The Monadic Theory of Toric Words

Valérie Berthé[a], Toghrul Karimov[b], Joël Ouaknine[b], Mihir Vahanwala[b], James Worrell[c]

[a] *Université de Paris, IRIF, CNRS, Paris, F-75013, France*
[b] *Max Planck Institute for Software Systems, Saarland Informatics Campus, Saarbrücken, 66123, Germany*
[c] *University of Oxford, Department of Computer Science, Oxford, OX1 3QG, United Kingdom*

Abstract

For which unary predicates $P_1, \ldots, P_m$ is the MSO theory of the structure $\langle \mathbb{N}; <, P_1, \ldots, P_m \rangle$ decidable? We survey the state of the art, leading us to investigate combinatorial properties of almost-periodic, morphic, and toric words. In doing so, we show that if each P_i can be generated by a toric dynamical system of a certain kind, then the attendant MSO theory is decidable.

Keywords: Monadic second-order logic, morphic words, toric words, Pisot conjecture, dynamical systems, linear recurrence sequences

1. Introduction

In 1962, Büchi proved in his seminal work [1] that the monadic second-order (MSO) theory of the structure $\langle \mathbb{N}; < \rangle$ is decidable. Shortly thereafter, in 1966, Elgot and Rabin [2] showed how to decide the MSO theory of $\langle \mathbb{N}; <, P \rangle$ for various interesting unary predicates P. On the other hand, it was known already in the 1960s that extending $\langle \mathbb{N}; < \rangle$ with the addition or even the doubling function yields a structure with an undecidable MSO theory [3, 4]. In this paper, we focus on the following question: Which unary predicates $P_1, \ldots, P_m$ can one add to $\langle \mathbb{N}; < \rangle$ whilst maintaining decidability of the MSO theory? We give an overview of the state of the art and provide some new answers. In particular, we identify a class of predicates generated by rotations on a torus, any number of which can be adjoined to $\langle \mathbb{N}; < \rangle$ and still preserve decidability of the attendant monadic theory.

By a predicate P we mean a function with type $\mathbb{N} \to \Sigma$, where Σ is a finite alphabet. The *characteristic word* of P is the word $\alpha \in \Sigma^\omega$ whose nth letter is $P(n)$. Let us take the primes predicate as an example, defined by $P(n) = 1$ if n is prime and $P(n) = 0$ otherwise. Recall that in a monadic second-order language we have access to the membership predicate $\in$ and quantification over elements (written Qx for a quantifier Q) as well as subsets of the universe (written QX), which is $\mathbb{N}$ in our case. Consider the sentence ψ given by

$$\begin{aligned} \varphi(X) &\coloneqq 1 \in X \ \wedge\ 0, 2 \notin X \ \wedge\ \forall x.\, x \in X \Leftrightarrow s(s(s(x))) \in X \\ \psi &\coloneqq \exists X \colon \varphi(X) \ \wedge\ \forall y.\, \exists z > y \colon z \in X \ \wedge\ P(z) \end{aligned}$$

where $s(\cdot)$ is the successor function defined by $s(x) = y$ if and only if

$$x < y \ \wedge\ \forall z.\, x < z \Rightarrow y \leq z.$$

The formula φ defines the subset $\{n \colon n \equiv 1 (\mathrm{mod}\ 3)\}$ of $\mathbb{N}$, and ψ is the sentence "there are infinitely many primes congruent to 1 modulo 3", which is true. Another example of a number-theoretic statement expressible in our setting would be the Twin Primes Conjecture, which is given by the first-order sentence

$$\forall x.\, \exists y > x \colon P(y) \ \wedge\ P(s(s(y))).$$

Unsurprisingly, decidability of the MSO theory of the structure $\langle \mathbb{N}; <, P\rangle$, where P is the primes predicate, remains open. Conditional decidability is known subject to Schinzel's Hypothesis, a number-theoretic conjecture that implies existence of infinitely many twin primes [5].

The MSO theory of $\mathbb{N}$ equipped with the order relation is intimately connected to the theory of finite automata. The *Acceptance Problem* for a word $\alpha \in \Sigma^\omega$, denoted Acc_α, is to decide, given a deterministic (e.g. Muller) automaton $\mathcal{A}$ over Σ, whether $\mathcal{A}$ accepts α. The previously mentioned result of Büchi establishes that the MSO theory of $\langle \mathbb{N}; <, P_1, \ldots, P_m\rangle$ is decidable if and only if Acc_α is decidable for the word $\alpha = \alpha_1 \times \ldots \times \alpha_m$, where each α_i is the characteristic word of P_i.[1] Hence our central question can be reformulated as follows: for which classes of words $\alpha_1, \ldots, \alpha_m$ is Acc_α decidable?

[1] The original formulation by Büchi was given in terms of non-deterministic Büchi automata. The formulations involving deterministic automata with a Muller, Rabin, or parity acceptance condition are equivalent.

In this work we consider the classes of *almost-periodic*, *morphic*, and *toric* words. Almost-periodic words were introduced by Semënov in [6]. He showed that for an *effectively* almost-periodic word α, the MSO theory of the structure $\langle \mathbb{N}; <, P_\alpha \rangle$ is decidable, where P_α is the predicate whose characteristic word is α. We discuss almost-periodic words in Section 3. We then move onto morphic words (Section 4), focussing on the result [7] of Carton and Thomas that for a morphic word α, the MSO theory of $\langle \mathbb{N}; <, P_\alpha \rangle$ is decidable. These two works provide answers to our main question for a single predicate, i.e. in the case of $m = 1$.

In Section 5, we introduce the class of toric words, which are codings of a rotation with respect to target sets consisting of finitely many connected components. In Theorem 5.12, we give a large class $\mathcal{K}$ of toric words such that the MSO theory of the structure $\langle \mathbb{N}; <, P_1, \ldots, P_m \rangle$ is decidable for any number m of predicates with characteristic words belonging to $\mathcal{K}$. We also study almost periodicity and closure properties toric words (Section 5.3), and give account of the overlap between toric words and various other well-known families of words. The latter is summarised below.

(a) Sturmian words are toric. In Section 6.1 we use the theory of toric words to show that for Sturmian words $\alpha_1, \ldots, \alpha_m$ that satisfy a certain effectiveness assumption, the MSO theory of $\langle \mathbb{N}; <, P_{\alpha_1}, \ldots, P_{\alpha_m} \rangle$ is decidable. This answers a question of [7].

(b) One of the central problems in symbolic dynamics is to understand the morphic words for which the associated *shift space* has a representation as a *geometric dynamical system* [8]. A slightly different (but similar in spirit) question is: Which morphic words are toric? The Pisot conjecture identifies a class of morphic words for which the answer to the first question is expected to be positive. We discuss the conjecture and its relation to the second question in Section 6.2.

(c) Toric words arise naturally in the study of linear recurrence sequences. In fact, specialised classes of toric words have already been used in the literature [9, 10, 11] to study *sign patterns* of linear recurrence sequences, discussed in Section 6.3. We construct sign patterns of LRS that prove that the product of an almost-periodic word with a toric word that is almost-periodic need not be almost-periodic.

(d) Finally, in Section 6.4 we give an overview of how modelling sign patterns of LRS using toric words yields decision procedures for the *Model-Checking Problem* for linear dynamical systems.

2. Mathematical background

By an alphabet Σ we mean a non-empty finite set. For a word $\alpha \in \Sigma^+ \cup \Sigma^\omega$ we denote by $\alpha(n)$ its nth letter. For $\alpha \in \Sigma^\omega$ we denote by P_α the predicate defined by $P_\alpha(n) = \alpha(n)$ for all n. We write $\alpha[n, m)$ for the finite word $u = \alpha(n) \cdots \alpha(m-1)$. Such u is called a *factor* of α. We write $\alpha[n, \infty)$ for the infinite word $\alpha(n)\alpha(n+1)\cdots$.

Let $\alpha_i \in \Sigma_i^\omega$ for $0 \leq i < L$. The *product* $\alpha_0 \times \ldots \times \alpha_{L-1}$ of $\alpha_0, \ldots, \alpha_{L-1}$ is the word α over the product alphabet $\Sigma_0 \times \ldots \times \Sigma_{L-1}$ defined by $\alpha(n) = (\alpha_0(n), \ldots, \alpha_{L-1}(n))$. The *merge* (alternatively, the *shuffling* or the *interleaving*) of $\alpha_0, \ldots, \alpha_{L-1}$ is the word α defined by $\alpha(nL + r) = \alpha_r(n)$ for all $n \in \mathbb{N}$ and $0 \leq r < L$. Let Σ_1, Σ_2 be two alphabets. A *morphism* $\tau \colon \Sigma_1^* \to \Sigma_2^*$ is a map satisfying $\tau(a_1, \ldots a_l) = \tau(a_1) \cdots \tau(a_l)$ for all $a_1, \ldots, a_l \in \Sigma_1$.

We write Log for the principal branch of the complex logarithm. That is, $\mathrm{Im}(\mathrm{Log}(z)) \in (-\pi, \pi]$ for all non-zero $z \in \mathbb{C}$. For $z = (z_1, \ldots, z_d) \in \mathbb{C}^d$ and $p \geq 1$, we denote by $\|z\|_p$ the ℓ_p norm $\sqrt[p]{|z_1|^p + \ldots + |z_d|^p}$.

By a $\mathbb{K}$-*semialgebraic* subset of $\mathbb{R}^d$, where $\mathbb{K} \subseteq \mathbb{R}$, we mean a set that can be defined by polynomial inequalities with coefficients belonging to $\mathbb{K}$; Recall that $p(\mathbf{x}) = 0 \Leftrightarrow p(\mathbf{x}) \geq 0 \wedge p(\mathbf{x}) \leq 0$. A set $X \subseteq \mathbb{C}^d$ is $\mathbb{K}$-*semialgebraic* if

$$\{(x_1, y_1, \ldots, x_d, y_d) \colon (x_1 + y_1 \boldsymbol{i}, \ldots, x_d + y_d \boldsymbol{i}) \in X\}$$

is a $\mathbb{K}$-semialgebraic subset of $\mathbb{R}^{2d}$.

A sequence $(u_n)_{n \in \mathbb{N}}$ over ring R is a *linear recurrence sequence* over R if there exist $d > 0$ and $(a_0, \ldots, a_{d-1}) \in R^d$ such that

$$u_{n+d} = a_0 u_n + \ldots + a_{d-1} u_{n+d-1}$$

for all $n \in \mathbb{N}$. The order of $(u_n)_{n \in \mathbb{N}}$ is the smallest d such that $(u_n)_{n \in \mathbb{N}}$ satisfies a recurrence relation from R^d. An LRS $(u_n)_{n \in \mathbb{N}}$ over R of order d can be written in the form $u_n = c^\top M^n s$, where $c, s \in R^d$ and $M \in R^{d \times d}$. If R is an integral domain, then for any $p \in R[x_1, \ldots, x_d]$, $u_n = p(M^n s)$ defines an LRS over R. This is a consequence of the Fatou Lemma [12, Chapter 7.2].

The most famous problem about LRS is the Skolem Problem (over $\mathbb{Q}$): given an LRS $(u_n)_{n \in \mathbb{N}}$ over $\mathbb{Q}$, decide whether there exists n such that $u_n = 0$. The Skolem Problem has been open for nearly ninety years, counting from the seminal work [13] of Skolem, and is currently known to be decidable for LRS (over $\mathbb{Q}$) of order 4 or less [14, 15]. A related result is the celebrated Skolem-Mahler-Lech theorem [13, 16, 17], which states that for any LRS over a field of characteristic zero, the set of zeros is a union of a finite set F and finitely many arithmetic progressions $a_1 + b_1\mathbb{N}, \ldots, a_k + b_k\mathbb{N}$, where $0 \leq a_i < b_i$ for all i. The values of k, a_i, b_i can all be effectively computed, whereas determining whether F is empty is exactly the Skolem Problem. Berstel and Mignotte showed in [18] that if an LRS $(u_n)_{n\in\mathbb{N}}$ is not identically zero, then there exists effectively computable $L \geq 1$ such that for all $0 \leq r < L$, the subsequence $(u_{nL+r})_{n\in\mathbb{N}}$ is either identically zero or has finitely many zeros. Consequently, if we assume existence of an oracle for the Skolem Problem, then we can effectively compute all elements of F in the Skolem-Mahler-Lech theorem: take L subsequences and repeatedly apply the Skolem oracle to each non-zero subsequence until all zeros have been found.

