of almost periodic sequences. Intuitively the model checking problem can be handled more directly there by exploiting the simplicty of LTL.

Paper [Beauquier et al. 2006] considers MSO over $\langle \mathbb{N}, < \rangle$ augmented with a probability quantifier. The semantics of the probability quantifier is defined relative to trajectories of a finite-state Markov chain. The setting is close to the present paper: Markov chains are a special case of linear loops, and the probability quantifier corresponds to having predicates that report the sign of an LRS at each index. However the results of [Beauquier et al. 2006] only apply in situations in which truth values of formulas are ultimately periodic. By working with the notion of almost periodicity we avoid the need for such semantic restrictions. Interestingly, [Beauquier et al. 2006, Section 8] notes the close relationship to the model checking problem for their logic and the Skolem Problem for linear recurrences.

Another similar work is [Agrawal et al. 2015], which considers the problem of model checking LTL formulas on a symbolic dynamics of a Markov chain that is induced by a finite polyhedral partition of the space of probability distributions on the states. Again, the key issue is ultimate periodicity: the authors of [Agrawal et al. 2015] note that their symbolic dynamics is not ultimately periodic in general, and therefore switch their attention of a notion of approximate model checking.

Decision problems on the positivity of LRS have been studied in [Ouaknine and Worrell 2013, 2014a,b]. Our second main result, Theorem 1.2, generalises the fact that it is decidable whether a simple LRS is ultimately positive [Ouaknine and Worrell 2014b]. In terms of the structure of the sign description of an LRS, [Bell and Gerhold 2007] show that the positivity set of an LRS (the set of indices where the LRS is positive) has a density and characterises the numbers that can appear as such a density. A classical result of Skolem, Mahler, Lech states that the set of zeros of an LRS over a field of characteristic zero is ultimately periodic.

Organisation. The rest of the paper is organised as follows. In Section 2, almost periodic sequences and sign descriptions of LRS are defined. There we also discuss two classical results: the Skolem-Mahler-Lech theorem and Semënov's theorem. The main technical lemma is proved in Section 3. In the last subsection, properties related to effectiveness of the objects defined in the proof are given. In Section 4 we show that the sign descriptions of general LRS need not be almost periodic. This section is independent and can be read out of order. In Section 5 we give the procedure and in Section 6 we show how the proof can be adapted to more complex predicates instead of sign descriptions. The pertinent notions of the first-order theory of real closed fields and related proofs are presented in Appendix A.

2 SIGN DESCRIPTIONS OF LINEAR RECURRENCE SEQUENCES

A *linear recurrence sequence (LRS)* is a sequence $\mathbf{u} = \langle u \rangle_{n \in \mathbb{N}}$ of rational numbers that satisfies a recurrence relation

$$u_n = a_1 u_{n-1} + a_2 u_{n-2} + \cdots + a_d u_{n-d}, \qquad n > d, \tag{1}$$

where $a_1, \ldots, a_d$ are rational constants and $d \in \mathbb{N}$ is the order of recurrence. Clearly such a sequence is determined by the recurrence and the initial values $u_1, \ldots, u_d$.

The *characteristic polynomial* of the recurrence (1) is

$$f(x) \overset{\text{def}}{=} x^d - a_1 x^{d-1} - \cdots - a_{d-1} x - a_d .$$

We refer to the roots of f as the *characteristic roots* of the recurrence. It is well known an LRS $\mathbf{u}$ admits a unique representation as an *exponential polynomial*

$$u_n = \sum_{i=1}^{m} C_i(n) \Lambda_i^n ,$$

where $\Lambda_1, \ldots, \Lambda_m$ are the distinct characteristic roots and the C_i are polynomials.

An LRS satisfies a unique recurrence of minimum order. We say that the recurrence is *simple* if the characteristic roots of this recurrence are simple. Equivalently $\mathbf{u}$ is simple if the coefficients C_i in its representation as an exponential polynomial are constant polynomials.

Let $\mathbf{u} = \langle u \rangle_{n \in \mathbb{N}}$ be a linear recurrence sequence. Define ζ, an infinite word over the alphabet $\{0, \pm\}$, as:

$$\zeta_n \overset{\text{def}}{=} 0 \qquad \Leftrightarrow \qquad u_n = 0.$$

In other words, we abstract away the terms of the sequence and only keep the information of whether or not they are equal to zero. The celebrated Skolem-Mahler-Lech theorem says that the word ζ is ultimately periodic.

THEOREM 2.1 (SKOLEM-MAHLER-LECH, [EVEREST ET AL. 2003, THEOREM 2.1]). *For any linear recurrence sequence* $\mathbf{u}$ *the word* ζ *is of the form*

$$\zeta = w_1 w_2^{\omega},$$

for $w_1, w_2 \in \{0, \pm\}^*$.

The word w_2 can be computed [Berstel and Mignotte 1976] from the description of $\mathbf{u}$; it is however a longstanding open problem whether the same is true for the prefix w_1.

In this paper we are interested in a slightly finer analysis:

Definition 2.2 (Sign description). The sign description of $\mathbf{u}$ is the infinite word σ over the alphabet $\{-, 0, +\}$ defined as:

$$\sigma_n \overset{\text{def}}{=} \operatorname{sgn}(u_n),$$

where for $x \in \mathbb{R}$,

$$\operatorname{sgn}(x) \overset{\text{def}}{=} \begin{cases} + & \text{if } x > 0, \\ - & \text{if } x < 0, \\ 0 & \text{otherwise}. \end{cases}$$

Unlike ζ, the word σ is not ultimately periodic in general, as the following example shows.

Example 2.3. Let $\mathbf{u}$ be an LRS given in the matrix form[2] as:

$$u_n = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ -b & a \end{pmatrix}^n \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad a, b \neq 0, \text{ and } a^2 + b^2 = 1.$$

Putting the square matrix above in Jordan normal form and using Euler's formula, we can deduce that $u_n = \cos(n\varphi)$, where $\varphi = \arg(a + b\mathbf{i})$ is not a rational multiple of π. If the sign description σ were to be ultimately periodic, then there would be some $p \in \mathbb{N}$ and $s \in \{-, 0, +\}$ such that $\sigma_{np} = s$ for all $n \in \mathbb{N}$. This is not the case, because $p\varphi$ is not a rational multiple of π, from which it easily follows (e.g., using Kronecker's theorem for inhomogeneous Diophantine approximation, Theorem 3.5) that $\{cos\,(n(p\varphi)) \;:\; n \in \mathbb{N}\}$ is dense in $[-1, 1]$.

However, in the case of simple LRS, the sign description is well-behaved. In the sequel we will prove that simple LRS have *almost periodic* sign descriptions.

2.1 Almost periodic words

We say that the pattern $w \in \Sigma^*$ occurs in a word $\alpha \in \Sigma^* \cup \Sigma^\omega$ if w occurs as an infix of α. More specifically, we say that w occurs in position p of α if

$$w = \alpha_p \alpha_{p+1} \cdots \alpha_{p+|w|-1}.$$

Definition 2.4 (Almost periodic). An infinite word $\alpha \in \Sigma^\omega$ is almost periodic if for every word $w \in \Sigma^*$, there exists $p \in \mathbb{N}$ such that either:

- w does not occur in α after the position p, or
- w occurs in every factor of α of length p, *i.e.* for every $n \in \mathbb{N}$, w occurs in

$$\alpha_n \alpha_{n+1} \cdots \alpha_{n+p}.$$

Intuitively, an almost periodic word is one with the property that any pattern that occurs infinitely often, does so in such a manner that the gaps between successive ocurrences of the pattern have bounded length. A typical *non-example* of almost periodic words is:

$$aba^2ba^3ba^4b \cdots .$$

Here the letter b occurs infinitely often, but the distances between consecutive occurrences are unbounded.

Almost periodic words are sometimes referred to in the literature as *uniformly recurrent sequences*, or *minimal sequences*. As examples of almost periodic words we have: ultimately periodic words, Sturmian words, and some morphic sequences such as the Thue-Morse sequence. Almost-periodic words enjoy good closure properties, low Kolmogorov complexity, *etc.*; [Muchnik et al. 2003] is an extensive study on the combinatorics of these words.

An almost periodic word $\alpha \in \Sigma^\omega$ is said to be *effectively almost periodic* if, given a pattern $w \in \Sigma^*$, we can decide whether or not w occurs infinitely often in α, and, if so, we can compute an upper bound p between successive occurrences of w in α. If the pattern does not occur infinitely often on

[2] This is an equivalent formulation for LRS, inter-reducible in polynomial time with the definition that we gave in the beginning of this section; see [Everest et al. 2003, Section 1.1.12].

the other hand, we can compute an upper bound on the threshold after which the pattern does not occur. Equivalently, α is effectively almost periodic if there is a procedure that inputs a pattern $w \in \Sigma^*$ and outputs an upper bound on the number p in Definition 2.4.

