

# Automata on $S$ -adic words

Valérie Berthé   

Université Paris Cité, IRIF, CNRS, France

Toghrul Karimov   

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

Mihir Vahanwala   

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

---

## Abstract

A fundamental question in logic and verification is the following: for which unary predicates  $P_1, \dots, P_k$  is the monadic second-order theory of  $\langle \mathbb{N}; <, P_1, \dots, P_k \rangle$  decidable? Equivalently, for which infinite words  $\alpha$  can we decide whether a given Büchi automaton  $\mathcal{A}$  accepts  $\alpha$ ? Carton and Thomas showed decidability in the case that  $\alpha$  is a fixed point of a letter-to-word substitution  $\sigma$ , i.e.,  $\sigma(\alpha) = \alpha$ . However, abundantly more words, e.g., Sturmian words, are characterised by a broader notion of self-similarity that uses a set  $S$  of substitutions. A word  $\alpha$  is said to be directed by a sequence  $s = (\sigma_n)_{n \in \mathbb{N}}$  over  $S$  if there is a sequence of words  $(\alpha_n)_{n \in \mathbb{N}}$  such that  $\alpha_0 = \alpha$  and  $\alpha_n = \sigma_n(\alpha_{n+1})$  for all  $n$ ; such  $\alpha$  are called  $S$ -adic. We study the automaton acceptance problem for such words and prove, among others, the following: given finite  $S$  and an automaton  $\mathcal{A}$ , we can compute an automaton  $\mathcal{B}$  that accepts  $s \in S^\omega$  if and only if  $s$  directs a word  $\alpha$  accepted by  $\mathcal{A}$ . Thus we can algorithmically answer questions of the form “Which  $S$ -adic words are accepted by a given automaton  $\mathcal{A}$ ?”

**2012 ACM Subject Classification** Theory of computation  $\rightarrow$  Logic and verification

**Keywords and phrases** Sturmian words,  $S$ -adic words, automata theory, word combinatorics

**Digital Object Identifier** 10.4230/LIPIcs.ICALP.2026.182

**Funding** Valérie Berthé: supported by the Agence Nationale de la Recherche through the projects “SymDynAr” (ANR-23-CE40-0024) and “IZES” (ANR-22-CE40-0011), as well as by the 2024 ERC Synergy Project DynAMiCs (101167561).

Toghrul Karimov: supported by the DFG grant 389792660 as part of TRR 248 (Foundations of Perspicuous Software Systems, see [perspicuous-computing.science](https://perspicuous-computing.science)).

Mihir Vahanwala: supported by the DFG grant 389792660 as part of TRR 248 (Foundations of Perspicuous Software Systems, see [perspicuous-computing.science](https://perspicuous-computing.science)).

**Acknowledgements** The authors are grateful to the anonymous reviewers for their helpful suggestions.

## 1 Introduction

In 1962, Büchi proved that the monadic second-order (MSO) theory of the structure  $\langle \mathbb{N}; < \rangle$  is decidable [8], and, in doing so, laid the foundations of the theory of automata over infinite words. Subsequently, Elgot and Rabin [12] adopted automata-theoretic techniques to show how to decide the MSO theory of  $\langle \mathbb{N}; <, P \rangle$  for various interesting unary predicates  $P$  including  $\{n! : n \in \mathbb{N}\}$  and  $\{2^n : n \in \mathbb{N}\}$ . By then, it was already known that unary predicates lay at the frontiers of decidability: expanding  $\langle \mathbb{N}; < \rangle$  with most natural functions (e.g., addition or doubling) or non-unary predicates yields undecidable MSO theories [23, 27, 26]. The question thus arose: for which unary predicates  $P_1, \dots, P_k$  is the MSO theory of  $\langle \mathbb{N}; <, P_1, \dots, P_k \rangle$  decidable? Equivalently, for which infinite words  $\alpha$  is the *automaton acceptance problem*<sup>1</sup>, which asks whether a given automaton  $\mathcal{A}$  accepts  $\alpha$ , decidable?

---

<sup>1</sup> Here the problem is parametrised by  $\alpha$ ; the only input is the automaton  $\mathcal{A}$ .