Other well-known open decision problems of LRS include the Positivity Problem (given $(u_n)_{n\in\mathbb{N}}$, decide if $u_n \geq 0$ for all n) and the Ultimate Positivity Problem (given $(u_n)_{n\in\mathbb{N}}$, decide if $u_n \geq 0$ for all sufficiently large n). These decision problems were already encountered in the 1970s by Salomaa and others when studying growth and related problems in formal languages [19, 20]. The Skolem Problem for LRS over $\mathbb{Q}$ can be reduced to the Positivity Problem for LRS over $\mathbb{Q}$, but the latter is also, independently from the Skolem Problem, hard with respect to certain open problems in Diophantine approximation [21].

3. Almost-periodic words

A word $\alpha \in \Sigma^\omega$ is *almost-periodic* if for every finite word $u \in \Sigma^*$, there exists $R(u) \in \mathbb{N}$ with the following property.

(a) Either u does not occur in $\alpha[R(u), \infty)$, or

(b) it occurs in every factor of α of length $R(u)$.

The word α is *effectively almost-periodic* if (i) $\alpha(n)$ can be effectively computed for every n, and (ii) given u, we can effectively compute a value $R(u)$ with the properties above. We represent an effectively almost-periodic word with two programs that compute $\alpha(n)$ on n and $R(u)$ on u, respectively. The word α is *strongly almost-periodic* if it is almost-periodic and every finite word u either does not occur in α, or occurs infinitely often. Strongly almost-periodic words are also known as *uniformly recurrent* words in the literature; see [22, 23]. For such words, $R(u)$ is an upper bound on the *return* *time* of u. We will see that certain morphic words, sign patterns of linear recurrence sequences, as well as large classes of toric words are almost-periodic. The characteristic word $\alpha_{n!} = 01100010000\cdots$ of the set $\{n! \mid n \in \mathbb{N}\}$ of all factorial numbers, on the other hand, is an example of a word that is not almost-periodic.

Remarkably, for an effectively almost-periodic word α the Acceptance Problem Acc_α and hence the MSO theory of the structure $\langle \mathbb{N}; <, P_\alpha \rangle$ are decidable. We refer to this result as **Semënov's theorem**:

Theorem 3.1. *Given a deterministic automaton $\mathcal{A}$ and an effectively almost-periodic word α, it is decidable whether $\mathcal{A}$ accepts α.*

See [24] for an elegant proof, showing that the sequence of states $\mathcal{A}(\alpha)$ obtained when a deterministic automaton $\mathcal{A}$ reads an effectively almost-periodic word α is also effectively almost-periodic. It remains to determine which states occur infinitely often in $\mathcal{A}(\alpha)$. This can be done by computing $R(q)$ for every state q and then checking whether q occurs in $\mathcal{A}(\alpha)[R(q), 2R(q))$.

An infinite word $\beta = \tau(\alpha)$, where α is (effectively) almost-periodic and $\tau\colon \Sigma_1^* \to \Sigma_2^*$ is a morphism, is (effectively) almost-periodic [24]. Furthermore, the product of an almost-periodic word with an ultimately periodic word is almost-periodic. However, by the result [6] of Semënov, the product of two effectively almost-periodic words need not be effectively almost-periodic. This tells us that we cannot immediately use Semënov's theorem to show decidability of the MSO theory of the structure $\langle \mathbb{N}; <, P_\alpha, P_\beta \rangle$ for effectively almost-periodic words α, β. In Section 6.3, we will give explicit words α, β that are sign patterns of linear recurrences sequences and effectively almost-periodic, whereas the product $\alpha \times \beta$ is not almost-periodic. The proof of [6], in comparison, is indirect: it constructs two effectively almost-periodic words α, β that encode information about Turing machines such that the MSO theory of $\langle \mathbb{N}; <, P_{\alpha\times\beta} \rangle$ is undecidable. It follows that the word $\alpha \times \beta$ cannot be *effectively* almost-periodic.

4. Morphic words

By *substitution* we mean a non-erasing morphism $\tau\colon \Sigma^* \to \Sigma^*$. That is, $\tau(a) \in \Sigma^+$ for all $a \in \Sigma$. Let τ be a substitution and $a \in \Sigma$ be a letter such that $\tau(a) = aw$ for some $w \in \Sigma^*$. Iterating τ on a, we obtain a sequence $(x_n)_{n\in\mathbb{N}}$ of words given by $x_0 = a$ and $x_{n+1} = aw\tau(w)\cdots\tau^n(w)$. For every $k, n \in \mathbb{N}$, x_n is a prefix of x_{n+k}. If $|\tau^n(a)| \to \infty$ as $n \to \infty$, then $(x_n)_{n\in\mathbb{N}}$

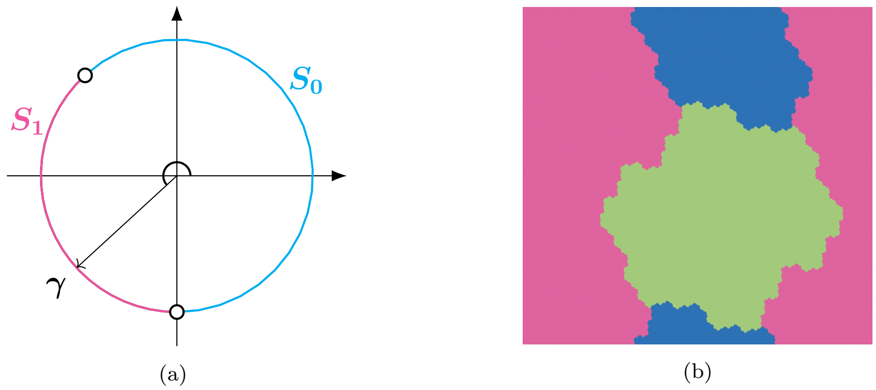

Figure 1: Target sets for the Fibonacci and Tribonacci words. In (b), the pink, green, and blue sets correspond to S_1, S_2, S_3, respectively.

converges to an infinite word $\alpha \in \Sigma^\omega$ that is a fixed point of τ. Such α is called a *substitutive* (alternatively, a *pure morphic*) word; see [23] for an account of dynamics of these words. Substitutive words are similar to and subsumed by words generated by D0L systems; the latter is obtained by iteratively applying a morphism to a word $w \in \Sigma^*$, as opposed to a single letter [25]. We next give a few well-known examples of substitutive words.

(a) The Thue-Morse sequence $0110100110\cdots$ is generated by the substitution $0 \to 01$ and $1 \to 10$, starting with the letter 0.

(b) The Fibonacci word $\alpha_F = 01001010010\cdots$, generated by the substitution $0 \to 01$ and $1 \to 0$. This famous sequence has many equivalent definitions, one of them as the coding of a rotation (Figure 1 (a)). Let $\mathbb{T} = \{z \in \mathbb{C} \colon |z| = 1\}$. Denote by $\varphi \approx 1.618$ and $\Phi = \varphi - 1$ the golden ratio and its multiplicative inverse, respectively, and write $\gamma = e^{\boldsymbol{i}2\pi/\varphi}$. The long-run ratio of zeros to ones in α_F is equal to $1/\Phi$, and α_F is the coding of $(\gamma^n)_{n\in\mathbb{N}}$ with respect to $\{S_0, S_1\}$, where S_0, S_1 are open interval subsets of $\mathbb{T}$ with lengths $2\pi\Phi$ and $2\pi\Phi^2$, respectively. That is, for all $n \in \mathbb{N}$ and $a \in \{0, 1\}$, $\alpha(n) = a \Leftrightarrow \gamma^n \in S_a$. We will see in Section 6 that α_F is also a Sturmian and a Pisot word.

(c) The Tribonacci word $\alpha_T = 121312112131\cdots$, generated by the substitution $1 \to 12$, $2 \to 13$, $3 \to 1$. Let $\beta \approx 1.839$ be the real root of $x^3 - x^2 - x - 1$ and $\Gamma = (e^{\boldsymbol{i}2\pi/\beta}, e^{\boldsymbol{i}2\pi/\beta^2}) \in \mathbb{T}^2$. The word α_T has a representation as the coding of $(\Gamma^n)_{n\in\mathbb{N}}$ with respect to three open subsets S_1, S_2, S_3 of $\mathbb{T}^2$ with fractal boundaries [26]. For $z \in \mathbb{T}$, let $f(z) = \frac{\mathrm{Log}(z)}{i2\pi} + \frac{1}{2}$. If we identify the multiplicative group $\mathbb{T}^2$ with the additive group $\mathbb{R}^2/\mathbb{Z}^2 = [0,1)^2$ via $(z_1, z_2) \to (f(z_1), f(z_2))$, the images of S_1, S_2, S_3 form the *Rauzy fractal*. See Figure 1 (b).

(d) (Carton and Thomas, [7].) Consider the substitution τ given by $a \to ab$, $b \to ccb$, $c \to c$, and let $x_n = \tau^n(a)$. We have that $x_1 = ab$, $x_2 = abccb$, $x_3 = abccbccccb$, and so on, with the fixed point $\alpha = abc^2bc^4bc^6bc^8 \cdots$ that is not almost-periodic.

(e) (Salomaa, [27].) Consider the morphism $a \to aab, b \to a$. The fixed point $\alpha = aabaabaaabaabaaab \cdots$ is also a Sturmian (see Section 6.1) and hence a toric word [28].

Let τ be a substitution, and order the letters of the alphabet Σ as $a_1, \ldots, a_k$. The matrix M_τ, where $(M_\tau)_{i,j}$ is the number of occurrences of a_j in $\tau(a_i)$, is called the *incidence matrix* of τ. Observe that M_τ^n counts the number of occurrences of each letter in $\tau^n(a_i)$ for $1 \le i \le k$. A substitution is called *primitive* if there exists n such that all entries of M_τ^n are positive.

The factorial word $\alpha_{n!} \in \{0,1\}^\omega$, i.e. the characteristic word of the set $\{n! \mid n \in \mathbb{N}\}$, is not substitutive. This can be shown by observing that every fixed point of a substitution τ can be factorised as $a\tau^0(w)\tau^1(w)\tau^2(w)\cdots$ where $(|\tau^n(w)|)_{n\in\mathbb{N}}$ grows at most exponentially. The blocks of zeros of $\alpha_{n!}$, however, grow super-exponentially. Substitutive words need not be almost-periodic (see example (d) above), but fixed points of primitive substitutions are strongly and effectively almost-periodic [22, Chapter 10.9].

We say that a word $\beta \in \Sigma_2^\omega$ is *morphic* if there exist a substitutive word $\alpha \in \Sigma_1^\omega$ and a renaming of letters $\mu\colon \Sigma_1 \to \Sigma_2$ such that $\beta = \mu(\alpha)$. As an example, if we apply the morphism μ given by $a \to 1$, $b \to 1$, $c \to 0$ to the word $\alpha = abc^2bc^4bc^6 \cdots$ above, the word $\beta = \mu(\alpha)$ we obtain is the characteristic word of the squares predicate: $\beta(n) = 1 \Leftrightarrow n = k^2$ for some $k \in \mathbb{N}$. Carton and Thomas [7] showed that, in fact, for any integer $m \ge 1$ and polynomial $p \in \mathbb{Z}[x]$ satisfying $p(n) \ge 0$ for $n \in \mathbb{N}$, the characteristic word of the set $\{p(n)m^n : n \in \mathbb{N}\}$ is morphic. Morphic words moreover subsume the class of automatic words [22, Chapter 6.3].

4.1. MSO decidability for morphic words

In this section we discuss the semigroup approach used in [7] to show that, for a predicate P whose characteristic word is morphic, the MSO theory of $\langle \mathbb{N}; <, P\rangle$ is decidable. Let $\mathcal{A}$ be a deterministic automaton over an alphabet Σ with the set of states Q. We can associate a semigroup with $\mathcal{A}$ as follows. Two words $u_1, u_2 \in \Sigma^*$ are equivalent with respect to $\mathcal{A}$, written $u_1 \equiv_{\mathcal{A}} u_2$, if for any state q, there exist $R \subseteq Q$ and $t \in Q$ with the following property. For $i \in \{1, 2\}$, when u_i is read in the state q, the run visits exactly the states in R and ends in the state t. Observe that $\Sigma^*/\equiv_{\mathcal{A}}$ consists of finitely many equivalence classes. Denote by $[u]_{\mathcal{A}}$ the equivalence class of $u \in \Sigma^*$, noting that $u \equiv_{\mathcal{A}} v$ implies $uw \equiv_w vw$ and $wu \equiv_{\mathcal{A}} wv$ for all finite words u, v, w. We define the semigroup $G_{\mathcal{A}} = \{[u]_{\mathcal{A}} \colon u \in \Sigma^*\}$ with $[u]_{\mathcal{A}} \cdot [v]_{\mathcal{A}} \coloneqq [uv]_{\mathcal{A}}$. The semigroup $G_{\mathcal{A}}$ associated with $\mathcal{A}$ has been known since the work of Büchi [1].