A key property of effectively almost periodic words is that they have a decidable monadic second-order theory. More specifically, a word $\alpha \in \Sigma^\omega$ determines a structure that expands $(\mathbb{N}, <)$ with a monadic predicate for every letter in Σ that denotes the positions in α where the letter occurs. Formulas of MSO over this structure are formulas of predicate logic with both first-order variables and monadic second-order variables. Then we have:

Theorem 2.5 ([Semënov 1984, Theorem 1]). *For any effectively almost periodic word α, the* MSO *theory of* $(\mathbb{N}, <)$ *expanded with unary predicates that define α, is decidable.*

One of the main results of this paper is that the sign description of a given simple LRS is an effectively almost periodic word. This effectiveness, however, is non-uniform, in the sense that we do not have a single algorithm that takes an LRS as input and witnesses the effectiveness of the corresponding sign description. Indeed such a uniform effectiveness result would allow to decide Skolem's Problem ("Does an LRS have a zero term?") and the Positivity Problem ("Are all terms of an LRS positive?"), both of which are open for simple LRS. This fact leads us to formulate and prove a variant of Theorem 2.5 that assumes a weaker notion of effectiveness that talks only about the asymptotic properties of the word. Specifcally this notion asks to compute an upper bound on the gap between all but finitely many succcessive occurrences of an infinitely recurring factor. Naturally, for such sequences we correspondingly weaken the conclusion of Theorem 2.5: we ask to decide any *prefix-independent* ω-regular property of the sign descriptions.

Issues of effectiveness will be discussed in Section 5. First we prove that sign descriptions of simple LRS are almost periodic.

3 SIMPLE LRS HAVE ALMOST PERIODIC SIGN DESCRIPTIONS

In this section we prove our first main result:

Theorem 3.1. *The sign description of a simple linear recurrence sequence is almost periodic.*

Fix a simple LRS $\mathbf{u}$. We first give a brief informal overview of the proof.[3] To set up the idea of the proof, recall that $\mathbf{u}$ admits a representation as an exponential polynomial

$$u_n = \sum_{i=1}^{d} c_i \, \Lambda_i^n, \tag{2}$$

where c_i, Λ_i are non-zero algebraic numbers, with Λ_i being characteristic roots of the recurrence defining $\mathbf{u}$. Now for each $i \in \{1, \ldots, d\}$, we factor each Λ_i as the product $\Lambda_i = \rho_i \lambda_i$ of a positive real number $\rho_i > 0$ and a complex number λ_i of absolute value 1. The first key idea is that for n sufficiently large, the sign of u_n is determined by $(\lambda_1^n, \ldots, \lambda_d^n)$, i.e., the absolute values of the characteristic roots can be ignored for large n. The second key idea is that the set $\{(\lambda_1^n, \ldots, \lambda_d^n) : n \in \mathbb{N}\}$ is the orbit of a point under a homeomorphism of a compact topological space, namely the d-fold product of the unit circle $\mathbb{T}$ in the complex plane. This transports us to a classical situation in symbolic dynamics.

As a preliminary step, we first decompose $\mathbf{u}$ as the interleaving of several so-called *non-degenerate* subsequences. Recall here that an LRS is said to be non-degenerate if no quotient of two distinct characteristic roots is a root of unity. To decompose $\mathbf{u}$, as given in (2), we take $P \in \mathbb{N}$ to be the least

[3] In fact the technical details, below, will depart slightly from this overview due to the need to handle the issue of degeneracy of LRS.

common multiple of the orders of all roots of unity among the quotients Λ_i/Λ_j for $1 \le i < j \le d$; then for all $\ell \in \mathbb{N}$, $0 \le \ell < P$, the sequence

$$\mathbf{u}^{(\ell)} \overset{\text{def}}{=} \langle u_{\ell+nP} \rangle_{n\in\mathbb{N}},$$

is a non-degenerate LRS with characteristic roots among $\{\Lambda_1^P, \dots, \Lambda_d^P\}$. We factor the characteristic roots as

$$\rho_i\lambda_i \overset{\text{def}}{=} \Lambda_i^P \qquad \rho_i \in \mathbb{R}_+,\ |\lambda_i| = 1,\ 1 \le i \le d. \tag{3}$$

The rationale behind this decomposition is that non-degenerate sequences have the following property:

Proposition 3.2 ([Shapiro 1959, Corollary 2.1]). *A non-degenerate* LRS *either has finitely many zeros, or it is identically zero.*

Next we will demonstrate that the sign description of $\mathbf{u}^{(\ell)}$ is asymptotically the same as that of a certain linear function on $(\lambda_1^n, \dots, \lambda_d^n)$, *i.e.*, it does not depend on the moduli ρ_i. We achieve this by applying the work of Evertse, van der Poorten, and Schlickewei on bounds of sums of S-units.

3.1 A lower bound on sums of S-units

We will prove the following lemma.

Lemma 3.3. *Let $\mathbf{v} \overset{\text{def}}{=} \mathbf{u}^{(\ell)}$, for some $0 \le \ell < P$. There exist $z_1, \dots, z_d \in \mathbb{C}$ such that $\sum_{i=1}^d z_i\lambda_i^n$ is real for all $n \in \mathbb{N}$, and furthermore exists $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$,*

$$\operatorname{sgn}(v_n) = \operatorname{sgn}\left(\sum_{i=1}^{d} z_i\lambda_i^n\right).$$

Remark 3.1. There is no known effective means of determining the constant n_0 above. For such a method we would need an effective version of Roth's theorem (consult Section 2.4 in [Everest et al. 2003]). It is as a consequence of the ineffectiveness of this constant that we are forced to restrict to *prefix-independent* ω-regular properties in the main theorem. It is worth noting, however, that in the presence of at most three dominant roots, this constant is effective [Tijdeman et al. 1984, Theorem 1].

The principal ingredient in the proof of Lemma 3.3 is the aforementioned lower bound on sums of S-units. We introduce this theorem first.

Let K be the field extension of $\mathbb{Q}$ generated by the characteristic roots $\Lambda_1, \dots, \Lambda_d$. The elements of K that are roots of monic polynomials in $\mathbb{Z}[x]$ (i.e., with leading coefficient one) form a subring, known as the *algebraic integers* of K, denoted O_K. Further, O_K is a Dedekind ring, so for every $x \in O_K$, the principal ideal generated by x can be written down as a product of a finite number of prime ideals. Let S be a finite set of prime ideals. An S-unit is any $x \in O_K$ such that the prime divisors of the principal ideal of x are in S.

If K has degree r over $\mathbb{Q}$ then there are r field embeddings from K to $\mathbb{C}$, denoted $h_1, \dots, h_r$.

Theorem 3.4 (Evertse, van der Poorten and Schlickewei, see *e.g.*[Everest et al. 2003, Section 1.5 and Section 2.4]). *Let S be a finite set of prime ideals in O_K, and $m \in \mathbb{N}$. Then for all $\epsilon > 0$ there exists $C > 0$, depending on ϵ and m, such that for any set of S-units $x_1, \dots, x_m \in O_K$, with the property that $\sum_{i\in I} x_i \ne 0$, $I \subseteq \{1, 2, \dots, m\}$, we have*

$$\left|\sum_{i=1}^{m} x_i\right| \ge CXY^{-\epsilon},$$

where $X \overset{\text{def}}{=} \max\{|x_i| \ : \ 1 \leq i \leq m\}$, *and* $Y \overset{\text{def}}{=} \max\{|h_j(x_i)| \ : \ 1 \leq i \leq m, 1 \leq j \leq r\}$.

We show how we can apply Theorem 3.4 to our setting.

Let $\ell \in \mathbb{N}$, $0 \leq \ell < P$, and $\mathbf{v} = \mathbf{u}^{(\ell)}$. Assume that $\mathbf{v}$ is not identically zero, then there exists $J \subseteq \{1, \ldots, d\}$, and $b_j \in \mathbb{C}$, $j \in J$ with $b_j \neq 0$ such that

$$v_n = \sum_{j \in J} b_j (\rho_j \lambda_j)^n.$$

Let $J' \subseteq J$ be the dominant roots (with modulus ρ) among the roots in J and write

$$v_n = \underbrace{\sum_{j \in J'} b_j (\rho \lambda_j)^n}_{D(n)} + \underbrace{\sum_{j \in J \setminus J'} b_j (\rho_j \lambda_j)^n}_{R(n)}.$$

We will use Theorem 3.4 to show that the sign of $\mathbf{v}$ asymptotically depends only on that of $D(n)$, by noticing that $D(n)$ is a sum of S-units.

By replacing the sequence $\mathbf{u}$ with $\langle ab^n u_n \rangle_{n \in \mathbb{N}}$ for certain positive integers a, b, which does not affect the sign, we can assume that numbers c_i and Λ_i in the exponential polynomial description (2) are algebraic integers.

Define S to be the set of prime divisors of $(\rho_j \lambda_j)$ and prime divisors of b_j. By definition of S all $(\rho_j \lambda_j), b_j$ are S-units and consequently $D(n)$ is a sum of S-units. To apply Theorem 3.4, we need now only show that any sub-sum of $D(n)$ vanishes for only finitely many n. To see this, observe that any sub-sum of $D(n)$ is itself a non-degenerate LRS, moreover we have assumed that it is not identically zero (because $b_j \neq 0$); as a consequence of Proposition 3.2, it cannot vanish for infinitely many n.

We now apply Theorem 3.4 to the sum of S-units $D(n)$. In this situation we clearly have $X = \rho^n$ and $Y \geq \rho^n$. It follows that for every $\epsilon > 0$ there exists $C > 0$ such that for all but finitely many n, we have

$$|D(n)| \geq C \rho^{n(1-\epsilon)}. \tag{4}$$

We are now ready to prove Lemma 3.3.