© Valérie Berthé, Toghrul Karimov, Mihir Vahanwala;  
licensed under Creative Commons License CC-BY 4.0

53rd International Colloquium on Automata, Languages, and Programming (ICALP 2026).

Editors: Sayan Bhattacharya, Danupon Nanongkai, Michael Benedikt, and Gabriele Puppis; Article No. 182; pp. 182:1–182:20



Leibniz International Proceedings in Informatics  
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The automaton acceptance problem under various assumptions on  $\alpha$  has been studied, among others, by Semënov, Carton and Thomas, and Rabinovich [24, 10, 22]. Semënov [24] showed decidability for words  $\alpha$  that are *effectively almost-periodic*. These include, for example, the Thue-Morse word and *toric words*, which are obtained from certain compact dynamical systems [7]. Carton and Thomas [10], on the other hand, used algebraic methods to show decidability for *morphic* words  $\alpha$ , which include  $\alpha$  that can be constructed by infinitely iterating a letter-to-word morphism  $\sigma$  on a starting letter  $a$ . Their result implies, in one fell swoop, the decidability of the MSO theory of  $\langle \mathbb{N}; <, \{p(n)a^n : n \in \mathbb{N}\} \rangle$  where  $a \geq 1$  and  $p$  is a polynomial with integer coefficients satisfying  $p(\mathbb{N}) \subseteq \mathbb{N}$ .

By now, the study of the automaton acceptance problem for various special classes of  $\alpha$  has led to rich interactions between word combinatorics, algebra (particularly monoid and group theory), number theory, and formal verification. For example, effectively almost-periodic words of Semënov have been identified as a powerful tool for analysing the behaviour of linear while loops in program verification [19]. More recently, [6] showed the decidability of the MSO theory of  $\langle \mathbb{N}; <, a^{\mathbb{N}}, b^{\mathbb{N}} \rangle$  using the fact that the order in which powers of coprime  $a$  and  $b$  occur is effectively almost-periodic by virtue of being captured by a computable *Sturmian word* (Sec. 3.2). Sturmian words are an extremely well-studied and fundamental class of “special words” that appear naturally in a range of fields including number theory, computer graphics, and astronomy [1, Chap. 9.6], [13, Chap. 6]. We refer the reader to [7] for a more detailed survey of the role of Sturmian words, and word combinatorics in general, in logic and verification.

In this paper, we are motivated by questions of the form, “Given an  $\omega$ -regular language  $L$ , decide whether it contains a Sturmian word.” (Of course, we can also ask about any other combinatorial class.) Our approach adopts the *S-adic perspective*, a powerful tool for elucidating combinatorial properties of infinite words. Akin to the continued fraction expansion of a real number, or even a Fourier decomposition of a signal, we write an infinite word  $\alpha$  as an infinite composition of (possibly different) *substitutions*. A substitution  $\sigma$  over an alphabet  $\Sigma$  gives rules to replace each letter  $a \in \Sigma$  with a non-empty word  $\sigma(a) \in \Sigma^+$ . For example, the Fibonacci substitution  $\sigma_{\text{fib}}$  over  $\{0, 1\}$  replaces 0 with 01 and 1 with 0; the Fibonacci word  $\alpha_{\text{fib}} = 01001010 \dots$  is obtained as the limit of iterating  $\sigma_{\text{fib}}$  infinitely on the letter 0 (or, alternatively, the letter 1). Hence we have the infinite decomposition  $\alpha_{\text{fib}} = \sigma_{\text{fib}} \circ \dots \circ \sigma_{\text{fib}} \dots$ . In general, we have a set  $S$  of substitutions, and say that a sequence  $s$  over  $S$  *directs*  $\alpha \in \Sigma^\omega$  if there exists a sequence of words  $(\alpha^{(n)})_{n \in \mathbb{N}}$  such that  $\alpha^{(0)} = \alpha$ , and  $\alpha^{(n)} = \sigma_n(\alpha^{(n+1)})$  for all  $n$ . This gives us the *S-adic decomposition*  $\alpha = \sigma_0 \circ \sigma_1 \circ \dots$ . We refer to any  $s \in S^\omega$  as a *directive sequence*.