Carton and Thomas [7] define the class of *profinitely ultimately periodic* words for which the Acceptance Problem is decidable. A word α is profinitely ultimately periodic if it has a factorisation $\alpha = u_0 u_1 u_2 \cdots$ into finite words $(u_n)_{n\in\mathbb{N}}$ such that for every morphism $\sigma\colon \Sigma \to G$ into a finite semigroup G, the sequence $(\sigma(u_n))_{n\in\mathbb{N}}$ is ultimately periodic. This property is *effective* if given σ, we can compute $a, b \in G^*$ such that $\sigma(\alpha) = \sigma(u_0)\sigma(u_1)\cdots = ab^\omega$.

Theorem 4.1. *If $\alpha \in \Sigma^*$ is effectively profinitely ultimately periodic, then the MSO theory of $\langle \mathbb{N}; <, P_\alpha\rangle$ is decidable.*

Proof. Recall that decidability of the MSO theory is equivalent to decidability of the Acceptance Problem for α: given a deterministic automaton $\mathcal{A}$, decide if $\mathcal{A}$ accepts α. Take σ to be the morphism that maps each $u \in \Sigma^*$ to $[u]_{\mathcal{A}}$. By the assumption on α, we can effectively compute $a, b \in (G_{\mathcal{A}})^*$ such that $\sigma(\alpha) = ab^\omega$. It remains to extract from a and b the set S of states that are visited infinitely often when $\mathcal{A}$ reads α, and check S against the acceptance condition of $\mathcal{A}$. $\square$

All morphic words are effectively profinitely ultimately periodic [7]. Hence the MSO theory of $\langle \mathbb{N}; <, P_\alpha\rangle$ for a morphic word α is decidable. Effectively profinitely ultimately periodic words also subsume all words α for which Elgot and Rabin [2] showed decidability of the MSO theory of $\langle \mathbb{N}; <, P_\alpha\rangle$ using their *contraction method*. The factorial word $\alpha_{n!}$, for example, is an effectively profinitely ultimately periodic word that is amenable to the approach of Elgot and Rabin. The factorisation of $\alpha_{n!}$ that yields profinite ultimate periodicity is $u_0 = 0$ and for $n \geq 1$, $u_{2n-1} = 1$ and $u_{2n} = 0^{n!-(n-1)!}$. Rabinovich ([29], see also [30]) showed that, in fact, the MSO theory of $\langle \mathbb{N}; <, P_1, \ldots, P_m\rangle$

is decidable *if and and only if* $\alpha = \alpha_1 \times \ldots \times \alpha_m$, where each α_i is the characteristic word of P_i, is effectively profinitely ultimately periodic. However, if we do not have any *a priori* information on the decidability of the MSO theory, this characterisation does not give us any means to determine whether the word α is effectively profinitely ultimately periodic or not.

Profinitely ultimately periodic words are not known to be closed under products, which makes the approach of [7] inapplicable to the case of multiple predicates. For $k \geq 1$, let α_k be the characteristic word of the predicate P such that $P(m) = 1$ if and only if $m = k^n$ for some $n \in \mathbb{N}$. As mentioned earlier, for every k the word α_k is morphic and hence the MSO theory of $\langle \mathbb{N}; <, P_{\alpha_k} \rangle$ is decidable. However, to the best of our knowledge, the following is open.

Open Problem. *Is the MSO theory of* $\langle \mathbb{N}; <, P_{\alpha_2}, P_{\alpha_3} \rangle$ *decidable?*

Equivalently, is it possible to algorithmically verify all order-theoretic statements (in the language of MSO) about the powers of 2 and 3?

5. Toric words

Recall that we denote by $\mathbb{T}$ the set $\{z \in \mathbb{C} \colon |z| = 1\}$, viewed as an abelian group under multiplication. A word $\alpha \in \Sigma^\omega$ is *toric* if there exist $d > 0$, a collection $\mathcal{S} = \{S_a \colon a \in \Sigma\}$ of pairwise disjoint subsets of $\mathbb{T}^d$, and $\Gamma \in \mathbb{T}^d$ with the following properties. Each S_a has finitely many connected components (in the Euclidean topology), and for all $n \in \mathbb{N}$ and $a \in \Sigma$,

$$\alpha(n) = a \Leftrightarrow \Gamma^n \in S_a.$$

In particular, $\Gamma^n \in \bigcup_{a \in \Sigma} S_a$ for all n. We say that α is generated by $(\Gamma, \mathcal{S})$. In the symbolic dynamics literature, α is referred to as the *coding* of the orbit $(\Gamma^n)_{n \in \mathbb{N}}$ with respect to the collection of sets $\mathcal{S}$. We denote the class of all toric words by $\mathcal{T}$.

The purpose of the topological restriction that each S_a must have finitely many connected components is to avoid the situation where every word is toric with $d = 1$. Below we define further special subclasses of toric words that will help us better classify Sturmian words, certain morphic words, sign patterns of linear recurrence sequences, and so on.

(a) We denote by $\mathcal{T}_O$ the class of toric words that are generated by $(\Gamma, \mathcal{S})$ where each set in $\mathcal{S}$ is open in the Euclidean topology on $\mathbb{T}^d$.

(b) The class $\mathcal{T}_{SA}$ comprises all toric words generated by $(\Gamma, \mathcal{S})$ where each set in $\mathcal{S}$ is an $\mathbb{R}$-semialgebraic subset of $\mathbb{T}^d$.

(c) Finally, we denote by $\mathcal{T}_{SA(\mathbb{Q})}$ the set of all words generated by $(\Gamma, \mathcal{S})$ such that $\Gamma \in (\mathbb{T} \cap \overline{\mathbb{Q}})^d$, i.e. Γ has algebraic entries, and each set in $\mathcal{S}$ is $\mathbb{Q}$-semialgebraic.

Clearly, $\mathcal{T}_{SA} \supseteq \mathcal{T}_{SA(\mathbb{Q})}$. A desirable property that the latter class has is that all operations we will need to perform on $\alpha \in \mathcal{T}_{SA(\mathbb{Q})}$ are effective, although $\mathcal{T}_{SA(\mathbb{Q})}$ is not the only subclass of $\mathcal{T}_{SA}$ with this property.

We have already seen that the Tribonacci word, which is generated by the morphism $1 \to 12, 2 \to 13, 3 \to 1$ and the starting letter 1, belongs to $\mathcal{T}_O$: It is generated by $(\Gamma, \mathcal{S})$ where $\Gamma \in \mathbb{T}^2$ and the sets in $\mathcal{S}$ constitute the Rauzy fractal. We will later show that Sturmian words belong to $\mathcal{T}_{SA}$, and the sign patterns of various linear recurrence sequences belong to $\mathcal{T}_{SA(\mathbb{Q})}$.

5.1. Orbits in $\mathbb{T}^d$

In order to understand toric words, we have to understand the time steps at which the orbit $\mathcal{O}(\Gamma) \coloneqq (\Gamma^n)_{n\in\mathbb{N}}$ of $\Gamma \in \mathbb{T}^d$ visits a given subset of $\mathbb{T}^d$. In this section we will show that unlike the discrete orbit $\mathcal{O}(\Gamma)$, its Euclidean closure $\mathbb{T}_\Gamma \coloneqq \mathrm{Cl}(\mathcal{O}(\Gamma))$ is $\mathbb{Q}$-semialgebraic and effectively computable under some assumptions on Γ. Moreover, $\mathcal{O}(\Gamma)$ visits every open subset of $\mathbb{T}_\Gamma$ infinitely often.

The key to proving these results is the notion of a *multiplicative relation*. We say that $(a_1, \ldots, a_d) \in \mathbb{Z}^d$ is a multiplicative relation of $z = (z_1, \ldots, z_d)$, $z \in (\mathbb{C}^\times)^d$ if $z_1^{a_1} \cdots z_d^{a_d} = 1$. For such z,

$$G(z) \coloneqq \{(a_1, \ldots, a_d) \in \mathbb{Z}^d \mid z_1^{a_1} \cdots z_d^{a_d} = 1\}$$

is called the *group of multiplicative relations* of z. For all z, $G(z)$ is a free abelian group under addition with a basis containing at most d vectors from $\mathbb{Z}^d$. If the entries of z are algebraic, then such a basis can be effectively computed: By a theorem of Masser [31], $G(z)$ has a basis $v_1, \ldots, v_m$ of vectors satisfying $\|v_i\|_2 < B$ for all i, where B is a bound that can be effectively computed from z. It remains to find by enumeration a maximally linearly independent set of vectors of the form $a = (a_1, \ldots, a_d) \in \mathbb{Z}^d$ with the property that $z_1^{a_1} \cdots z_d^{a_d} = 1$ and $\|a\|_2 < B$.

To describe $\mathbb{T}_\Gamma$ we will employ Kronecker's theorem in simultaneous Diophantine approximation. For $x, y \in \mathbb{R}$, denote by $[\![x]\!]_y$ the distance from x

to a nearest integer multiple of y. Further write $[\![x]\!]$ for $[\![x]\!]_1$. The following is a classical version of Kronecker's theorem [32].

Theorem 5.1. *Let $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and $y = (y_1, \ldots, y_d) \in \mathbb{R}^d$ be such that for all $b \in \mathbb{Z}^d$,*

$$b \cdot x \in \mathbb{Z} \Rightarrow b \cdot y \in \mathbb{Z}.$$

For every $\epsilon > 0$ there exist infinitely many values $n \in \mathbb{N}$ satisfying

$$\sum_{j=1}^{d} [\![nx_j - y_j]\!] < \epsilon.$$

Writing $X = (e^{\boldsymbol{i}2\pi x_1}, \ldots, e^{\boldsymbol{i}2\pi x_d})$ and $Y = (e^{\boldsymbol{i}2\pi y_1}, \ldots, e^{\boldsymbol{i}2\pi y_d})$, the condition that for all $b \in \mathbb{Z}^d$, $b \cdot x \in \mathbb{Z} \Rightarrow b \cdot y \in \mathbb{Z}$ is equivalent to $G(X) \subseteq G(Y)$. That is, "every multiplicative relation of X is also a multiplicative relation of Y". We can now prove the main result of this section.

Lemma 5.2. *Let $\Gamma = (\gamma_1, \ldots, \gamma_d) \in \mathbb{T}^d$.*

(a) If $z \in \mathbb{T}^d$ is such that $G(\Gamma) \subseteq G(z)$, then for any open $O \subset \mathbb{T}_\Gamma$ containing z there exist infinitely many values $n \in \mathbb{N}$ such that $\Gamma^n \in O$.

(b) $\mathbb{T}_\Gamma$ is equal to $\{z \in \mathbb{T}^d \colon G(\Gamma) \subseteq G(z)\}$, $\mathbb{Q}$-semialgebraic, and effectively computable given a basis of $G(\Gamma)$.

Proof. Consider $z = (z_1, \ldots, z_d) \in \mathbb{T}^d$ with $G(\Gamma) \subseteq G(z)$. Define $x_j = \frac{\mathrm{Log}(\gamma_j)}{\boldsymbol{i}2\pi}$ and $y_j = \frac{\mathrm{Log}(z_j)}{\boldsymbol{i}2\pi}$ for $1 \leq j \leq d$. We have $x_j, y_j \in (-1/2, 1/2]$. For all $n \in \mathbb{N}$,

$$\begin{aligned}
\|\Gamma^n - z\|_1 &= \sum_{j=1}^{d} |\gamma_j^n - z_j| \\
&\leq \sum_{j=1}^{d} |\,\mathrm{Log}(\gamma_j^n / z_j)| \\
&= \sum_{j=1}^{d} [\![n\,\mathrm{Log}(\gamma_j)/\boldsymbol{i} - \mathrm{Log}(z_j)/\boldsymbol{i}]\!]_{2\pi} \\
&= 2\pi \sum_{j=1}^{d} [\![nx_j - y_j]\!]
\end{aligned}$$

where the last equality follows from the fact that $[\![x]\!]_{2\pi} = 2\pi[\![x/(2\pi)]\!]$ for all $x \in \mathbb{R}$. Applying Kronecker's theorem, for each $\epsilon > 0$ there exist infinitely many values n such that $\|\Gamma^n - z\|_1 < \epsilon$. This proves (a).