Proof of Lemma 3.3. If $\mathbf{v}$ is identically zero the lemma clearly holds. Assume that $\mathbf{v}$ is not identically zero. Since $\rho > \rho_j$ for $j \in J \setminus J'$, we have that there exists some $\epsilon_1 > 0$ such that for all but finitely many n,

$$|R(n)| < \rho^{n(1-\epsilon_1)}.$$

Since (4) holds for any $\epsilon > 0$ it follows now that for all but finitely many n,

$$|D(n)| > |R(n)|.$$

For all but fintely many n we thus have

$$\operatorname{sgn}(v_n) = \operatorname{sgn}(D(n)) = \operatorname{sgn}\left(\sum_{j \in J'} b_j \lambda_j^n \right).$$

This completes the proof the lemma. □

3.2 Orbits in $\mathbb{T}^d$

Lemma 3.3 tells us that the information about the sign description σ can be found in the set $\{(\lambda_1^n, \ldots, \lambda_d^n) \ : \ n \in \mathbb{N}\}$. We will recall a classical result that says that the set above is a dense subset of the set of points in the d-dimensional torus that have all the multiplicative relations as $\lambda_1, \ldots, \lambda_d$.

Consider the set of multiplicative relations of $\lambda = (\lambda_1, \ldots, \lambda_d)$:

$$\mathcal{M}_\lambda \overset{\text{def}}{=} \{\mathbf{v} \in \mathbb{Z}^d \ : \ \lambda_1^{v_1}\lambda_2^{v_2}\cdots\lambda_d^{v_d} = 1\}.$$

The one-dimensional torus is the unit circle $\mathbb{T} \overset{\text{def}}{=} \{z \in \mathbb{C} \ : \ |z| = 1\}$. Define the set of points in $\mathbb{T}^d$ having all the multiplicative relations of λ as follows:

$$\mathbb{T}_\lambda \overset{\text{def}}{=} \{\mathbf{z} \in \mathbb{T}^d \ : \ z_1^{v_1}z_2^{v_2}\cdots z_d^{v_d} = 1 \text{ for all } \mathbf{v} \in \mathcal{M}_\lambda\}.$$

Denote by $s \ : \ \mathbb{T}_\lambda \to \mathbb{T}_\lambda$ the map

$$(z_1, \ldots, z_d) \mapsto (z_1\lambda_1, \ldots, z_d\lambda_d).$$

With this new notation we are interested in the set $\{s^n(1, \ldots, 1) \ : \ n \in \mathbb{N}\}$. To prove that it is a dense subset of $\mathbb{T}_\lambda$ we will use Kronecker's theorem on simultaneous Diophantine approximation.

Theorem 3.5 (Kronecker, see e.g. [Cassels 1957, Page 53]). *Let $\theta_1, \ldots, \theta_k, \varphi_1, \ldots, \varphi_k \in \mathbb{R}$ such that for any integers $a_1, \ldots, a_k$,*

$$\sum_{i=1}^{k} a_i\theta_i \in \mathbb{Z} \qquad \Rightarrow \qquad \sum_{i=1}^{k} a_i\varphi_i \in \mathbb{Z}.$$

Then for every $\epsilon > 0$, there exists $n \in \mathbb{N}$ and integers $r_1, \ldots, r_k$ such that

$$|n\theta_i - r_i - \varphi_i| \leq \epsilon,$$

for all $i \in \{1, \ldots, k\}$.

Lemma 3.6. *For all $\mathbf{z} \in \mathbb{T}_\lambda$, the set $O(\mathbf{z}) \overset{\text{def}}{=} \{s^n(z_1, \ldots, z_d) \ : \ n \in \mathbb{N}\}$ is dense in $\mathbb{T}_\lambda$.*

Proof. Let $\mathbf{y} \in \mathbb{T}_\lambda$. We have to prove that $O(\mathbf{z})$ intersects every ϵ-ball around $\mathbf{y}$. Let us write

$$\lambda_i = e^{\theta_i \, 2\pi \mathrm{i}}, \qquad z_i = e^{\alpha_i \, 2\pi \mathrm{i}}, \qquad y_i = e^{\beta_i \, 2\pi \mathrm{i}},$$

and set $\varphi_i = \beta_i - \alpha_i$, for $i \in \{1, \ldots, d\}$. Because $\mathbf{y}$ and $\mathbf{z}$ belong to $\mathbb{T}_\lambda$, and the multiplicative relations of λ correspond to additive relations of θ, the hypothesis of Theorem 3.5 is fulfilled and the theorem can be applied. It tells us that there exists $n \in \mathbb{N}$ and integers $r_1, \ldots, r_d$ such that for every $i \in \{1, \ldots, d\}$,

$$\left|z_i\lambda_i^n - y_i\right| = \left|e^{(\alpha_i + n\theta_i - r_i)\, 2\pi \mathrm{i}} - e^{\beta_i \, 2\pi \mathrm{i}}\right| \leq 2\pi \ |\alpha_i + n\theta_i - r_i - \beta_i| \leq 2\pi\epsilon.$$

□

Compactness of $\mathbb{T}_\lambda$ together with Lemma 3.6 entail that any open set in $\mathbb{T}_\lambda$ can be reached in a bounded number of steps from any other point.

Lemma 3.7. *Let $U \subseteq \mathbb{T}_\lambda$ be an open set. There exists $B \in \mathbb{N}$ such that for every $\mathbf{x} \in \mathbb{T}_\lambda$, there exists $n \leq B$ such that $s^n(\mathbf{x}) \in U$.*

Proof. Lemma 3.6 implies that for any $\mathbf{z} \in \mathbb{T}_\lambda$, there exists some $n \in \mathbb{N}$ such that $s^n(\mathbf{z}) \in U$. Whence by continuity of the successor function s, we have that

$$\{s^{-n}(U) \ : \ n \in \mathbb{N}\} = \mathbb{T}_\lambda,$$

is an open cover of $\mathbb{T}_\lambda$. Since $\mathbb{T}_\lambda$ is bounded and closed as a subset of $\mathbb{T}^d$, it is compact. It follows that it admits a finite sub-cover, *i.e.* there exists $B \in \mathbb{N}$ such that

$$\{s^{-n}(U) \ : \ n \in \{1, 2, \ldots, B\}\} = \mathbb{T}_\lambda.$$

□

3.3 The proof of Theorem 3.1

It is tempting to try to prove Theorem 3.1 by showing that the sign descriptions of every subsequence $\mathbf{u}^{(\ell)}$, where $0 \leq \ell < P$, is almost periodic and combining the results. Unfortunately the proof cannot be modular in this respect, for the simple fact that the product of two almost periodic sequences need not be almost periodic itself; see [Muchnik et al. 2003, Theorem 22]. We must directly prove almost periodicity for the whole sequence, which is done as follows.

Let $\mathbf{u}$ be a simple LRS, $\sigma \in \{-, 0, +\}^\omega$ its sign description, and $w \in \{-, 0, +\}^*$, a pattern that occurs infinitely many times in σ. We have to prove that the distances between between consecutive occurrences are bounded.

Since w occurs infinitely many times in σ, there is some $m \in \mathbb{N}$ such that for infinitely many n,

$$w \text{ occurs in } \sigma_{nP}\sigma_{nP+1}\cdots\sigma_{(n+m)P-1}, \tag{5}$$

where we recall that P was defined as the least common multiple of orders of roots of unity among the ratios of roots of $\mathbf{u}$. Since the right-hand side of (5) is a word over a finite alphabet, there exists a word $w' \in \{-, 0, +\}^*$ that has w as an infix such that for infinitely many n,

$$w' = \sigma_{nP}\sigma_{nP+1}\cdots\sigma_{(n+m)P-1}.$$

We prove that there is an upper bound for the distances among successive such n, which clearly implies almost periodicity of σ.

Cut the word w' into m factors of length P such that

$$w' = w'(1)w'(2)\cdots w'(m).$$

Applying Lemma 3.3 to each subsequence $\mathbf{u}^{(\ell)}$, and combining the resulting linear functions together, we obtain a linear function $f \ : \ \mathbb{T}_\lambda \to \mathbb{R}^P$ such that for all but finitely many n, if $f\left(s^n(1, \ldots, 1)\right) = (a_1, \ldots, a_P)$, then

$$\operatorname{sgn}(a_1)\operatorname{sgn}(a_2)\cdots\operatorname{sgn}(a_P) = \sigma_{nP}\sigma_{nP+1}\cdots\sigma_{(n+1)P-1}.$$

Denote by $g \ : \ \mathbb{T}_\lambda \to \{-, 0, +\}^P$ the composition of f and sgn applied component-wise.

Since $\mathbf{u}^{(\ell)}$ are all non-degenerate, because of Proposition 3.2, some coordinates of $f(s^n(1, \ldots, 1))$ are identically zero, and some are zero only for finitely many n. Denote by $Z \subseteq \{1, 2, \ldots, P\}$ the former. On components in Z, f is a constant function mapping to zero.

It follows from the continuity of f that for elements of $v \in \{-, 0, +\}^P$ that have zeros exactly in coordinates Z, $f^{-1}(v)$ is an open subset of $\mathbb{T}_\lambda$. Since $w'(1), \ldots, w'(m)$ are words that occur infinitely often, they must have zeros exactly in positions in Z, hence the set

$$U(w') \overset{\text{def}}{=} \{\mathbf{x} \in \mathbb{T}_\lambda \ : \ g(\mathbf{x}) = w'(1), g\left(s(\mathbf{x})\right) = w'(2), \ldots, g\left(s^{m-1}(\mathbf{x})\right) = w'(m)\},$$

is open. By applying Lemma 3.7 we know that there exists $B \in \mathbb{N}$ such that for any $\mathbf{y} \in \{s^n(1, \ldots, 1) \ : \ n \in \mathbb{N}\}$, there exists $k \leq B$ with $s^k(\mathbf{y}) \in U(w')$. So from any point, in fewer than B steps, we enter the set $U(w')$ from where g outputs w' (in the next m steps). This proves that the distances between consecutive n for which (5) holds is at most $B \cdot P$.