An observant reader might argue that every word  $\alpha = a_0 a_1 a_2 \dots$  is vacuously *S-adic* for  $S = \{\sigma_1, \dots, \sigma_{|\Sigma|}, \tau_1, \dots, \tau_{|\Sigma|}\}$  over the alphabet  $\Sigma \cup \{b\}$ , where  $\sigma_i$  replaces the distinguished letter  $b$  with  $a_i \in \Sigma$  and is the identity function elsewhere, and  $\tau_i$  replaces  $b$  with  $ba_i$  and is the identity function elsewhere. We then see that  $\alpha$  is directed by  $\sigma_{a_0} \tau_{a_1} \tau_{a_2} \dots$ . The vacuous construction notwithstanding, the point of directive sequences is to provide a perspective that conveys more structural information about the word. E.g., characteristic Sturmian words are *parametrised* by directive sequences over the alphabet  $\{\lambda_0, \lambda_1\}$  such that both letters occur infinitely often; a directive sequence encodes the continued fraction expansion of the *slope* of its Sturmian word. We show that this alternate perspective can still be used for verification, even when it does not give a vacuous reformulation of the word.

Indeed, entire classes of words such as Sturmian and Arnoux-Rauzy words (Sec. 3.2) can be parametrised in terms of directive sequences over a specific  $S$ ; such a class  $\mathcal{W}$  is called *S-adic*. Our central question is the following: given an  $\omega$ -regular language  $L \subseteq \Sigma^\omega$  and

an  $S$ -adic class of words  $\mathcal{W} \subseteq \Sigma^\omega$ , what is the set of all  $\alpha \in \mathcal{W}$  that are contained in  $L$ ? Does there exist at least one such  $\alpha$ ? Equivalently, given an MSO formula  $\varphi$ , can we decide whether there exists  $\alpha \in \mathcal{W}$  that induces a structure in which  $\varphi$  holds? We remark that this question is similar in spirit to the problem solved in [17]: given a *first-order* formula  $\varphi$ , it is decidable whether there exists a Sturmian word  $\alpha$  such that  $\varphi$  holds in the induced first-order structure  $\langle \mathbb{N}; <, +, P_\alpha := \{n: \alpha(n) = 1\} \rangle$ . The key idea there is that the class of structures  $\langle \mathbb{N}; <, +, P_\alpha \rangle$ , where  $\alpha$  ranges over all Sturmian words, is *automatic* [25].

Our main contribution is that  $\omega$ -regular specifications on words in the  $\Sigma$ -space translate to  $\omega$ -regular constraints on the directive sequences in the  $S$ -space of substitutions. Hence the common MSO theory of an  $S$ -adic class of words is decidable. This generalises the result of [2] (in particular, Cor. 3 *ibidem*) that, given finite  $S$  and an automaton  $\mathcal{A}$  whose language is closed (i.e., a Büchi automaton whose states are all accepting), the set of sequences over  $S$  that direct a word accepted by  $\mathcal{A}$  constitutes a closed  $\omega$ -regular language, which is moreover computable. We consider arbitrary  $\omega$ -regular languages, are equipped to handle both directed words and generated words, and relax the finiteness assumption on  $S$ .

Our algorithms thoroughly answer questions of the kind “Which  $S$ -adic words are accepted by a given automaton  $\mathcal{A}$ ?” by translating it to, “Which parametrising directive sequences are accepted by the constructed automaton  $\mathcal{B}$ ?” For example, given two words  $u, v \in \{0, \dots, d-1\}^*$  and  $a, b$  with  $b \neq 0$ , we can compute an effective representation (as an  $\omega$ -regular language of directive sequences) of all Arnoux-Rauzy words in which between any two consecutive occurrences of  $u$ , the number of occurrences of  $v$  is congruent to  $a$  modulo  $b$ .

## Outline and contributions of the paper

In Sec. 2 we establish the necessary mathematical background. In Sec. 3 we formally define what it means for a directive sequence to generate and to direct a word. Briefly, a sequence of substitutions  $(\sigma_n)_{n \in \mathbb{N}}$  generates  $\alpha$  if there exists a sequence of letters  $(a_n)_{n \in \mathbb{N}}$  such that  $\alpha = \lim_{n \rightarrow \infty} \sigma_0 \cdots \sigma_n(a_n)$ ; generating a word is a strong form of directing it. We then recall well-known properties of directive sequences, the most important of them being *weak primitivity*, and then describe Sturmian and Arnoux-Rauzy words, which are the best-known examples of  $S$ -adic words (see e.g., [13, 4]).