To prove (b), let $V = \{v_1, \ldots, v_m\}$ be a basis of $G(\Gamma)$, where for all $1 \le k \le m$, $v_k = (v_{k,1}, \ldots, v_{k,d})$. Since for $z = (z_1, \ldots, z_d)$,

$$G(\Gamma) \subseteq G(z) \quad \Leftrightarrow \quad \bigwedge_{k=1}^{m} z_1^{v_{k,1}} \cdots z_d^{v_{k,d}} = 1,$$

the set $\{z \in \mathbb{T}^d \colon G(\Gamma) \subseteq G(z)\}$ is closed and $\mathbb{Q}$-semialgebraic. It moreover contains the orbit $\mathcal{O}(\Gamma)$ as $G(\Gamma) \subseteq G(\Gamma^n)$ for all $n \in \mathbb{N}$. Invoking (a), the orbit $\mathcal{O}(\Gamma)$ is dense in $\{z \in \mathbb{T}^d \colon G(\Gamma) \subseteq G(z)\}$. Hence the latter must be exactly the closure of $\mathcal{O}(\Gamma)$. □

5.2. Closure properties of toric words

We now investigate closure properties of toric words under various word operations. First we will show that unlike the class of almost-periodic words, all classes of toric words that we have defined are closed under products.

Theorem 5.3. *Let $\alpha_0, \ldots, \alpha_{L-1} \in \mathcal{K}$, where $\mathcal{K}$ is one of $\mathcal{T}, \mathcal{T}_O, \mathcal{T}_{SA}, \mathcal{T}_{SA(\mathbb{Q})}$. The product word $\alpha = \alpha_0 \times \ldots \times \alpha_{L-1}$ also belongs to $\mathcal{K}$.*

Proof. Suppose each $\alpha_i \in \Sigma_i^\omega$ and is generated by $(\Gamma_i, \{S_a^{(i)} \colon a \in \Sigma_i\})$, where $\Gamma_i \in \mathbb{T}^{d_i}$. Let Σ be the product alphabet $\Sigma_0 \times \ldots \times \Sigma_{L-1}$, noting that $\alpha \in \Sigma^\omega$. Further let $d = d_0 + \ldots + d_{L-1}$ and $\Gamma = \prod_{r=0}^{L-1} \Gamma_i \in \mathbb{T}^d$. For each letter $b = (a_0, \ldots, a_{L-1}) \in \Sigma$, define $S_b = \prod_{r=0}^{L-1} S_{a_r}^{(r)}$. The word α is toric and generated by $(\Gamma, \{S_b \colon b \in \Sigma\})$. It remains to observe that if every $S_a^{(i)}$ is open, or $\mathbb{K}$-semialgebraic for $\mathbb{K} = \mathbb{Q}$ or $\mathbb{K} = \mathbb{R}$, then the same applies to S_b for every $b \in \Sigma$. □

The classes of toric words we consider are also closed under applications of uniform morphisms.

Theorem 5.4. *Let $\alpha \in \mathcal{K}$, where $\mathcal{K}$ is one of $\mathcal{T}, \mathcal{T}_O, \mathcal{T}_{SA}, \mathcal{T}_{SA(\mathbb{Q})}$. Suppose $\alpha \in \Sigma_1^\omega$, and let $\tau \colon \Sigma_1 \to \Sigma_2$ be a k-uniform morphism, i.e. $|\tau(a)| = k$ for all $a \in \Sigma_1$. The word $\beta \coloneqq \tau(\alpha)$ also belongs to $\mathcal{K}$.*

Proof. Suppose α is generated by $(\Gamma, \{S_a \colon a \in \Sigma_1\})$. The idea is to "slow down Γ by a factor of k" and "add a counter modulo k". Let $\Gamma = (\gamma_1, \ldots, \gamma_d)$,

$\lambda_j = e^{\boldsymbol{i}\operatorname{Log}(\gamma_j)/k}$ for $1 \le j \le d$, and $\Lambda = (\lambda_1, \ldots, \lambda_d)$. Observe that $\gamma_j = \lambda_j^k$ for all j. Further let $\omega = e^{\boldsymbol{i}2\pi/k}$ and $B_j = \{z \in \mathbb{C} \colon |z - \omega^j| < 1/k\}$. The sets $B_0, \ldots, B_{k-1}$ are open, $\mathbb{Q}$-semialgebraic and pairwise disjoint. Moreover, $\omega^n \in B_j$ if and only if $n \equiv j (\operatorname{mod} k)$.

For a letter $b \in \Sigma_2$, define

$$S_b = \bigcup_{\substack{a \in \Sigma_1,\ 0 \le j < |\tau(a)| \\ \tau(a)(j) = b}} \Gamma^j S_a \times B_j.$$

We will show that for all $n \in \mathbb{N}$ and $b \in \Sigma_2$, $\beta(n) = b$ if and only if $\Lambda_1^n \in S_b$, where $\Lambda_1 = (\lambda_1, \ldots, \lambda_d, \omega)$. Fix $n = qk + r$ where $0 \le r < k$. By construction,

$$\Lambda_1^n \in S_b \Leftrightarrow \exists a, j \colon \tau(a)(j) = b,\ \Lambda^n \in \Gamma^j S_a,\ \text{and } \omega^n \in B_j.$$

Recall that $\omega^n \in B_j$ is equivalent to $j = r$. Hence

$$\Lambda^n \in \Gamma^j S_a \ \Leftrightarrow\ \Gamma^{-j}\Lambda^n \in S_a \ \Leftrightarrow\ \Gamma^q \in S_a \ \Leftrightarrow\ \alpha(q) = a.$$

Above we used the fact that $\Gamma = \Lambda^k$. We have thus shown that $\Lambda_1^n \in S_b$ if and only if $\alpha(q) = a$ for some $a \in \Sigma_1$ satisfying $\tau(a)(r) = b$. Since $\beta(n) = \tau(\alpha(q))(r)$, it follows that $\Gamma^n \in S_b$ if and only if $\beta(n) = b$. That is, β is the toric word generated by $(\Lambda_1, \{S_b \colon b \in \Sigma_2\})$. $\square$

Corollary 5.5. *The merge α of $\alpha_0, \ldots, \alpha_{L-1} \in \mathcal{K}$, where $\mathcal{K}$ is one of the classes of toric words as above, also belongs to $\mathcal{K}$.*

Proof. Suppose $\alpha_i \in \Sigma_i$. Denote by τ the L-uniform morphism that maps each $(a_0, \ldots, a_{L-1}) \in \Sigma_0 \times \ldots \times \Sigma_{L-1}$ to the concatenation of $a_0, \ldots, a_{L-1}$. Observe that $\alpha = \tau(\alpha_0 \times \ldots \times \alpha_{L-1})$. $\square$

Finally, we show that our classes of toric words are closed under taking suffixes. This property is shared with the classes of almost-periodic words.

Theorem 5.6. *All four classes of toric words are closed under taking suffixes.*

Proof. If α is generated by $(\Gamma, \{S_a \colon a \in \Sigma\})$, then $\alpha[N, \infty)$ is generated by $(\Gamma, \{\Gamma^{-N} S_a \colon a \in \Sigma\})$. $\square$

5.3. Almost-periodicity of toric words

We will now show that toric words belonging to the classes $\mathcal{T}_O$ and $\mathcal{T}_{SA}$ are almost-periodic, albeit for somewhat different reasons. The proof for the former class is topological, whereas the proof for $\mathcal{T}_{SA}$ relies on the Skolem-Mahler-Lech theorem for linear recurrence sequences. Combined with closure under products, almost periodicity of toric words will allow us to apply Semënov's theorem to the problem of deciding the MSO theory of $\langle \mathbb{N}; <, P_1, \ldots, P_m \rangle$, where each P_i is a predicate associated with a toric word.

Theorem 5.7. *Every $\alpha \in \mathcal{T}_O$ is strongly almost-periodic.*

Proof. Consider $\alpha \in \mathcal{T}_O$ that is generated by $(\Gamma, \{S_a : a \in \Sigma\})$ where $\Gamma = (\gamma_1, \ldots, \gamma_d) \in \mathbb{T}^d$ and each S_a is an open subset of $\mathbb{T}^d$. Let $\mathbb{T}_\Gamma$ denote the closure of $(\Gamma^n)_{n \in \mathbb{N}}$, and consider a finite word $w = w(0) \ldots w(l-1) \in \Sigma^l$. The latter occurs at the position n in α if and only if

$$\bigwedge_{i=0}^{l-1} \Gamma^{n+i} \in S_{w(i)}$$

which is equivalent to $\Gamma^n \in S_w$ where

$$S_w \coloneqq \mathbb{T}_\Gamma \cap \bigcap_{i=0}^{l-1} \Gamma^{-i} S_{w(i)}.$$

Since each $S_{w(i)} \subseteq \mathbb{T}^d$ is open, S_w is an open subset of $\mathbb{T}_\Gamma$. If S_w is empty, then w does not occur in α. Suppose therefore S_w is not empty.

For $k \in \mathbb{N}$, let $X_k = \{z \in \mathbb{T}_\Gamma : \Gamma^k z \in S_w\}$. Each X_k is an open subset of $\mathbb{T}_\Gamma$, and since $(\Gamma^n)_{n \in \mathbb{N}}$ visits every open subset of $\mathbb{T}_\Gamma$ infinitely often, $\{X_k : k \in \mathbb{N}\}$ is an open cover of $\mathbb{T}_\Gamma$. By compactness of $\mathbb{T}_\Gamma$, there exists $K \in \mathbb{N}$ such that $\bigcup_{k=0}^{K} X_k$ covers $\mathbb{T}_\Gamma$. That is, the orbit of any point in $\mathbb{T}_\Gamma$ under the action of $z \to \Gamma z$ visits S_w in at most K steps. Hence for every $n \in \mathbb{N}$ there exists $0 \le k \le K$ such that $\Gamma^{n+k} \in S_w$. Therefore, the word w is guaranteed to occur in $\alpha[n, n+K+l)$ for every n. $\square$

Corollary 5.8. *If $\alpha \in \mathcal{T}_O \cap \mathcal{T}_{SA(\mathbb{Q})}$, then α is strongly and effectively almost-periodic.*

Proof. Suppose α is generated by $(\Gamma, \{S_a : a \in \Sigma\})$ where $\Gamma \in (\mathbb{T} \cap \overline{\mathbb{Q}})^d$ and each S_a is open and $\mathbb{Q}$-semialgebraic. As discussed in Section 5.1, we can compute the $\mathbb{Q}$-semialgebraic set $\mathbb{T}_\Gamma$ effectively. Hence, given w, we can effectively compute a representation of X_k (see the proof of Theorem 5.7) as a $\mathbb{Q}$-semialgebraic set using tools of semialgebraic geometry. We can then determine K by checking for increasing values of m, starting with $m = 0$, whether $\bigcup_{k=0}^{m} X_k$ covers $\mathbb{T}_\Gamma$. Hence given w, we can effectively compute $K+l$ as a bound on consecutive occurrences of w in α. $\square$

We now move onto the classes $\mathcal{T}_{SA}$ and $\mathcal{T}_{SA(\mathbb{Q})}$.

Theorem 5.9. *Let $\alpha \in \mathcal{K}$, where $\mathcal{K}$ is either $\mathcal{T}_{SA}$ or $\mathcal{T}_{SA(\mathbb{Q})}$.*

(a) There exists a suffix $\beta := \alpha[N, \infty)$ of α such that $\beta \in \mathcal{T}_O \cap \mathcal{K}$.

(b) The word α is almost-periodic.