3.4 Effectiveness

We make a closer inspection of the proof of almost periodicity above, in order to gather three lemmas which pull out what can be effectively computed about the sign description.

Lemma 3.8. *Let $w' = w'(1)w'(2)\cdots w'(m)$ be such that $w'(i)$ are factors of length P, then the following two statements are equivalent*

- *for infinitely many n,*

$$\sigma_{nP}\sigma_{nP+1}\cdots\sigma_{(n+m)P-1} = w'$$

- *$U(w')$ is non-empty.*

Proof. ($\Rightarrow$) Let $k \in \mathbb{N}$ be such that the equation in Lemma 3.3 holds for all $\ell \in \{0, 1, \ldots, P-1\}$. From the hypothesis there exists some $n > k$ such that $\sigma_{nP}\cdots\sigma_{(n+m)P-1} = w'$. Now the definition of the set $U(w')$ implies that it is not empty.
($\Leftarrow$) Since $U(w')$ is non-empty and open, we can apply Lemma 3.7, which gives us a bound B on how many steps we have to take in the walk in $\mathbb{T}^d$ before we enter again the set $U(w')$. Therefore we enter the set $U(w')$ infinitely many times, and hence the word w' occurs infinitely often in σ. □

The next lemma says that modulo a finite prefix, the word σ is strongly recurrent, which means that if some word occurs in it, it does so infinitely often. This stems from the fact that after some threshold, the sign description only depends on the walk in $\mathbb{T}^d$, which is repetitive.

Lemma 3.9. *There exists a threshold $c \in \mathbb{N}$ such that any word that occurs in the suffix $\sigma_c\sigma_{c+1}\cdots$, occurs infinitely often in σ.*

Proof. Let $n_1 \in \mathbb{N}$ be such that for all $n \geq n_1$ and $0 \leq \ell < P$, the equation in Lemma 3.3 holds. Let $n_2 \in \mathbb{N}$ be such that for all $0 \leq \ell < P$, $\mathbf{u}^{(\ell)}$ is either identically zero or has no zeros after n_2 (well defined thanks to Proposition 3.2). Set $c = \max\{n_1, n_2\}$. Let w be some word that occurs after the threshold c in σ. Then there is some n and m such that w occurs in

$$\sigma_{nP}\sigma_{nP+1}\cdots\sigma_{(n+m)P-1}.$$

Call this word w' and let $w'(1), \ldots, w'(m)$ be its decomposition into factors of length P. Since w' occurs after n_2 the factors $w'(1), \ldots, w'(m)$ have zeros exactly in the same positions $Z \subseteq \{1, \ldots, P\}$. This, together with the fact that w' occurs after n_1 implies that $U(w')$ is open and non-empty. Now Lemma 3.7 gives us a bound for the distances between consecutive occurrences of w'. □

The last crucial property is that for all w the set $U(w)$ is semi-algebraic and effective. We give here a sketch and relegate the full proof, as well as the relevant definitions, to Appendix A.

Lemma 3.10. *For all w, $U(w)$ is semi-algebraic, and we can compute the first-order formula that defines it.*

Proof sketch. The set $U(w)$ is a subset of $\mathbb{C}^d$ while semi-algebraic sets are subsets of $\mathbb{R}^n$. However there is a simple first-order interpretation of $\mathbb{C}$ in $\mathbb{R}$, we take for every complex number two real variables, one for the real part and one for the imaginary part.

The set of normalized roots $\lambda_1, \ldots, \lambda_d$ are algebraic numbers whose first-order formulas we can effectively construct given a lrs. Their multiplicative relations, *i.e.* the set $\mathcal{M}_\lambda$ has a finite basis, which can be computed using a result of Masser. Whence it follows that $\mathbb{T}_\lambda$ is semi-algebraic and that we can effectively construct the first-order formula that defines it.

The lemma now follows because the coefficients of the linear map $f : \mathbb{T}_\lambda \to \mathbb{R}^P$ (in the definition of $U(w)$ and g), are algebraic and we can compute their first-order formula. For a full proof see Appendix A. □

4 A COUNTER-EXAMPLE FOR GENERAL LRS

We have proved that simple LRS have almost periodic sign descriptions. In this section we show that the same does not hold for general LRS. The additional structure of simple LRS is consonant with what is known about decidability: e.g., it is decidable whether a simple LRS is ultimately positive, whereas the decidability of the same question for general LRS is open and a positive result would imply computability of Lagrange constants of certain transcendental numbers, see [Ouaknine and Worrell 2013, Theorem 5.1].

Let $\lambda \in \mathbb{T}$ be any algebraic number in the unit circle that is not a root of unity. Consider the generalized power sum

$$u_n \overset{\text{def}}{=} \frac{n}{2}\,\lambda^n + \frac{n}{2}\,\overline{\lambda}^n + (1-n)\,1^n,$$

where $\overline{\lambda}$ is the complex conjugate of λ. As discussed before, such sums are equivalent to linear recurrence sequences (see [Everest et al. 2003, Section 1.1.6]), so we can easily extract an LRS of order six from the sum above. This example has been designed in such a way as to have the following property. Set $\theta := \arg(\lambda)$, and using Euler's formula deduce that $u_n = 1 - n + n\cos(n\theta)$. Consequently for all n, we have:

$$u_n > 0 \qquad \Leftrightarrow \qquad \cos(n\theta) > 1 - \frac{1}{n}. \tag{6}$$

We will prove that this sequence has a sign description that is not almost periodic. More precisely we will prove that (a) the letter '+' occurs infinitely often in the sign description and (b) that the distances between consecutive occurrences can be arbitrarily large.

The intuition is as follows. Since λ is not a root of unity, θ is not a rational multiple of π and hence $\{\cos(n\theta) : n \in \mathbb{N}\}$ is a dense subset of $[0,1]$. Using basic properties of the cosine function we can prove that the right-hand inequality in (6), *i.e.* $\cos(n\theta) > 1 - 1/n$, is true for infinitely many n. However, since the interval $(1 - 1/n, 1]$ becomes arbitrarily tight as n increases, we have to wait longer and longer until $\cos(n\theta)$ enters it.

We give now the proofs of the two claims (a) and (b) above.

Proposition 4.1. *For infinitely many $n \in \mathbb{N}$, $u_n > 0$.*

Proof. For a real number $x \in \mathbb{R}$ denote by $[x]$ its distance to the closest integer, and by $[x]_{2\pi}$ its distance to the closest integer multiple of 2π, *i.e.*

$$[x] \overset{\text{def}}{=} \min_{k\in\mathbb{Z}} |x - k|, \qquad\qquad [x]_{2\pi} \overset{\text{def}}{=} \min_{k\in\mathbb{Z}} |x - 2k\pi|.$$

We will first prove that for infinitely many $n \in \mathbb{N}$,

$$[n\theta]_{2\pi} < \frac{2\pi}{n}.$$

This is a corollarly of Dirichlet's Theorem [Lang 1995, Chapter 2, Theorem 1], which states that for every $x \in \mathbb{R}$ there exist infinitely many $n \in \mathbb{N}$ such that $[nx] < 1/n$. Indeed since for any $x \in \mathbb{R}$, $[x]_{2\pi} = 2\pi[x/(2\pi)]$, Dirichlet's Theorem implies that for infinitely many $n \in \mathbb{N}$,

$$[n\theta]_{2\pi} = 2\pi \left[n\frac{\theta}{2\pi} \right] < \frac{2\pi}{n}. \tag{7}$$

By the monotonicity of the cosine function on $[0, \pi]$ we have that for all $n \geq 2$, $[n\theta]_{2\pi} < 2\pi/n$ if and only if $\cos(n\theta) > \cos(2\pi/n)$. As a consequence of (7), the inequality $\cos(n\theta) > \cos(2\pi/n)$ holds for infinitely many n.

Using a Taylor series expansion of cosine, we can prove that for x sufficiently close to 0,

$$\cos x \geq 1 - \frac{x^2}{2} \geq 1 - \frac{|x|}{2\pi}.$$

Applying this bound to $x = 2\pi/n$, we derive that for infinitely many $n \in \mathbb{N}$,

$$\cos(n\theta) > \cos\left(\frac{2\pi}{n}\right) \geq 1 - \frac{1}{n}.$$

Therefore, from (6) it follows that $\langle u \rangle_{n\in\mathbb{N}}$ is positive in infinitely many positions, or equivalently the letter '+' occurs infinitely often in its sign description. □

Proposition 4.2. *For every $p \in \mathbb{N}$ we can find p consecutive entries in $\langle u \rangle_{n\in\mathbb{N}}$ that are negative or zero.*

Proof. Fix some $p \in \mathbb{N}$ and let $N \in \mathbb{N}$ be such that if $\cos x > 1 - 1/N$ then $x \leq 2\pi/(p+1)$, for all $x \in [-\pi, \pi]$. Define:

$$I_0 \overset{\text{def}}{=} \left\{ z \in \mathbb{T} \ : \ \cos(\arg z) > 1 - \frac{1}{N}, \ -\pi \leq \arg z \leq \pi \right\}.$$

Using (6), clearly for all $n \geq N$,

$$u_n > 0 \qquad \Rightarrow \qquad e^{n\theta \mathrm{i}} = \lambda^n \in I_0.$$

For all $k \in \mathbb{N}$, we denote by I_k the rotation of I_0 by $-k\theta$, *i.e.* $I_k \overset{\text{def}}{=} e^{-nk\theta\mathrm{i}} I_0$. Similarly to above, for any $k \in \mathbb{N}$ and $n \geq N$,

$$u_{n+k} > 0 \qquad \Rightarrow \qquad e^{n\theta \mathrm{i}} = \lambda^n \in I_k.$$

Define $I \overset{\text{def}}{=} I_0 \cup I_1 \cup \cdots \cup I_{p-1}$, so that for all $n \geq N$,

$$u_{n+k} > 0 \text{ for some } k \in \{0, \ldots, p-1\} \qquad \Rightarrow \qquad e^{n\theta\mathrm{i}} = \lambda^n \in I. \tag{8}$$

By construction of N above, the set $\mathbb{T} \setminus I$ is non-empty and it is a finite union of intervals, so in particular it has non-empty interior U. It follows from an analogue of Lemma 3.6 that for some $n \in \mathbb{N}$, $\lambda^n \in U$. This in turn means, using the contrapositive of (8), that starting from u_n, the next p consecutive entries are either zero or negative. □

5 DECIDING ω-REGULAR PROPERTIES

In this section we prove the main theorem, which we recall here.

Theorem 1.2. *Given a prefix-independent ω-regular language $\mathcal{L}$ and a simple* LRS $\mathbf{u}$, *it is decidable whether the sign description of* $\mathbf{u}$ *belongs to $\mathcal{L}$.*

We will first explain what prefix-independent languages are, and how does an automaton that accepts such a language look like. It will be simpler to work with a finite monoid (similar to the syntactic monoid) that has all the relevant information about the automaton, so we will define this afterwards. In the end of the section we will describe the algorithm.

Prefix-independent languages are those that have the property that by modifying a word in finitely many places it is not possible to change its membership in the language. More precisely:

Definition 5.1 (Prefix-independent language). A language $\mathcal{L} \subseteq \Sigma^\omega$ is prefix-independent if for all infinite words α, α' such that we can get α' from α with finitely many insertions and deletions we have that

$$\alpha \in \mathcal{L} \qquad \Leftrightarrow \qquad \alpha' \in \mathcal{L}.$$

These are languages that have trivial right-congruence [Angluin and Fisman 2020, Section 4].

We will assume that the ω-regular language $\mathcal{L}$ is given as a deterministic Müller automaton $\mathcal{A}$, which is a tuple:

$$\langle \underbrace{Q}_{\text{states}}, \quad \underbrace{q_0}_{\text{initial state}}, \quad \underbrace{\Sigma}_{\text{alphabet}}, \quad \underbrace{\delta : Q \times \Sigma \to Q,}_{\text{transition function}} \quad \underbrace{F \subseteq \mathcal{P}(Q)}_{\text{accepting states}} \rangle.$$

The automaton accepts a word α if the unique run is such that the set of states that appear infinitely often belongs to F. We also assume that any state can be reached from the initial state.

Lemma 5.2. *Let $\mathcal{A}$ be a deterministic Müller automaton that recognizes a prefix-independent language, and α an infinite word. Then the following are equivalent:*

1. *$\mathcal{A}$ accepts α,*
2. *$\mathcal{A}$ accepts some suffix of α starting from some state,*
3. *$\mathcal{A}$ accepts every suffix of α starting from any state.*

Proof. We prove $2 \Rightarrow 3$, other directions are trivial. Let β be a suffix of α that is accepted starting from some state. Since the latter can be reached from q_0, it follows that $w\alpha$ is accepted by the automaton (started in q_0) for some finite word w. For 3, take any suffix β' of α and any state q_1; and let w' be the word that takes the automaton from state q_0 to state q_1. The words $w\beta$ and $w'\beta'$ have a suffix in common, since the language is prefix-independent, $w'\beta'$ is accepted by the automaton. □

As a consequence of this lemma, to show that the automaton accepts the sign description, we only have to prove that *some* suffix of the sign description σ is accepted by the given automaton $\mathcal{A}$ started at *some* state q_1.

Before we turn our attention to the monoid associated with the automaton, we gather here from the prequel some properties of the sign description of a given simple LRS. These are the essential properties that will be used by the algorithm.

Proposition 5.3. *Let $\mathbf{u}$ be a LRS and σ its sign description, then the following hold:*

1. *σ is almost-periodic,*
2. *there is a threshold $c \in \mathbb{N}$ such that any word that occurs in the suffix $\sigma_c\sigma_{c+1}\cdots$, occurs infinitely often in σ,*
3. *there is a procedure that inputs a finite word w and decides whether w occurs infinitely often in σ,*
4. *there is a procedure that inputs a finite word w that occurs infinitely often in σ and outputs the bound on the distances between consecutive occurrences.*

Proof. Property 1 is Theorem 3.1, Property 2 is Lemma 3.9. We prove 3.

Recall that P is the least common multiple of orders of roots of unity among the ratios of roots of the given LRS. We can effectively determine it, and moreover for $\ell \in \{0, 1, \ldots, P-1\}$ we can decide which subsequence $\mathbf{u}^{(\ell)}$ is identically zero; which we denote by $Z \subseteq \{0, 1, \ldots, P-1\}$. We construct a word w' such that

$$w' = w'(1)w'(2)\cdots w'(m), \text{ for some } m \in \mathbb{N},$$

where the factors $w'(i)$ are of length P, w occurs in w', and the factors $w'(i)$ have zeros exactly in positions Z. If this is not possible, the procedure returns **no**. The word w occurs infinitely often if and only if for infinitely many n, we have

$$\sigma_{nP}\sigma_{nP+1}\cdots\sigma_{(n+m)P-1} = w'.$$

The latter is true if and only if $U(w')$ is non-empty, according to Lemma 3.8. Since $U(w')$ is semi-algebraic (Lemma 3.10) and we can effectively construct its formula, to test whether $U(w')$ is empty we can use Tarski's algorithm, see Theorem A.1.

We now prove item 4. We continue as above and define w', so that $U(w')$ is non-empty. Since the normalized roots $\lambda_1, \ldots, \lambda_d$ are algebraic with effective descriptions, for all $k \in \mathbb{N}$, the set of points $\mathbf{z} \in \mathbb{T}_\lambda$ such that $(z_1\lambda_1^k, \ldots, z_d\lambda_d^k) \in U(w')$ is semi-algebraic, and we can effectively construct its formula φ_k. The formula for the set of points that enter $U(w')$ in at most j steps, $\varphi_{\leq j}$, is just the disjunction of formulas $\varphi_1, \ldots, \varphi_j$. Since $U(w')$ is open and non-empty, Lemma 3.7 implies that there exists some B such that any point in $\mathbb{T}_\lambda$ enters $U(w')$ in fewer than B steps. In the language above this means that $\Phi(B)$ which says:

$$\text{every element of } \mathbb{T}_\lambda \text{ satisfies } \varphi_{\leq B},$$

is true. Now to compute B, or a different (stronger) bound, we only need to find the first formula in $\langle \Phi(1), \Phi(2), \ldots \rangle$ that is true. We can do this using Tarski's algorithm. □

Fix a deterministic Müller automaton $\mathcal{A}$ and a LRS $\mathbf{u}$ with sign description σ for the rest of this section. We provide a "wrapper" for the procedures in properties 3 and 4 in the proposition above. There is a procedure "inter" that inputs a word w and outputs a finite set of words, or **no**:

$$w \mapsto \begin{cases} \textbf{no} & \text{if } w \text{ does not occur infinitely often in } \sigma, \\ \{w_1, w_2, \ldots, w_k\} & \text{otherwise,} \end{cases}$$

such that, in the case when w occurs infinitely often in σ,

$$\sigma = r\ w\ w_{i_1}\ w\ w_{i_2}\ \cdots,$$

where r is some finite prefix and $i_1, i_2, \ldots$ take values in $\{1, 2, \ldots, k\}$.

Intuitively, from the almost periodicity of σ, when w occurs infinitely often, the distance between the occurrences is bounded, hence there can be only finitely many words that appear between consecutive occurrences of w, using the procedure in Property 3 of Proposition 5.3, we can find these words that appear between occurrences of w, and it is this set of words that the procedure "inter" returns.

5.1 The finite monoid associated to $\mathcal{A}$

Denote by $\mathfrak{M}$ the following monoid. Its elements are directed and labeled graphs, where the set of vertices is Q (the set of states of the automaton), and the edges are labeled by subsets of Q. The product of the element $\bar{x}$ with the element $\bar{y}$ is defined as follows: for some $q_1, q_2 \in Q$ and $S_1, S_2 \subseteq Q$

$$\underbrace{q_1 \xrightarrow{S_1 \cup S_2} q_2,}_{\text{in } \bar{x}\cdot\bar{y}}$$

if and only if there exists some $q' \in Q$ such that:

$$\underbrace{q_1 \xrightarrow{S_1} q'}_{\text{in } \bar{x}} \text{ and } \underbrace{q' \xrightarrow{S_2} q_2}_{\text{in } \bar{y}}$$

The homomorphism h, is defined as follows. For any letter a of the alphabet (which in our case is $\{-,0,+\}$), $h(a)$ is such that for all $q_1, q_2 \in Q$

$$\underbrace{q_1 \xrightarrow{\{q_1,q_2\}} q_2,}_{\text{in } h(a)}$$

if and only if there is a transition in the automaton $\mathcal{A}$ from q_1 to q_2 with the letter a. The monoid $\mathfrak{M}$ is the monoid that is generated by $\{h(a) \ : \ a \in \{-,0,+\}\}$ as well as $h(\epsilon)$, where ϵ is the empty word, for the neutral element.