Proof. Suppose α is generated by $((\gamma_1, \ldots, \gamma_d), \{S_a : a \in \Sigma\})$ where each S_a is a $\mathbb{K}$-semialgebraic subset of $\mathbb{T}^d$; if $\mathcal{K} = \mathcal{T}_{SA}$, then $\mathbb{K} = \mathbb{R}$, and $\mathbb{K} = \mathbb{Q}$ otherwise. Recall the definition of a semialgebraic subset of $\mathbb{C}^d$. For each letter a and $z = (z_1, \ldots, z_d) \in \mathbb{T}^d$ it holds that $z \in S_a$ if and only if

$$\bigvee_{i \in I_a} \bigwedge_{j \in J_a} p_{i,j}(\mathrm{Re}(z_1), \mathrm{Im}(z_1), \ldots, \mathrm{Re}(z_d), \mathrm{Im}(z_d)) \;\Delta_{i,j}\; 0$$

where each $p_{i,j}$ is a polynomial with real coefficients and $\Delta_{i,j} \in \{\geq, >\}$. Define

$$u_n^{a,i,j} = p_{i,j}(\mathrm{Re}(\gamma_1^n), \mathrm{Im}(\gamma_1^n), \ldots, \mathrm{Re}(\gamma_d^n), \mathrm{Im}(\gamma_d^n)).$$

Observe that each $(u_n^{a,i,j})_{n\in\mathbb{N}}$ is a linear recurrence sequence over $\mathbb{R}$. Applying the Skolem-Mahler-Lech theorem, for each a, i, j there exist $\nu := N_{a,i,j}$ and $\lambda := L_{a,i,j}$ such that for each $0 \leq r < \lambda$, the subsequence $(u_{n\lambda+\nu}^{a,i,j})_{n\in\mathbb{N}}$ is either identically zero or does not have any zero terms. Take $N = \max_{a,i,j} N_{a,i,j}$ and $L = \prod_{a,i,j} L_{a,i,j}$. It holds that for every a, i, j and $0 \leq r < L$, the subsequence $(v_n^{a,i,j,r})_{n\in\mathbb{N}}$ of $(u_n^{a,i,j})_{n\in\mathbb{N}}$ given by

$$v_n^{a,i,j,r} = u_{N+nL+r}^{a,i,j}$$

is either identically zero or is never zero.

Consider $\beta = \alpha[N, \infty)$ and for $0 \leq r < L$, define β_r by

$$\beta_r(n) := \beta(nL + r) = \alpha(N + nL + r)$$

for all $n \in \mathbb{N}$. We will show that $\beta_r \in \mathcal{T}_O \cap \mathcal{K}$ for all r. Thereafter, from Theorem 5.4 it follows that $\beta \in \mathcal{T}_O \cap \mathcal{K}$, proving (a). Invoking Theorem 5.7, β is strongly almost-periodic. Since β is a suffix of α, we conclude that α is almost-periodic.

Fix $0 \le r < L$. For every $a \in \Sigma$ and $n \in \mathbb{N}$ we have that $\beta_r(n) = a$ if and only if

$$\bigvee_{i\in I_a} \bigwedge_{j\in J_a} v_n^{a,i,j,r} \ \Delta_{i,j} \ 0$$

where $\Delta_{i,j} \in \{\ge, >\}$. By construction of N, L, for each $i \in I_a$ and $j \in J_a$, the (i,j)th inequality above either holds for all n (in case $v_n^{a,i,j,r}$ is identically zero and $\Delta_{i,j}$ is equality), or holds if and only if $v_n^{a,i,j,r} > 0$. Hence there exist $K_a \subseteq I_a$ and $M_a \subseteq J_a$ such that for all n, $\beta_r(n) = a$ if and only if

$$\bigvee_{i\in K_a} \bigwedge_{j\in M_a} v_n^{a,i,j,r} > 0.$$

Let $\lambda_k = \gamma_k^L$ for $1 \le k \le d$, and observe that we can write $v_n^{a,i,j,r} > 0$ as

$$q_{a,i,j,r}(\mathrm{Re}(\lambda_1^n), \mathrm{Im}(\lambda_1^n), \ldots, \mathrm{Re}(\lambda_d^n), \mathrm{Im}(\lambda_d^n)) > 0$$

for a polynomial $q_{a,i,j}$ with real coefficients. For each a, define $S_a^{(r)} \subseteq \mathbb{T}^d$ by

$$(z_1, \ldots, z_d) \in S_a^{(r)} \Leftrightarrow \bigvee_{i\in K_a} \bigwedge_{j\in M_a} q_{a,i,j}(\mathrm{Re}(z_1), \mathrm{Im}(z_1), \ldots, \mathrm{Re}(z_d), \mathrm{Im}(z_d)).$$

We have that β_r is the toric word generated by $((\lambda_1, \ldots, \lambda_d), \{S_a^{(r)} \colon a \in \Sigma\})$. Since each $S_a^{(r)}$ is open, $\beta_r \in \mathcal{T}_O$. As discussed above, it follows that β is strongly almost-periodic and α is almost-periodic. □

Corollary 5.10. *Assuming decidability of the Skolem Problem for LRS over $\mathbb{R} \cap \overline{\mathbb{Q}}$, every $\alpha \in \mathcal{T}_{SA(\mathbb{Q})}$ is effectively almost-periodic.*

Proof. Suppose α is generated by $(\Gamma, \{S_a \colon a \in \Sigma\})$, where $\Gamma \in (\mathbb{T} \cap \overline{\mathbb{Q}})^d$ and each S_a is $\mathbb{Q}$-semialgebraic. In this case, in the proof of Theorem 5.9 each $(u_n^{a,i,j})_{n\in\mathbb{N}}$ is an LRS over $\mathbb{R} \cap \overline{\mathbb{Q}}$. If we assume decidability of the Skolem Problem for LRS over $\mathbb{R} \cap \overline{\mathbb{Q}}$, then using the Skolem-Mahler-Lech theorem (see Section 2) we can effectively compute the values of $N_{a,i,j}, L_{a,i,j}$ and hence N, L in the proof above. We can therefore effectively compute $(\Gamma_1, \mathcal{S}_1)$ that generates the toric word $\beta = \alpha[N, \infty)$, where $\Gamma_1 \in (\mathbb{T} \cap \overline{\mathbb{Q}})^d$ and each set in $\mathcal{S}_1$ is open and $\mathbb{Q}$-semialgebraic. Invoking Corollary 5.8, β is strongly and effectively almost-periodic. Hence α is effectively almost-periodic. □

Theorem 5.7 tells us that words belonging to the class $\mathcal{T}_{SA}$ are, in a sense, not too different from words in the class $\mathcal{T}_O$. In fact, we can combine words across the two classes by taking a product, while maintaining almost periodicity.

Theorem 5.11. *Let $\alpha_0, \ldots, \alpha_{L-1} \in \mathcal{T}_O$ and $\beta_0, \ldots, \beta_{M-1} \in \mathcal{T}_{SA}$. The word $\delta \coloneqq \prod_{i=0}^{L-1} \alpha_i \times \prod_{j=0}^{M-1} \beta_j$ is almost-periodic.*

Proof. Let $\alpha \coloneqq \prod_{i=0}^{L-1} \alpha_i$ and $\beta \coloneqq \prod_{j=0}^{M-1} \beta_j$. The word δ, up to a renaming of letters, is equal to $\alpha \times \beta$. By Theorem 5.4, $\alpha \in \mathcal{T}_O$ and $\beta \in \mathcal{T}_{SA}$. By Corollary 5.8, there exists N such that $\beta[N, \infty) \in \mathcal{T}_O$. By closure under taking suffixes (Theorem 5.6), $\alpha[N, \infty) \in \mathcal{T}_O$. Applying Theorem 5.4, $\delta[N, \infty) = \alpha[N, \infty) \times \beta[N, \infty)$ belongs to $\mathcal{T}_O$ and hence is strongly almost-periodic. It follows that δ is almost-periodic. $\square$

We have thus uncovered a myriad of structures with potentially decidable MSO theories: Suppose $P_1, \ldots P_m$ are predicates with characteristic words $\alpha_1, \ldots, \alpha_m$ that belong to $\mathcal{T}_O \cup \mathcal{T}_{SA}$. Then the word $\alpha \coloneqq \alpha_1 \times \ldots \times \alpha_m$ is almost-periodic by Theorem 5.11. Recall that by Semënov's theorem, a sufficient condition for decidability of the MSO theory of $\langle \mathbb{N}; <, P_1, \ldots, P_m \rangle$ is *effective* almost periodicity of α. Hence the questions arises: For which toric predicates $P_1, \ldots, P_m$ is it possible to prove effective almost periodicity of the product word? As a concrete example, let α_1 be the sign pattern of LRS $u_n = \sin(n\theta)$, where $e^{i\theta} \in \overline{\mathbb{Q}} \cap \mathbb{T}$ is not a root of unity, and α_2 be the Tribonacci word generated by the morphism $1 \to 12, 2 \to 23, 3 \to 1$ and the starting letter 1. As discussed earlier, $\alpha_1 \in \mathcal{T}_{SA(\mathbb{Q})}$ and $\alpha_2 \in \mathcal{T}_O$, and both words are effectively almost-periodic. Hence we can separately decide the MSO theories of $\langle \mathbb{N}; <, P_1 \rangle$ and $\langle \mathbb{N}; <, P_2 \rangle$, but at the moment do not have a solution to the following.

Open Problem. *Is the word $\alpha \coloneqq \alpha_1 \times \alpha_2$ effectively almost-periodic? Is the MSO theory of $\langle \mathbb{N}; <, P_1, P_2 \rangle$ decidable?*

Note that it is possible that the latter theory be decidable while at the same time α not being effectively almost-periodic. A similar open problem is decidability of the MSO theory of $\langle \mathbb{N}; < \rangle$ extended with a morphic predicate P_3 and the predicate P_2 above. In this case once again we can separately decide the MSO theories of $\langle \mathbb{N}; <, P_2 \rangle$ and $\langle \mathbb{N}; <, P_3 \rangle$ by [7] and Semënov's theorem, respectively.

We conclude by isolating a class of toric words which we can combine while maintaining effective almost periodicity of the product word and decidability of the resulting MSO theory. It turns out that this family of toric words is powerful enough for proving decidability of various subclasses of the Model-Checking Problem for linear dynamical systems, discussed in Section 6.4.

Theorem 5.12. *Let $\alpha_1, \ldots, \alpha_m \in \mathcal{T}_O \cap \mathcal{T}_{SA(\mathbb{Q})}$.*

(a) The product $\alpha = \alpha_1 \times \ldots \times \alpha_m$ is effectively almost-periodic.

(b) The MSO theory of the structure $\langle \mathbb{N}; <, P_{\alpha_1}, \ldots, P_{\alpha_m} \rangle$ is decidable.

Proof. Apply Theorem 5.4 and corollary 5.8 to prove (a). To prove (b), recall that by Büchi's construction, the decision problem for the MSO theory of the structure above reduces to the Acceptance Problem for α. The latter is decidable by Corollary 5.8 and Semënov's theorem. □

We can do better if we assume existence of a Skolem oracle.

Theorem 5.13. *Let $\alpha_1, \ldots, \alpha_m \in \mathcal{T}_{SA(\mathbb{Q})}$. Assuming decidability of the Skolem Problem for LRS over $\mathbb{R} \cap \overline{\mathbb{Q}}$, the MSO theory of $\langle \mathbb{N}; <, P_{\alpha_1}, \ldots, P_{\alpha_m} \rangle$ is decidable.*

Proof. Let $\alpha = \alpha_1 \times \ldots \times \alpha_m$. By theorem 5.4, $\alpha \in \mathcal{T}_{SA}$, and by Corollary 5.10, α is effectively almost-periodic under the assumption that the Skolem Problem is decidable for real algebraic LRS. It remains to invoke Semënov's theorem. □

6. Applications

In this section we discuss MSO decidability and almost periodicity properties of Sturmian words, Pisot words, sign patterns of linear recurrence sequences, and words arising from linear dynamical systems.

6.1. Sturmian words

An infinite word over the alphabet $\Sigma = \{0, 1\}$ is Sturmian if the number of its distinct factors of length n is equal to $n + 1$ for all $n \in \mathbb{N}$. We refer the reader to [22, Chapter 10.5] for a detailed discussion of Sturmian words. It is known that if a word has at most n distinct factors of length n for some $n > 0$, then it is eventually periodic. Hence Sturmian words have the smallest factor complexity among words that are not ultimately periodic.

Sturmian words have many equivalent characterisations, including one as a family of toric word. For $z \in \mathbb{T}$ and $x \in \mathbb{R}_{>0}$, denote by $\mathcal{I}(z,x)$ the open interval subset of the unit circle $\mathbb{T}$ generated by starting at z and rotating counter-clockwise until $ze^{\boldsymbol{i}x}$ is reached. Further define $\mathcal{I}[z,x) \coloneqq \{z\} \cup \mathcal{I}(z,x)$ and $\mathcal{I}(z,x] \coloneqq \mathcal{I}(z,x) \cup \{ze^{\boldsymbol{i}x}\}$. A word α is Sturmian if and only if there exist $\gamma \in \mathbb{T}$ not a root of unity and $\xi \in \mathbb{T}$ such that for all n, $\alpha(n) = 1$ if and only if $\gamma^n \in \mathcal{I}[\xi, \theta)$, where $\theta = |\operatorname{Log}(\gamma)|$. That is, a Sturmian word is the coding of $(\gamma^n)_{n \in \mathbb{N}}$ for some γ that is not a root of unity with respect to a partition $\{S_0, S_1\}$ of $\mathbb{T}$ where S_1 is a semi-open interval of length exactly θ.[2] Hence all Sturmian words belong to $\mathcal{T}_{SA}$, and are almost-periodic by Theorem 5.9. In fact, they are strongly almost-periodic [22].