This monoid is a classical object in algebraic formal language theory, it gathers all the information needed from the automaton $\mathcal{A}$, *e.g.* if there is an edge from s_1 to s_2 in $h(w)$, labeled by S, it means that in the automaton, we can go from state s_1 to state s_2 with the word w while visiting all the states in S.

In our case, where the automaton $\mathcal{A}$ is deterministic, the elements of the monoid $\mathfrak{M}$ are particularly simple in that in every $\bar{x}$, and every $q \in Q$, there is only one outgoing edge from q. Therefore it makes sense to talk about the *states that are seen from* q in $\bar{x}$, *i.e.* the set S that is the label of the unique outgoing transition of from q in $\bar{x}$.

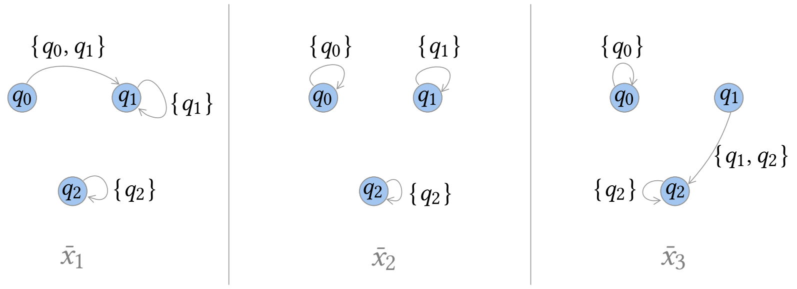
Definition 5.4 (Increasing product). Let $\bar{x}_1, \ldots, \bar{x}_k \in \mathfrak{M}$. We say that the product

$$\bar{x} = \bar{x}_1\bar{x}_2\cdots\bar{x}_k$$

is increasing, if there exists some $q \in Q$ such that the states that are seen from q in $\bar{x}_1$ is a strict subset of the states that are seen from q in $\bar{x}$.

In terms of the automaton $\mathcal{A}$, this definition means that we visit strictly more states by reading a word associated to $\bar{x}_2\cdots\bar{x}_k$, than we do by reading a word associated to $\bar{x}_1$.

Example 5.5. Consider the following elements.



The product $\bar{x}_1\bar{x}_2$ is not increasing, however the product $\bar{x}_1\bar{x}_2\bar{x}_3$ is increasing, because of the path:

$$q_0 \xrightarrow{\{q_0,q_1\}} q_1 \xrightarrow{\{q_1\}} q_1 \xrightarrow{\{q_1,q_2\}} q_2.$$

So in the product $\bar{x}_1\bar{x}_2\bar{x}_3$, the states that are seen from q_0 are $\{q_0,q_1,q_2\}$. But in $\bar{x}_1$, the states that are seen from q_0 are $\{q_0,q_1\}$, a strict subset. Similarly, the product $\bar{x}_2\bar{x}_3$ is also increasing.

We make the following observation about increasing products before we move on to the description of the algorithm.

Lemma 5.6. *Let $\bar{x}_1\bar{x}_2\cdots\bar{x}_k$ be an increasing product. Then there exists $i \in \{1,\ldots,k-1\}$ such that:*

$$\bar{x}_i\bar{x}_{i+1} \textit{ is increasing.}$$

Moreover, for all $1 \le r \le i$ and $i < r' \le k$ the product

$$\bar{x}_r\bar{x}_{r+1}\cdots\bar{x}_{r'} \textit{ is increasing.}$$

Proof. Since $\bar{x}_1\bar{x}_2\cdots\bar{x}_k$ is increasing there exists a state q and a path

$$\underbrace{q \xrightarrow{S_1} q_1}_{\text{in } \bar{x}_1} \underbrace{\xrightarrow{S_2} q_2}_{\text{in } \bar{x}_2} \cdots \underbrace{\xrightarrow{S_j} q_j}_{\text{in } \bar{x}_j},$$

such that S_j is the first set that is not a subset of S_1. Let $i \overset{\text{def}}{=} j-1$. Now the lemma follows from the fact that multiplying an increasing product to the right with elements of the monoid gives us again an increasing product. □

5.2 Description of the algorithm

The algorithm only uses the monoid $\mathfrak{M}$ and the procedure inter. It starts with some word w that occurs infinitely often in σ (*e.g.* a letter). Let $\text{inter}(w) = \{w_1, \ldots, w_k\}$. Among words:

$$w\ w_i\ w\ w_j,$$

where $i, j \in \{1, \ldots, k\}$ that occur infinitely often in σ, it tries to find one such that

$$h(w)\ h(w_i)\ h(w)\ h(w_j) \text{ is increasing.}$$

If it manages to find such a word ww_iww_j, it calls $\text{inter}(ww_iww_j)$ and repeats the steps above. Since for every state $q \in Q$, the states seen from q in $h(w)$ is a subset of the states that are seen from q in $h(ww_iww_j)$, it follows that in fewer than $|Q|^2$ iterations, a fix-point is reached: a word u, with $\text{inter}(u) = \{u_1, \ldots, u_c\}$ such that among all words

$$u\ u_i\ u\ u_j,$$

that occur infinitely often in σ,

$$h(u)\ h(u_i)\ h(u)\ h(u_j) \text{ is } \textit{not} \text{ increasing.} \tag{9}$$

We give the crucial property of the word u before proceeding with the last step of the algorithm.

Since $\text{inter}(u) = \{u_1, \ldots, u_c\}$ there exists some suffix of σ that is equal to

$$\sigma' \overset{\text{def}}{=} u\ u_{i_1}\ u\ u_{i_2} \cdots,$$

where i_n take values in $\{1, \ldots, c\}$, and moreover from Property 2 in Proposition 5.3, we can assume that σ' is such that any word that occurs in σ', occurs infinitely often.

Observe that for all $n \in \mathbb{N}$

$$h(u)\ h(u_{i_1})\ h(u)\ h(u_{i_2}) \cdots h(u)\ h(u_{i_n}) \text{ is } \textit{not} \text{ increasing.} \tag{10}$$

Indeed, if it were an increasing product, we would be able to find a short one, because of Lemma 5.6, in particular we would be able to find an increasing product of the form $h(u)h(u_i)h(u)h(u_j)$. By construction of σ' the word uu_iuu_j occurs infinitely often in σ' (and therefore also in σ), which contradicts (9).

In the last step, the algorithm picks an edge

$$q \xrightarrow{S} q'$$

in $h(u)$, such that S has minimal cardinality out of all the other labels of edges in $h(u)$. It returns **yes** if and only if $S \in F$, where F is the collection of accepting sets of states in the definition of the automaton $\mathcal{A}$.

For the correctness of the algorithm we prove the following claim.

Lemma 5.7. *When the automaton $\mathcal{A}$ reads the word σ' starting from the state q, the set of states that is seen infinitely often is S.*

This suffices because the automaton $\mathcal{A}$ accepts the infinite word σ if and only if it accepts the word σ' starting from state q, thanks to Lemma 5.2.

Proof. Observation (10) implies that the run of the automaton $\mathcal{A}$ starting from q and reading the infinite word σ' looks as follows:

$$q \xrightarrow{\text{visits } S} q_1 \xrightarrow{\text{visits } S_1} q_2 \xrightarrow{\text{visits } S} q_3 \xrightarrow{\text{visits } S_2} q_4 \xrightarrow{\text{visits } S} \cdots,$$

where $S_n \subseteq S$; and the reason why in every second edge the label is S is because there is no label that is a subset of S in $h(u)$ (except itself), because we have chosen S to have minimal cardinality out of all other labels in $h(u)$. □

We have thus proved Theorem 1.2. As for Theorem 1.1, the circumstances are simpler. When the LRS $\mathbf{u}$ is fixed, so is the bound $\tilde{n}_0$, defined as the maximum of the bounds from Lemma 3.3 over all $0 \leq \ell < P$. The number $\tilde{n}_0$ together with Property 4 in Proposition 5.3 mean that for any word w we can effectively determine the bound p in Definition 2.4. As a consequence, we can apply Theorem 2.5.

6 EXTENSIONS

Theorem 1.2 can be extended in a couple of ways: (a) it is possible to decide properties of the product of multiple sign descriptions, corresponding to different LRS, and (b) instead of predicates that speak about the sign of the entries, we can have general polynomial inequalities. We have chosen to present the specialized Theorem 1.2, so as not to obscure the principal ideas.

6.1 Multiple LRS

We are given m simple linear recurrence sequences:

$$\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(m)}, \tag{11}$$

of orders $d^{(1)}, d^{(2)}, \ldots, d^{(m)}$ respectively. The product of their sign descriptions $\sigma^{(1)} \times \cdots \times \sigma^{(m)}$ is an infinite word $\tilde{\sigma}$ over the alphabet $\{-, 0, +\}^m$. We explain how the proof in the prequel can be adapted to prove a generalisation of Theorem 1.2, where $\mathbf{u}$ is replaced by (11), and instead of the sign description σ, we have $\tilde{\sigma}$, the product of sign descriptions.