Carton and Thomas [7] asked: Is the MSO theory of $\langle \mathbb{N}; <, P_\alpha \rangle$, where α is a Sturmian word, decidable? Call the Sturmian word with parameters γ and ξ *effective* if there exists an algorithm for approximating $\operatorname{Log}(\xi)$ and $\theta \coloneqq |\operatorname{Log}(\gamma)|$ to arbitrary precision. We will show that such α is effectively almost-periodic and hence the MSO theory of $\langle \mathbb{N}; <, P_\alpha \rangle$ is decidable. Note that by the assumption that γ is not a root of unity, the equation $\gamma^n = \xi$ can have at most one solution in n. Moreover, $\gamma^n = \xi e^{\boldsymbol{i}\theta}$ if and only if $\gamma^{n+1} = \xi$ or $\gamma^{n-1} = \xi$. Hence for every effective Sturmian word α there exists an algorithm that computes $\alpha(n)$ given n. The algorithm simply stores the value N (if any) such that $\gamma^N = \xi$, as well as the values of $\alpha(N-1), \alpha(N), \alpha(N+1)$.[3] On $n \notin \{N-1, N, N+1\}$, it determines $\alpha(n)$ by approximating $\operatorname{Log}(\gamma^n)$ to sufficient precision and comparing it to approximations of $\operatorname{Log}(\xi)$ and $\operatorname{Log}(\xi e^{\boldsymbol{i}\theta})$.

Theorem 6.1. *An effective Sturmian word α is effectively almost-periodic.*

Proof. Suppose α is generated by γ and ξ. Define θ, S_0 and S_1 as above. As mentioned earlier, all Sturmian words are strongly almost-periodic. Moreover, under the assumption on α, there exists a program that computes $\alpha(n)$ given n. Hence we have to show existence of a program that, given a finite word u, determines whether u occurs in α, and in case it does, computes an upper bound on the gaps between consecutive occurrences. If $\gamma^N = \xi$

[2] Note that $\theta = |\operatorname{Log}(\overline{\gamma})|$ and $\gamma^n \in \mathcal{I}(\xi, \theta]$ if and only if $\overline{\gamma}^n \in \mathcal{I}[\overline{\xi\gamma}, \theta)$. Hence it suffices to only consider closed-open intervals when defining Sturmian words.

[3] Here we only show existence of the desired algorithm. If we want to write such an algorithm down, we have to first determine, if any, the value of N. Techniques for accomplishing this depend on the values of ξ, γ and how they are presented.

for some N, then let $M = N + 2$. Otherwise, let $M = 0$. For $n \geq M$, $\gamma^n \neq \xi, \xi e^{\boldsymbol{i}\theta}$. That is, γ^n does not hit the endpoints of S_0, S_1. It suffices to prove effective almost periodicity of $\beta \coloneqq \alpha[M, \infty)$. As in the proof of Theorem 5.7, a word $w = w(0) \cdots w(l-1)$ occurs at a position $n \geq M$ in α if and only if $\gamma^n \in S_w$, where

$$S_w = \bigcap_{i=0}^{l-1} \gamma^{-i} S_{w(i)}$$

and each $S_{w(i)}$ is the open interval $\mathcal{I}(\xi, \xi e^{\boldsymbol{i}\theta})$ if $w(i) = 1$ and $S_{w(i)} = \mathcal{I}(\xi e^{\boldsymbol{i}\theta}, \xi)$ otherwise. Since γ is not a root of unity, no two distinct intervals $\gamma^{-i} S_{w(i)}$ and $\gamma^{-j} S_{w(j)}$ share an endpoint. Hence by approximating $\operatorname{Log}(z)$ to sufficient precision for every endpoint z of $\gamma^{-i} S_{w(i)}$ for $0 \leq i < l$, we can decide whether S_w is empty. If $S(w) = \emptyset$, then w does not occur in β. If $S_w \neq \emptyset$, then we can compute, using the approximate positions of the endpoints, an open semialgebraic interval subset J of $\mathbb{T}$ that is contained in S_w. Similarly to the proofs of Theorem 5.7 and corollary 5.8, let K be such that $\bigcup_{i=0}^{K} \gamma^{-i} J$ covers $\mathbb{T}$; Such K can be computed using a trial-and-error method and tools of semialgebraic geometry. Thus for every $m \in \mathbb{N}$ there exists $n \in [m, m+K]$ such that $\gamma^n \in J$, which implies $\gamma^n \in S_w$. It follows that for every $m \in \mathbb{N}$ the word w occurs in $\beta[m, m + K + l)$. □

What about decidability of the MSO theory of $\langle \mathbb{N}; <, P_{\alpha_1}, \ldots, P_{\alpha_m} \rangle$, where each α_i is Sturmian? Suppose each α_i is the effective Sturmian word with parameters γ_i, ξ_i and $\theta_i = |\operatorname{Log}(\gamma_i)|$. Further suppose that $\gamma_1, \ldots, \gamma_m$ are multiplicatively independent. Importantly, under this assumption, $\mathbb{T}_\Gamma = \mathbb{T}^d$ for $\Gamma = (\gamma_1, \ldots, \gamma_d)$.

Theorem 6.2. *Under the assumptions above, $\alpha \coloneqq \alpha_1 \times \ldots \times \alpha_m$ is effectively almost-periodic and hence the MSO theory of $\langle \mathbb{N}; <, P_{\alpha_1}, \ldots, P_{\alpha_m} \rangle$ is decidable.*

Proof sketch. Let $\Sigma = \{0, 1\}^m$ and M be such that for all $n \geq M$ and $1 \leq j \leq m$, $\gamma_j^n \neq \xi_j$ and $\gamma_j^n \neq \xi_j e^{\boldsymbol{i}\theta_j}$. For each $a \in \Sigma$, there exists $S_a \subset \mathbb{T}^m$ that is a product of open interval subsets of $\mathbb{T}$ (henceforth called a *box*) such that for all $n \in \mathbb{N}$, $\alpha(n) = a$ if and only if $\Gamma^n \in S_a$. Let $w \in \Sigma^l$. For $n \geq M$, the word w occurs at the position n in α if and only if $\Gamma^n \in S_w$, where $S_w = \bigcap_{i=0}^{l-1} \Gamma^{-i} S_{w(i)}$ and each $S_{w(i)}$ is of the form $\prod_{j=1}^{m} T_j^{(i)}$ for open intervals

$T_1^{(i)}, \dots, T_m^{(i)} \subset \mathbb{T}$. Therefore,

$$S_w = \prod_{j=1}^{m} \bigcap_{i=0}^{l-1} \gamma_j^{-i} T_j^{(i)}$$

itself is an open box. As argued in the proof of Theorem 6.1, using the oracles for approximating $\mathrm{Log}(\gamma_i), \mathrm{Log}(\xi_i)$ to sufficient precision we can decide whether each $\bigcap_{i=0}^{l-1} \gamma_j^{-i} T_j^{(i)}$ is empty. In case S_w is non-empty, we compute an open semialgebraic box J such that $J \subset S_w$. It remains to bound the return time of $(\Gamma^n)_{n\in\mathbb{N}}$ in J by computing K such that $\bigcup_{i=0}^{K} \Gamma^{-i} J$ covers $\mathbb{T}_\Gamma$, which is the whole of $\mathbb{T}^d$ by the multiplicative independence assumption.

Since J is $\mathbb{Q}$-semialgebraic, such K can be computed effectively by trial-and-error. In the end it holds that for every $m \geq M$, the word w occurs in $\alpha[m, m+K+l)$. □

6.2. Pisot words

We now discuss a class of morphic words called *Pisot words* and the related Pisot conjecture. The conjecture identifies a class of morphic words that are expected to have, in a specific sense, a toric representation.

A *Pisot–Vijayaraghavan number*, also called a *Pisot number*, is a real algebraic integer greater than 1 whose Galois conjugates all have absolute value less than 1. A *Pisot* substitution $\tau\colon \Sigma^* \to \Sigma^*$ has the property that the incidence matrix M_τ of τ has a single real dominant eigenvalue that is a Pisot number. A morphic word generated by a Pisot substitution is called a *Pisot word.* The Fibonacci and Tribonacci words we encountered are both Pisot words that also belong to $\mathcal{T}_O$. The Fibonacci word is the coding of a rotation with respect to two interval subsets of $\mathbb{T}$, whereas the Tribonacci word is the coding of $(\Gamma^n)_{n\in\mathbb{N}}$, where $\Gamma = (e^{\boldsymbol{i}\frac{2\pi}{x}}, e^{\boldsymbol{i}\frac{2\pi}{x^2}})$ and $x \approx 1.839$ is the largest root of the polynomial $x^3 - x^2 - x - 1$, with respect to $\mathcal{S} = \{S_1, S_2, S_3\}$ with fractal boundaries (see Section 4).

To state the Pisot conjecture, we first need a few definitions. The *language* $\mathcal{L}(\alpha)$ of $\alpha \in \Sigma^\omega$ is the set of all factors of α. Recall that a substitution $\tau\colon \Sigma^* \to \Sigma^*$ is primitive if there exists $k \in \mathbb{N}$ such that starting from any letter a, $\tau^k(a)$ contains all possible letters. Further recall that a fixed point of a primitive substitution is strongly and effectively almost-periodic. A substitution τ is *unimodular* if $\det(M_\tau) = \pm 1$. Finally, τ is irreducible if the characteristic polynomial of M_τ is irreducible. The **Pisot conjecture** states that if α is a fixed point of a unimodular, primitive and irreducible

Pisot substitution over a k-letter alphabet, then there exists a word β with the following properties.

(a) $\mathcal{L}(\beta) = \mathcal{L}(\alpha)$, and

(b) β is the toric word generated by $(\Gamma, \mathcal{S})$ where $\Gamma \in \mathbb{T}^{k-1}$ and each set in $\mathcal{S}$ is open.

Statement (b) implies $\beta \in \mathcal{T}_O$. Note that by (a), the word β is also strongly and effectively almost-periodic. The Pisot conjecture is widely believed to be true but has only been proven for $k = 2$; see [8] for a detailed account.

6.3. Sign patterns of linear recurrence sequences

The *sign pattern* of a real-valued LRS $(u_n)_{n\in\mathbb{N}}$ is the word $\alpha \in \{+, 0, -\}^\omega$ such that $\alpha(n)$ is defined by $\text{sign}(u_n)$ for all $n \in \mathbb{N}$. The Skolem, Positivity and Ultimate Positivity problems introduced in Section 2 are all decision problems about such sign patterns. We will see that sign patterns of LRS can have distinctive combinations of toricity and almost periodicity properties.

We start with *diagonalisable* (also known as *simple*) sequences. An LRS $(u_n)_{n\in\mathbb{N}}$ over R is called diagonalisable if it can be expressed in the form $u_n = c^\top M^n s$ where $c, s \in R^d$ and $M \in R^{d\times d}$ is diagonalisable. Using a deep result [33] of Evertse on the sums of S-units, we can show that the sign pattern α of a diagonalisable LRS $(u_n)_{n\in\mathbb{N}}$ has a suffix that belongs to $\mathcal{T}_O$.

Theorem 6.3 (Theorem 11 in [9])**.** *Let $(u_n)_{n\in\mathbb{N}}$ be a diagonalisable LRS over $\mathbb{R} \cap \overline{\mathbb{Q}}$ with the sign pattern $\alpha \in \{+, 0, -\}^\omega$.*

(a) There exist integers d, N, open semialgebraic subsets S_+, S_0, S_- of $\mathbb{T}^d$, and $\Gamma \in (\mathbb{T} \cap \overline{\mathbb{Q}})^d$ such that $\alpha[N, \infty) \in \mathcal{T}_O \cap \mathcal{T}_{SA(\mathbb{Q})}$ and is generated by $(\Gamma, \{S_+, S_0, S_-\})$.

(b) The value of N and representations of S_+, S_0, S_- can be effectively computed assuming decidability of the Positivity Problem for LRS over $\mathbb{Q}$.

Sign patterns of non-diagonalisable LRS, however, do not have such properties. We next give an example of a sign pattern of a non-diagonalisable LRS that is almost-periodic but provably does not belong to $\mathcal{T}_O$ nor to $\mathcal{T}_{SA}$. Let $\gamma = 0.6 + 0.8\boldsymbol{i} \in \mathbb{T} \cap \overline{\mathbb{Q}}$ and $\theta = \text{Log}(\gamma)/\boldsymbol{i}$, noting that γ is not a root of unity. Consider the linear recurrence sequences $u_n = \sin(n\theta)$ and $v_n = n\sin(n\theta) - 7\cos(n\theta)$. Write $\alpha, \beta \in \{+, 0, -\}^\omega$ for their sign patterns, respectively.

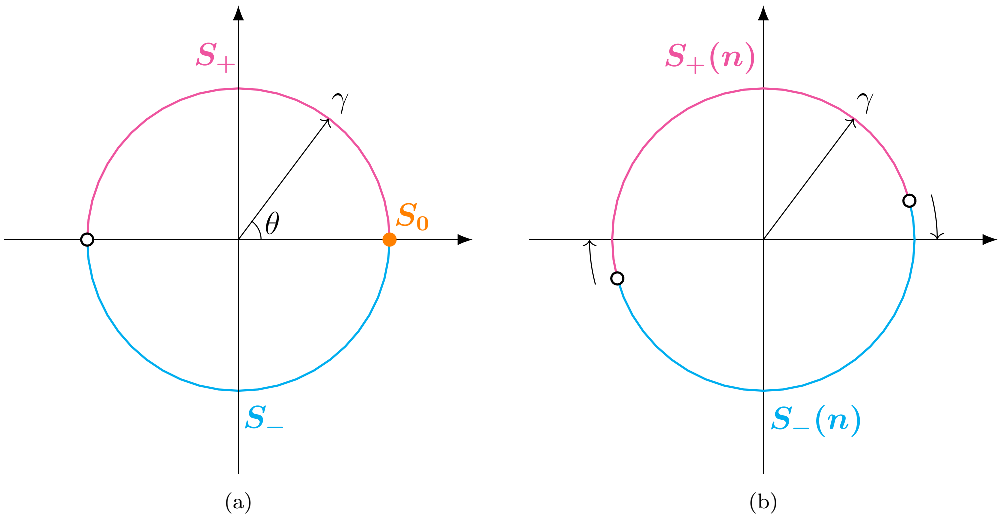

Figure 2: Target intervals for $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ in the proof of Lemma 6.4

Lemma 6.4. *Both α and β are effectively almost-periodic.*

Proof. Sequences $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ are non-degenerate LRS of order 2 and 4, respectively. Hence by [18] both sequences have finitely many zeros. In fact, we can identify all of them. Our sequences satisfy recurrence relations $u_{n+2} = 1.2u_{n+1} + u_n$ and

$$v_{n+4} = 2.4v_{n+3} - 3.44v_{n+2} + 2.4v_{n+1} - v_n.$$

Since γ is not a root of unity, it is immediate that $u_n = 0$ only for $n = 0$. We can determine all zeros of $(v_n)_{n\in\mathbb{N}}$ either using the general algorithm for solving the Skolem Problem for LRS over $\mathbb{R} \cap \overline{\mathbb{Q}}$ of order four [14, 15], or a simple approach based on the *(absolute logarithmic) Weil height*. The Weil height $h(z)$ of an algebraic number has the following properties.

(a) $h(z) > 0$ for every non-zero z that is not a root of unity;

(b) $h(k) = \log(|k|)$ for $k \in \mathbb{Z}$;

(c) $h(z^n) = nh(z)$ for every $z \in \overline{\mathbb{Q}}$ and $n \in \mathbb{Z}$;

(d) $h(z \cdot y), h(z + y) \leq h(z) + h(y) + \log(2)$ for all $z, y \in \overline{\mathbb{Q}}$.

See [34] for a detailed discussion of the Weil height. We have that $v_n = 0$ if and only if $z^n = y_n$, where $z = \gamma/\overline{\gamma}$ and $y_n = \frac{7-n\boldsymbol{i}}{7+n\boldsymbol{i}}$. Both z and y_n for all n are algebraic numbers of degree at most 2. From (c) and (d), $h(y_n) < C \log n$ for an effectively computable constant C, whereas $h(z^n) = nh(z)$ by (b). Since γ is non-zero and not a root of unity, $h(\gamma) \neq 0$. Therefore, $h(z^n)$ grows linearly, whereas $h(y_n)$ grows logarithmically in n. Equating $h(z^n)$ to $h(y_n)$, we conclude that $v_n \neq 0$ for all $n \geq N$, where N is effectively computable. Checking all $n \leq N$ individually, we find that for $n \geq 1$, $v_n \neq 0$. Therefore, $z(n), y(n) \in \{+, -\}$ for all $n \geq 1$.

Figure 2 (a) describes how $\alpha \in \mathcal{T}_{SA}$ is generated. Both S_+ and S_- are open subsets of $\mathbb{T}$, and $S_0 = \{1\}$. For all $n \in \mathbb{N}$, $\alpha(n)$ is $+$ if and only if $\gamma^n \in S_+$ and $\alpha(n)$ is $-$ if and only if $\gamma^n \in S_-$. Since $\alpha(n) \in \{+, -\}$ for $n \geq 1$, $\alpha[1, \infty) \in \mathcal{T}_O \cap \mathcal{T}_{SA}$, and is generated by $(\gamma, \{\gamma^{-1}S_+, \gamma^{-1}S_-\})$. Applying Corollary 5.8, $\alpha[1, \infty)$ and hence α are both effectively almost-periodic.

Let us consider β next. Let $\delta_n = \arctan(7/n) \in (0, \pi/2)$, $S_+(n) = e^{\boldsymbol{i}\delta_n} S_+$, and $S_-(n) = e^{\boldsymbol{i}\delta_n} S_-$. We have that for $n \geq 1$, $v_n > 0$ if and only if $\gamma^n \in S_+(n)$ and $v_n < 0$ if and only if $\gamma^n \in S_-(n)$. Figure 2 (b) depicts $S_+(n)$ and $S_-(n)$ for $n = 30$. Since $(e^{-\boldsymbol{i}\delta_n})_{n \in \mathbb{N}}$ converges to 1, as $n \to \infty$, $S_+(n)$ uniformly approaches the upper half S_+ of the unit circle, whereas $S_-(n)$ approaches S_-.

To prove effective almost periodicity of β, consider a finite word

$$w = w(0) \cdots w(l-1) \in \{+, -\}^l.$$

It occurs at the position $n \geq 1$ in β if and only if

$$\bigwedge_{j=0}^{l-1} \gamma^{n+j} \in S_{w(j)}(n+j) \;\Leftrightarrow\; \gamma^n \in \bigcap_{j=0}^{l-1} \gamma^{-j} S_{w(j)}(n+j).$$

Define $S_w(n) = \bigcap_{j=0}^{l-1} \gamma^{-j} S_{w(j)}(n+j)$. We will argue that either w occurs finitely often in β, or there exists an open interval subset K of $\mathbb{T}$ such that $K \subset S_w(n)$ for all sufficiently large n.

Recall that for distinct $z_1, z_2 \in \mathbb{T}_\Gamma$ we denote by $\mathcal{I}(z_1, z_2)$ the open interval subset of $\mathbb{T}$ with endpoints z_1 and z_2, generated by rotating counter-clockwise starting at z_1. Each $\gamma^{-j} S_{w(j)}(n+j)$ is of the form $e^{\boldsymbol{i}\delta(n+j)} \gamma^{-j} I_j$, where I_j is S_+ if $w(j)$ is the letter $+$ and $I_j = S_-$ otherwise. Since $\delta(n) = \Theta(1/n)$, $\gamma^{-j} S_{w(j)}(n+j)$ uniformly approaches the interval $\gamma^{-j} I_j$ as $n \to \infty$.

The endpoints of $\gamma^{-j} I_j$ are γ^{-j} and $-\gamma^{-j}$. As γ is not a root of unity, for every $j_1 \neq j_2$, γ^{-j_1} is not equal to γ^{-j_2} and $-\gamma^{-j_2}$. Hence the limit intervals $\gamma^{-j} I_j$ for $0 \leq j < l$ have $2l$ distinct endpoints in total. Therefore,

(a) either there exists N such that $S_w(n)$ is empty for all $n \geq N$ (which happens if and only if the "limit shape" $\bigcap_{j=0}^{l-1} \gamma^{-j} I_j$ is empty), or

(b) there exists N such that for all $n \geq N$, $S_w(n) = \mathcal{I}(z_1 e^{\boldsymbol{i}\delta_1(n)}, z_2 e^{\boldsymbol{i}\delta_2(n)})$ is non-empty, where z_1, z_2 are distinct and of the form $\pm\gamma^{-j}$ for some $0 \leq j < l$ and $\delta_1(n), \delta_2(n) = \Theta(1/n)$.

Since all steps above are effective, we can effectively compute N in both cases, and in case (b), construct a $\mathbb{Q}$-semialgebraic interval J such that for all $n \geq N$, $J \subset S_w(n)$. In case (a) the word w does not occur in $\beta[N, \infty)$ and we are done. Otherwise, observe that for $n \geq N$, $\gamma^n \in J \Rightarrow \beta[n, n+l) = w$. Since the endpoints of J are algebraic, we can compute K such that for all $m \in \mathbb{N}$, $\gamma^n \in J$ for some $m \leq n \leq m + K$; see the proof of Theorem 5.7 for the usual topological construction, or [35, Lemma 2] for a direct formula. We conclude that the word w occurs in every subword of β of length $N+K+l$. $\square$

The discussion above suggests to think of β as being "toric with moving targets". We next show that $\alpha \times \beta$ is radically different from both α and β, and far from belonging to $\mathcal{T}_O$ or $\mathcal{T}_{SA}$.

Theorem 6.5. *For α, β as in Lemma 6.4, the word $\delta = \alpha \times \beta$ is not almost-periodic and hence does not belong to $\mathcal{T}_O \cup \mathcal{T}_{SA}$.*

Proof. Recall from Theorems 5.7 and 5.9 that all words belonging to $\mathcal{T}_O$ or $\mathcal{T}_{SA}$ are almost-periodic. We thus have to only prove the first statement. We will show that (a) the letter $(+,-)$ occurs infinitely often in δ, and (b) the length of the gaps between its consecutive occurrences is not bounded.

We start with (a). The letter $(+,-)$ occurs at a position $n > 0$ if and only if $\sin(n\theta) > 0$ and $n\sin(n\theta) - 7\cos(n\theta) < 0$, which is equivalent to $0 < \mathrm{Log}(\gamma^n) < \arctan(7/n)$. We will show that $0 < \mathrm{Log}(\gamma^n) < 2\pi/n$ is satisfied for infinitely many $n \in \mathbb{N}$. Since $\arctan(7/n) > 2\pi/n$ for $n > 11$, this proves that $(+,-)$ occurs infinitely often in δ.

Let $t = \frac{1}{2} + \mathrm{Log}(\gamma)/(2\pi\boldsymbol{i})$, which is an irrational number between 0 and 1. For $n \geq 1$, $\mathrm{Log}(\gamma^n) \in (0, 2\pi/n)$ if and only if $nt - \lfloor nt \rfloor < 1/n$. We find infinitely many values of n satisfying the latter inequality using the *continued fraction expansion* of t:

$$t = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \dots}}}$$

where each a_i is a positive integer; see [36]. Let p_n/q_n be the rational approximation of t obtained by truncating the expansion at the nth level. For all n, it holds that

$$q_{n+1}t - p_{n+1} = \frac{(-1)^{n+1}}{a_{n+2}q_{n+1} + q_n}.$$

In particular, truncations with odd n are over-approximations, and truncations with even n are under-approximations of t. Moreover, for every n, $|p_n/q_n - t| < 1/q_n$, and $(q_n)_{n\in\mathbb{N}}$ is strictly increasing. Therefore, for every even $n \geq 1$,

$$0 < t - \frac{p_n}{q_n} < \frac{1}{q_n^2}$$

and hence $nt - \lfloor nt \rfloor < 1/q_n$.