Define:

$$\tilde{\lambda} \stackrel{\text{def}}{=} \left(\lambda_1^{(1)}, \ldots, \lambda_{d^{(1)}}^{(1)}, \lambda_1^{(2)}, \ldots, \lambda_{d^{(2)}}^{(2)}, \quad \ldots \quad, \lambda_1^{(m)}, \ldots, \lambda_{d^{(m)}}^{(m)}\right),$$

where $\lambda_i^{(k)}$ are the normalized roots of $\mathbf{u}^{(k)}$ as in (3). The subgroup of the torus, $\mathbb{T}_{\tilde{\lambda}}$, and the successor function $\tilde{s} : \mathbb{T}_{\tilde{\lambda}} \to \mathbb{T}_{\tilde{\lambda}}$ are defined as expected[4]. Lemma 3.6 can be applied to $\mathbb{T}_{\tilde{\lambda}}$, consequently the set $\{\tilde{s}^n(1, \ldots, 1) : n \in \mathbb{N}\}$ is a dense subset of $\mathbb{T}_{\tilde{\lambda}}$, and any open subset of the torus can be reached in bounded number of steps.

Let $k \in \{1, \ldots, m\}$, and denote by $P^{(k)}$ the least common multiple of the orders of all roots of unity among the ratios of the roots of $\mathbf{u}^{(k)}$, as defined in the beginning of Section 3. Applying Lemma 3.3 to the subsequences of $\mathbf{u}^{(k)}$ and combining the results yields the following. There exists a linear function f, from $\mathbb{T}_{\lambda^{(k)}}$ to $\mathbb{R}^{P^{(k)}}$ and a threshold $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,

$$\operatorname{sgn}\Big(f\left(s^n(1, \ldots, 1)\right)\Big) = \sigma^{(k)}\left[nP^{(k)},\ (n+1)P^{(k)}\right],$$

[4] There is another option of defining $\mathbb{T}_{\tilde{\lambda}}$ as the product of $\mathbb{T}_{\lambda^{(i)}}$. However this is a different object, not suitable for our needs; in particular the hypothesis of Theorem 3.5 is invalidated.

where sgn is applied component-wise and the word on the right hand side is the factor of $\sigma^{(k)}$ that starts in position $nP^{(k)}$ and ends in position $(n+1)P^{(k)}$. Let $\tilde{P}$ be the least common multiple of $P^{(1)}, \dots, P^{(m)}$. We can glue together the linear functions above to build another linear function $g : \mathbb{T}_{\tilde{\lambda}} \to \mathbb{R}^{\tilde{P}}$ such that the following holds. There exists a threshold, after which, for all n,

$$\operatorname{sgn}\Big(g\big(\tilde{s}^n(1,\dots,1)\big)\Big) = \tilde{\sigma}\big[n\tilde{P},\ (n+1)\tilde{P}\big].$$

After establishing this fact, the rest of the proof depends only on the objects $\mathbb{T}_{\tilde{\lambda}}$, $\tilde{s}$ and the linear function g. As a consequence we have the following theorem.

Theorem 6.1. *Given a prefix-independent ω-regular language $\mathcal{L}$ and simple linear recurrence sequences $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(m)}$, it is decidable whether the product of sign descriptions $\tilde{\sigma}$, belongs to $\mathcal{L}$.*

6.2 Semi-algebraic predicates

Theorem 6.1 already allows us to decide properties such as: "only finitely many times, does the sequence pass from a value smaller than 3, to a value larger than 10, without having the value 5 in between".

Example 6.2. More precisely, there are only finitely many $n \in \mathbb{N}$ such that

$$\exists m > n, u_n < 3 \text{ and } u_m > 10, \text{ moreover for every } r, n \le r \le m, u_r \ne 5. \tag{12}$$

This can be done as follows. Define:

$$u^{(1)} \stackrel{\text{def}}{=} \langle u_n - 3\rangle_{n\in\mathbb{N}}, \qquad u^{(2)} \stackrel{\text{def}}{=} \langle u_n - 5\rangle_{n\in\mathbb{N}}, \qquad u^{(3)} \stackrel{\text{def}}{=} \langle u_n - 10\rangle_{n\in\mathbb{N}}.$$

These sequences are linear recurrence sequences, because LRS are closed under addition and product of sequences. Consider the product $\tilde{\sigma}$, of sign descriptions of the LRS above. This is an infinite word over the alphabet $\{-,0,+\}^3$, so a letter looks like: $(-,+,-)$. The property (12), in terms of $\tilde{\sigma}$, can be expressed as: "only finitely many times does a letter of the type $(-,*,*)$ appear followed by a letter of the type $(*,*,+)$ without having a letter of type $(*,0,*)$ in between". This is a prefix-independent ω-regular property.

In the example above we have shifted the sequence by constants 3, 5, and 10, to produce new LRS. More generally, LRS are closed under sequence addition and product:

Proposition 6.3. *Let $\mathbf{u}, \mathbf{v}$ be two linear recurrence sequences. Both $\langle u_n + v_n\rangle_{n\in\mathbb{N}}$ and $\langle u_n v_n\rangle_{n\in\mathbb{N}}$ are linear recurrence sequences as well.*

As a consequence of this proposition, given LRS $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(m)}$, and a polynomial $F \in \mathbb{Z}[x_1, \dots, x_m]$, the sequence $\mathbf{u}(F) := \langle F(u_n^{(1)}, \dots, u_n^{(m)})\rangle_{n\in\mathbb{N}}$ is a LRS. It follows that we can have predicates that are polynomial inequalities. In other words, for polynomials $F_1, \dots, F_k$, we can construct the sequences $\mathbf{u}(F_1), \dots, \mathbf{u}(F_k)$ and apply Theorem 6.1.

We can go one step further, and define predicates which are equal to membership in a semi-algebraic set (Appendix A). We explain this more precisely. Let

$$S_1, \dots, S_k \subseteq \mathbb{R}^m,$$

be semi-algebraic sets. Define $\mathbf{S_i}$, $i \in \{1, \dots, k\}$, a predicate on naturals, to be true for $n \in \mathbb{N}$ if and only if

$$\left(u_n^{(1)}, \dots, u_n^{(m)}\right) \in S_i.$$

The *zone description* of $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(m)}$ with respect to $\mathrm{S_1}, \ldots, \mathrm{S_k}$ is the infinite word τ over the alphabet $\mathcal{P}(\{\mathrm{S_1}, \ldots, \mathrm{S_k}\})$, defined for all $n \in \mathbb{N}$ as:

$$\tau_n \text{ is the subset of predicates } \{\mathrm{S_1}, \ldots, \mathrm{S_k}\} \text{ that are true in } n.$$

Since quantifiers can be eliminated in first order logic of real closed fields, membership in a semi-algebraic set reduces to fulfilling a finite set of polynomial inequalities. As we have already shown how we are able to apply Theorem 6.1 on predicates that are polynomial inequalities we have the following theorem.

Theorem 6.4. *Given semi-algebraic sets $S_1, \ldots, S_k$, a prefix-independent ω-regular language $\mathcal{L}$, and simple linear recurrence sequences $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(m)}$, it is decidable whether the zone description of the sequences with respect to $\mathrm{S_1}, \ldots, \mathrm{S_k}$ belongs to $\mathcal{L}$.*

REFERENCES

Manindra Agrawal, S. Akshay, Blaise Genest, and P. S. Thiagarajan. 2015. Approximate Verification of the Symbolic Dynamics of Markov Chains. *J. ACM* 62, 1 (2015), 2:1–2:34.

Dana Angluin and Dana Fisman. 2020. Regular omega-languages With an Informative Right Congruence. *Information and Computation* (2020). https://doi.org/10.1016/j.ic.2020.104598

P. T. Bateman, C. G. Jockusch, and A. R. Woods. 1993. Decidability and undecidability of theories with a predicate for the primes. *Journal of Symbolic Logic* 58, 2 (June 1993), 672–687.

Danièle Beauquier, Alexander Moshe Rabinovich, and Anatol Slissenko. 2006. A Logic of Probability with Decidable Model Checking. *J. Log. Comput.* 16, 4 (2006), 461–487.

J. P. Bell and S. Gerhold. 2007. On the Positivity Set of a Linear Recurrence. *Israel Jour. Math.* 57 (2007).

Jean Berstel and Maurice Mignotte. 1976. Deux propriétés décidables Des Suites récurrentes linéaires. *Bulletin de la Societe mathematique de France* 79 (1976), 175–184. https://doi.org/10.24033/bsmf.1823

J. Richard Büchi. 1962. On a decision method in restricted second order arithmetic. In *Logic, Methodology and Philosophy of Science (Proc. 1960 Internat. Congr .).* Stanford Univ. Press, Stanford, Calif., 1–11.

Olivier Carton and Wolfgang Thomas. 2002. The Monadic Theory of Morphic Infinite Words and Generalizations. *Information and Computation* 176, 1 (2002), 51–65. https://doi.org/10.1006/inco.2001.3139

John William Scott Cassels. 1957. *An introduction to Diophantine approximation.* Number 45. CUP Archive.

Calvin C. Elgot and Michael O. Rabin. 1966. Decidability and Undecidability of Extensions of Second (first) Order Theory of (generalized) Successor. *The Journal of Symbolic Logic* 31, 02 (1966), 169–181. https://doi.org/10.2307/2269808

Graham Everest, Alf van der Poorten, Igor Shparlinski, and Thomas Ward. 2003. Recurrence Sequences. https://doi.org/10.1090/surv/104

Toghrul Karimov, Joel Ouaknine, and James Worrel. 2020. On LTL Model Checking for Low-Dimensional Discrete Linear Dynamical Systems. In *45th International Symposium on Mathematical Foundations of Computer Science, MFCS 2020, LIPIcs 170.*

Serge Lang. 1995. *Introduction to Diophantine Approximations.* Springer New York. https://doi.org/10.1007/978-1-4612-4220-8

David W Masser. 1988. Linear relations on algebraic groups. *New Advances in Transcendence Theory* (1988), 248–262.

Marston Morse and Gustav A. Hedlund. 1938. Symbolic Dynamics. *American Journal of Mathematics* 60, 4 (1938), 815. https://doi.org/10.2307/2371264

An. Muchnik, A. Semenov, and M. Ushakov. 2003. Almost Periodic Sequences. *Theoretical Computer Science* 304, 1-3 (2003), 1–33. https://doi.org/10.1016/s0304-3975(02)00847-2

Joël Ouaknine and James Worrell. 2013. Positivity Problems for Low-Order Linear Recurrence Sequences. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms.* https://doi.org/10.1137/1.9781611973402.27

Joël Ouaknine and James Worrell. 2014a. On the Positivity Problem for Simple Linear Recurrence Sequences. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 8573),* Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias (Eds.). Springer, 318–329.

Joël Ouaknine and James Worrell. 2014b. Ultimate Positivity is Decidable for Simple Linear Recurrence Sequences. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 8573).* Springer, 330–341.

Alexander Rabinovich. 2007. On Decidability of Monadic Logic of Order Over the Naturals Extended By Monadic Predicates. *Information and Computation* 205, 6 (2007), 870–889. https://doi.org/10.1016/j.ic.2006.12.004

A. Salomaa and M. Soittola. 1978. *Automata-theoretic aspects of formal power series.* Springer.

A L Semënov. 1984. Logical Theories of One-Place Functions on the Set of Natural Numbers. *Mathematics of the USSR-Izvestiya* 22, 3 (1984), 587–618. https://doi.org/10.1070/im1984v022n03abeh001456

Harold N. Shapiro. 1959. On a Theorem Concerning Exponential Polynomials. *Communications on Pure and Applied Mathematics* 12, 3 (1959), 487–500. https://doi.org/10.1002/cpa.3160120306

Alfred Tarski. 1951. A decision method for elementary algebra and geometry. (1951).

R. Tijdeman, M. Mignotte, and T.N. Shorey. 1984. The Distance Between Terms of an Algebraic Recurrence Sequence. *Journal für die reine und angewandte Mathematik (Crelles Journal)* 1984, 349 (1984), 63–76. https://doi.org/10.1515/crll.1984.349.63

A FIRST-ORDER THEORY OF REAL CLOSED FIELDS

In the first-order logic of real closed fields the atomic formulas are of the form:

$$p(x_1, \ldots, x_k) \sim 0,$$

where p is a polynomial in $\mathbb{Z}[x_1, \ldots, x_k]$ and $\sim \in \{>, =\}$. In this logic we are allowed to quantify over real numbers, and use the Boolean connectives. Subsets of $\mathbb{R}^k$ that satisfy a formula $\Phi(x_1, \ldots, x_k)$ of this logic are called *semi-algebraic sets.* The first-order logic of real closed fields admits effective quantifier elimination, a fact known as the Tarski's theorem, which we state as follows.

Theorem A.1 ([Tarski 1951, Theorem 37]). *There is an algorithm that inputs a sentence Φ (a formula without free variables) from the language above, and returns yes if and only if Φ is true over the real numbers.*

There is a natural first-order interpretation of the field of complex numbers into the field of real numbers, where every complex variable $z = x + \mathbf{i}y$ is replaced by two real variables x and y.

Example A.2. Consider the polynomial $p(z) := z^3 + 5z$. Its roots are $\lambda_1 := 0$, $\lambda_2 := \mathbf{i}\sqrt{5}$, and $\lambda_3 := -\mathbf{i}\sqrt{5}$. The algebraic number λ_2 can be identified with the formula (to be interpreted over $\mathbb{C}$) $\phi_2(z)$ which says $p(z) = 0$ and $\mathrm{Im}(z) > 2$. Using the interpretation alluded above, we replace $\phi_2(z)$ with $\phi_2'(x, y)$ (where x and y range over reals) which says $p_1(x, y) = p_2(x, y) = 0$ and $y > 2$ where

$$p_1(x, y) \stackrel{\text{def}}{=} x^3 - 3xy^2 + 5x, \qquad p_2(x, y) \stackrel{\text{def}}{=} 3x^2y - y^3 + 5y.$$

Clearly

$$\{z \in \mathbb{C} \ : \ p(z) = 0 \text{ and } \mathrm{Im}(z) > 2\} = \{x + \mathbf{i}y \in \mathbb{C} \ : \ p_1(x, y) = p_2(x, y) = 0 \text{ and } y > 2\}.$$

The intervals where the normalized roots of the characteristic polynomial associated to a LRS lay, can be computed, therefore we assume that $\lambda = (\lambda_1, \ldots, \lambda_d)$ are given by formulas as in the example above. One can take products and sums of such numbers, *i.e.* compute a different formula which defines the product or sum. We will prove that the set $\mathbb{T}_\lambda$ (defined in Section 3.2) is semi-algebraic. After this, we will give a full proof of Lemma 3.10.

Lemma A.3. *The set*

$$\left\{(x_1, y_1, \ldots, x_d, y_d) \in \mathbb{R}^{2d} \ : \ (x_1 + \mathbf{i}y_1, \ldots, x_d + \mathbf{i}y_d) \in \mathbb{T}_\lambda\right\}$$

is semi-algebraic.

Proof. The set of multiplicative relations of $\lambda = (\lambda_1, \ldots, \lambda_d)$ is:

$$\mathcal{M}_\lambda \stackrel{\text{def}}{=} \{\mathbf{v} \in \mathbb{Z}^d \ : \ \lambda_1^{v_1}\lambda_2^{v_2} \cdots \lambda_d^{v_d} = 1\}.$$

This is an Abelian subgroup of $(\mathbb{Z}^d, +)$, hence it has a finite basis $B \subset \mathbb{Z}^d$. Masser gave an explicit upper bound on the components of this basis in [Masser 1988, Section 4]. As a consequence, B is computable: to test whether some $(b_1, \ldots, b_d)$ is in B, or equivalently whether $\lambda_1^{b_1} \cdots \lambda_d^{b_d} = 1$, use Theorem A.1. We recall the definition of $\mathbb{T}_\lambda$:

$$\mathbb{T}_\lambda \overset{\text{def}}{=} \{\mathbf{z} \in \mathbb{T}^d \ : \ z_1^{v_1} z_2^{v_2} \cdots z_d^{v_d} = 1 \text{ for all } \mathbf{v} \in \mathcal{M}_\lambda\}.$$

Since B is a basis of $\mathcal{M}_\lambda$, we can replace $\mathcal{M}_\lambda$ by B in the definition above. So $\mathbb{T}_\lambda$ is the set of all $(x_1, y_1, \ldots, x_d, y_d)$ such that for every $(b_1, \ldots, b_d) \in B$ we have:

$$(x_1 + \mathbf{i}y_1)^{b_1}(x_2 + \mathbf{i}y_2)^{b_2} \cdots (x_d + \mathbf{i}y_d)^{b_d} - 1 = 0.$$

Since this is a finite set of equations, the lemma follows. □

We give a full proof of Lemma 3.10.

Lemma 3.10. *For all w, $U(w)$ is semi-algebraic, and we can compute the first-order formula that defines it.*

Proof. Let $P, m \in \mathbb{N}$ and $w = w(1)w(2)\cdots w(m) \in \{-, 0, +\}^*$ be such that $w(i)$ are factors of length P. We recall the definitions from Section 3.3.

$$U(w) \overset{\text{def}}{=} \{\mathbf{x} \in \mathbb{T}_\lambda \ : \ g(\mathbf{x}) = w(1), g\left(s(\mathbf{x})\right) = w(2), \ldots, g\left(s^{m-1}(\mathbf{x})\right) = w(m)\},$$

where $g \ : \ \mathbb{T}_\lambda \to \{-, 0, +\}^P$ is the composition of f and sgn applied component-wise, f is the linear map $f \ : \ \mathbb{T}_\lambda \to \mathbb{R}^P$ that we get after applying Lemma 3.3 to every subsequence $\mathbf{u}_\ell$, and finally s maps $(x_1, \ldots, x_d)$ to $(\lambda_1 x_1, \ldots, \lambda_d x_d)$. Since we have formulas for λ_i, we can compute formulas for $s^k(\mathbf{x})$, for any $k \in \mathbb{N}$. Inspecting the proof of Lemma 3.3, reveals that the constants $z_1, \ldots, z_d$ have computable formulas, hence the same holds for $f(s^k(\mathbf{x}))$, *i.e.* we can compute $\varphi_k(\mathbf{x}, \mathbf{y})$, such that

$$\varphi_k(x_1, \ldots, x_d, y_1, \ldots, y_P) \Leftrightarrow \mathbf{y} = f(s^k(\mathbf{x})).$$

To require that $g(s^k(\mathbf{x})) = v$ for some $v \in \{-, 0, +\}^P$, we use the formula φ_k and ask that $y_i \sim 0$ where $\sim \in \{<, =, >\}$ depending on whether v_i is "−", "0", or "+". Since the set of $\mathbf{x} \in \mathbb{T}_\lambda$ is semi-algebraic thanks to Lemma A.3, and we have constructed formulas that require $g(s^k(\mathbf{x})) = v$, the lemma is proved. □