We move on to proving (b). Let $J_n = S_+ \cap S_-(n)$. Recall that the letter $(+,-)$ occurs at the position n in β if and only if $\gamma^n \in J_n$, and the length of J_n is $\Theta(1/n)$. Let $B \in \mathbb{N}$. We show how to construct n such that letter $(+,-)$ does not occur in $\beta[n, n+B)$. Let m be sufficiently large that $\mathbb{T} \setminus \bigcup_{i=0}^{B} \gamma^{-i} J_m$ contains a non-empty open subset O of $\mathbb{T}$. Further let $n \geq m$ be such that $\gamma^n \in O$. By construction, for every $0 \leq i \leq B$, $\gamma^{n+i} \notin J_m$. Since $J_{m+i} \subset J_m$ for all $i \in \mathbb{N}$, it holds that for all $0 \leq i < B$, $\gamma^{n+i} \notin J_{n+i}$. That is, for all $0 \leq i < B$, $\delta(n+i)$ is not the letter $(+,-)$. □

Corollary 6.6. *The word δ does not belong to $\mathcal{T}_O \cup \mathcal{T}_{SA}$.*

Proof. Recall that α belongs to both $\mathcal{T}_O$ and $\mathcal{T}_{SA}$, and both classes are closed under products. Since $\alpha \times \beta$ does not belong to $\mathcal{T}_O \cup \mathcal{T}_{SA}$, neither does β. □

6.4. Characteristic words of linear dynamical systems

One application of toric words and MSO decidability that has recently received significant attention is the *Model-Checking Problem (MCP)* for Linear Dynamical Systems [10]. An LDS is given by a pair (M, s) where $M \in \mathbb{Q}^{d\times d}$ is the update matrix and $s \in \mathbb{Q}^d$ is the starting configuration. The *orbit* of (M, s) is the infinite sequence $(M^n s)_{n\in\mathbb{N}}$. Let $\mathcal{S} = \{S_1, \ldots, S_m\}$ be a collection of $\mathbb{Q}$-semialgebraic subsets of $\mathbb{R}^d$. Writing $\Sigma = 2^{\mathcal{S}}$, the *characteristic word* of (M, s) with respect to $\mathcal{S}$ is the word $\alpha \in \Sigma^\omega$ defined by $S_i \in \alpha(n) \Leftrightarrow M^n s \in S_i$ for all $1 \leq i \leq m$ and $n \in \mathbb{N}$. The Model-Checking Problem is to decide, given (M, s) and a deterministic automaton $\mathcal{A}$, whether $\mathcal{A}$ accepts α. If we fix $M, s, \mathcal{S}$, and only let $\mathcal{A}$ vary, by Büchi's result [1], the resulting problem is Turing-equivalent to the decision problem for the MSO theory of $\langle \mathbb{N}; <, P_1, \ldots, P_m \rangle$, where each $P_i \colon \mathbb{N} \to \{0, 1\}$ is the binary predicate defined by $P_i(n) = 1$ if and only if $M^n s \in S_i$ for all $n \in \mathbb{N}$.

Let $p_1, \ldots, p_K$ be all polynomials (with rational coefficients) appearing in the definition of $\mathcal{S}$. For each $1 \leq j \leq K$, the sequence $u_n = p_j(M^n s)$ is an LRS over $\mathbb{Q}$. Denote its sign patter by $\alpha_j \in \{+, 0, -\}^\omega$. Since each S_i is generated by a Boolean combination of polynomial inequalities, we have that $\alpha = \sigma(\alpha_1 \times \ldots \times \alpha_K)$, where σ is a 1-uniform morphism. Hence understanding the characteristic word of an LDS with respect to a collection of semialgebraic sets $\mathcal{S}$ boils down to understanding sign patterns of a collection of linear recurrence sequences.

The Model-Checking Problem for LDS subsumes, among many others, the Skolem Problem, the Positivity Problem, and the Ultimate Positivity Problem for LRS over $\mathbb{Q}$. Unsurprisingly, decidability of the full Model-Checking Problem is currently open. However, decidability can be proven if we place certain restrictions on $M, \mathcal{A}$ and $\mathcal{S}$.

(A) Call a $\mathbb{Q}$-semialgebraic set T *low-dimensional* if it either has intrinsic (i.e. semialgebraic) dimension 1, or is contained in a three-dimensional linear subspace. The set T is *tame* if it can be generated from a collection of low-dimensional sets through the usual set operations. If all targets in $\mathcal{S}$ are tame, then the characteristic word α of any LDS (M, s) with respect to $\mathcal{S}$ is effectively almost-periodic [37, 10]. In particular, α has a suffix belonging to the class $\mathcal{T}_O \cap \mathcal{T}_{SA(\mathbb{Q})}$ that is fully effective. Hence the MCP with tame targets (but arbitrary (M, s) and $\mathcal{A}$) is decidable.

(B) An automaton $\mathcal{A}$ is *prefix-independent* if for any infinite word β, whether it is accepted does not change if we perform finitely many insertions and deletions on β. It is shown in [11] that the Model-Checking Problem is decidable if we assume M is diagonalisable and $\mathcal{A}$ is prefix-independent.

From (A) it follows that the MCP is decidable in dimension at most 3. The result (B), on the other hand, is closely related to Theorem 6.3. Too see this, suppose M is diagonalisable. Then $u_n = p(M^n s)$ is a diagonalisable LRS for every polynomial p. From the connection between the characteristic word α and the sign patterns of LRS defining $\mathcal{S}$ discussed above, the closure properties of toric words, as well as Theorem 6.3 (a), it follows that α has a suffix that belongs to $\mathcal{T}_O \cap \mathcal{T}_{SA}$. Unfortunately, it is not known how to determine the starting position of such a suffix in α, which is the reason why in (B) we impose the prefix-independence restriction. However, similarly to Theorem 6.3 (b), it is shown in [9] that the MCP is decidable for diagonalisable LDS if we assume decidability of the Positivity Problem for diagonalisable LRS over $\mathbb{Q}$.

References

[1] J. R. Büchi, On a decision method in restricted second order arithmetic, in: The Collected Works of J. Richard Büchi, Springer New York, 1990, pp. 425–435. doi:10.1007/978-1-4613-8928-6_23.

[2] C. C. Elgot, M. O. Rabin, Decidability and undecidability of extensions of second (first) order theory of (generalized) successor, The Journal of Symbolic Logic 31 (02) (1966) 169–181. doi:10.2307/2269808.

[3] R. M. Robinson, Restricted set-theoretical definitions in arithmetic, Proceedings of the American Mathematical Society 9 (2) (1958) 238–242. doi:0.1090/S0002-9939-1958-0093479-4.

[4] B. Trahtenbrot, Finite automata and the logic of one-place predicates. russian, Siberian Mathematical Journal 3 (1962) 103–131.

[5] P. T. Bateman, C. G. Jockusch, A. R. Woods, Decidability and undecidability of theories with a predicate for the primes, The Journal of symbolic logic 58 (2) (1993) 672–687. doi:10.2307/2275227.

[6] A. L. Semënov, Logical theories of one-place functions on the set of natural numbers, Mathematics of the USSR-Izvestiya 22 (3) (1984) 587–618. doi:10.1070/im1984v022n03abeh001456.

[7] O. Carton, W. Thomas, The monadic theory of morphic infinite words and generalizations, Information and Computation 176 (1) (2002) 51–65. doi:10.1006/inco.2001.3139.

[8] S. Akiyama, M. Barge, V. Berthé, J.-Y. Lee, A. Siegel, On the pisot substitution conjecture, in: Mathematics of Aperiodic Order, Springer Basel, 2015, pp. 33–72. doi:10.1007/978-3-0348-0903-0_2.

[9] T. Karimov, E. Kelmendi, J. Nieuwveld, J. Ouaknine, J. Worrell, The power of positivity, in: 2023 38th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS), 2023, pp. 1–11. doi:10.1109/LICS56636.2023.10175758.

[10] T. Karimov, E. Kelmendi, J. Ouaknine, J. Worrell, What's Decidable About Discrete Linear Dynamical Systems?, Springer Nature Switzerland, Cham, 2022, pp. 21–38. doi:10.1007/978-3-031-22337-2_2.

[11] S. Almagor, T. Karimov, E. Kelmendi, J. Ouaknine, J. Worrell, Deciding ω-regular properties on linear recurrence sequences, Proc. ACM Program. Lang. 5 (POPL) (2021) 1–24. doi:10.1145/3434329.

[12] J. Berstel, C. Reutenauer, Noncommutative Rational Series with Applications, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2010. doi:10.1017/CBO9780511760860.

[13] T. Skolem, Einige sätze über π-adische potenzreihen mit anwendung auf gewisse exponentielle gleichungen, Mathematische Annalen 111 (1) (1935) 399–424. doi:10.1007/BF01472228.

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

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

[16] K. Mahler, An arithmetic property of taylor coefficients of rational functions (1935), in: M. Baake, Y. Bugeaud, M. Coons (Eds.), The Legacy of Kurt Mahler, EMS Press, 2019, pp. 437–448. doi:10.4171/DMS/8.

[17] C. Lech, A note on recurring series, Arkiv för Matematik 2 (5) (1953) 417–421. doi:10.1007/BF02590997.

[18] J. Berstel, M. Mignotte, Deux propriétés décidables des suites récurrentes linéaires, Bulletin de la Societe mathematique de France 79 (1976) 175–184. doi:10.24033/bsmf.1823.

[19] G. Rozenberg, A. Salomaa, Cornerstones of Undecidability, Prentice-Hall, Inc., 1995.

[20] A. Salomaa, M. Soittola, Automata-Theoretic Aspects of Formal Power Series, Springer New York, 1978. doi:10.1007/978-1-4612-6264-0.

[21] J. Ouaknine, J. Worrell, Positivity problems for low-order linear recurrence sequences, in: Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '14, Society for Industrial and Applied Mathematics, USA, 2014, p. 366–379. doi:10.1137/1.9781611973402.2.

[22] J.-P. Allouche, J. Shallit, Automatic Sequences: Theory, Applications, Generalizations, Cambridge University Press, 2003. doi:10.1017/CBO9780511546563.

[23] M. Queffélec, Substitution Dynamical Systems – Spectral Analysis, 2nd Edition, Vol. 1294 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2010. doi:10.1007/BFb0081890.

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

[25] L. Kari, G. Rozenberg, A. Salomaa, L Systems, Springer Berlin Heidelberg, 1997, pp. 253–328. doi:10.1007/978-3-642-59136-5_5.

[26] G. Rauzy, Nombres algébriques et substitutions, Bull. Soc. Math. France 110 (2) (1982) 147–178. doi:10.24033/bsmf.1957.

[27] A. Salomaa, Jewels of Formal Language Theory, Computer software engineering series, Computer Science Press, 1981.

[28] J. Berstel, P. Séébold, A remark on morphic sturmian words, RAIRO-Theoretical Informatics and Applications 28 (3-4) (1994) 255–263. doi:10.1051/ITA/1994283-402551.

[29] A. Rabinovich, On decidability of monadic logic of order over the naturals extended by monadic predicates, Information and Computation 205 (6) (2007) 870–889. doi:10.1016/j.ic.2006.12.004.

[30] A. Rabinovich, W. Thomas, Decidable theories of the ordering of natural numbers with unary predicates, in: International Workshop on Computer Science Logic, Springer, 2006, pp. 562–574. doi:10.1007/11874683_37.

[31] D. W. Masser, Linear relations on algebraic groups, Cambridge University Press, 1988, p. 248–262. `doi:10.1017/CBO9780511897184.016`.

[32] S. M. Gonek, H. L. Montgomery, Kronecker's approximation theorem, Indagationes Mathematicae 27 (2) (2016) 506–523, in Memoriam J.G. Van der Corput (1890–1975) Part 2. `doi:https://doi.org/10.1016/j.indag.2016.02.002`.

[33] J.-H. Evertse, On sums of S-units and linear recurrences, Compositio Mathematica 53 (2) (1984) 225–244.

[34] M. Waldschmidt, Heights of Algebraic Numbers, Springer Berlin Heidelberg, Berlin, Heidelberg, 2000, pp. 65–114. `doi:10.1007/978-3-662-11569-5_3`.

[35] T. Karimov, J. Ouaknine, J. Worrell, On LTL Model Checking for Low-Dimensional Discrete Linear Dynamical Systems, in: J. Esparza, D. Kráľ (Eds.), 45th International Symposium on Mathematical Foundations of Computer Science (MFCS 2020), Vol. 170 of Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2020, pp. 54:1–54:14. `doi:10.4230/LIPIcs.MFCS.2020.54`.

[36] J. Borwein, A. van der Poorten, J. Shallit, W. Zudilin, Neverending Fractions: An Introduction to Continued Fractions, Australian Mathematical Society Lecture Series, Cambridge University Press, 2014. `doi:10.1017/CBO9780511902659`.

[37] T. Karimov, E. Lefaucheux, J. Ouaknine, D. Purser, A. Varonka, M. A. Whiteland, J. Worrell, What's decidable about linear loops?, Proc. ACM Program. Lang. 6 (POPL) (jan 2022). `doi:10.1145/3498727`.