In Sec. 4 we study the structure of words directed or generated by directive sequences. Our key insight is the augmentation of a directive sequence  $s$  over<sup>2</sup>  $S$  into a *congenial* expansion  $\hat{s}$  over  $S \times \Sigma$  (Def. 9), which generates a word incrementally and predictably (Lem. 10). We prove the following pivotal results.

- If  $s$  generates  $\alpha$ , then it also directs  $\alpha$  (Lem. 6).
- A word  $\alpha$  directed by  $s$  is a concatenation of words generated congenially by  $s$  (Lem. 12).
- For  $s$  weakly primitive,  $s$  congenially generates  $\alpha$  if and only if it directs  $\alpha$  (Lem. 14).

In Sec. 5 we introduce our second key ingredient, namely a suitable equivalence relation of finite index on substitutions. Intuitively, we take the “quotient” of the monoid of substitutions modulo the syntactic monoid of a given  $\omega$ -regular language  $L$  (Prop. 16); we denote the finite set of equivalence classes by  $\Xi_L$ . Given a sequence  $s$  over the set of substitutions  $S$  or a sequence  $\hat{s}$  over  $S \times \Sigma$ , we naturally define its trace to be, respectively, a word over  $\Xi_L$

<sup>2</sup> Our set  $S$  of substitutions could possibly be infinite. The technically advanced reader might identify that a directive sequence induces an ordered Bratteli diagram whose vertices are letters of  $\Sigma$ , and there is an edge between  $b$  in level  $n+1$  and  $a$  in level  $n$  labelled  $r$  if  $a$  occurs in the  $r$ -th position of  $\sigma_n(b)$ . Congenial expansions mark minimal paths in this diagram, and are guaranteed to exist by König’s lemma. This remark may be ignored by a reader unfamiliar with this terminology.

or over  $\Xi_L \times \Sigma$ . In Sec. 6 and 7 we use our two key ingredients, the notion of congeniality together with the monoid  $\Xi_L$  of equivalence classes, to prove our main results.

**Morphic Words** Let  $\mathcal{A}$  be an automaton over  $\Sigma$ , and  $\sigma, \pi$  be substitutions. Using only the respective equivalence classes  $\xi, \zeta \in \Xi_L$  of the latter, we can compute a regular language  $L \subseteq \Sigma^+$  such that the word  $\pi(\sigma^\omega(u))$  is well-defined and accepted by  $\mathcal{A}$  if and only if  $u \in L$  (Thm. 21). We can thus characterise all such  $\pi, \sigma, u$ , which generalises the result of [10] that it is decidable whether a given morphic word is accepted by  $\mathcal{A}$ .

**Generated Words** Given an automaton  $\mathcal{A}$  over  $\Sigma$ , we can construct an automaton  $\mathcal{B}$  over  $\Xi_L \times \Sigma$  such that  $\mathcal{B}$  accepts the trace of  $\hat{s}$  if and only if  $\hat{s}$  is congenial and generates a word accepted by  $\mathcal{A}$  (Thm. 23).

**Directed Words** Given  $\mathcal{A}$  as above, we can construct an automaton  $\mathcal{B}$  over  $\Xi_L$  such that  $\mathcal{B}$  accepts the trace of  $s$  if and only if  $s$  directs a word accepted by  $\mathcal{A}$  (Thm. 26). This further generalises the result of [10] from infinite compositions of a single substitution  $\sigma$  to arbitrary infinite compositions over a set of substitutions  $S$ .

In Sec. 8, we refine our main results for Sturmian and Arnoux-Rauzy words, which have an *a priori* known *factor complexity* (Sec. 2.3). We show that for such classes, acceptance by  $\mathcal{A}$  is completely determined by finitely many (depending only on  $\mathcal{A}$  and not  $\alpha$ ) *partial quotients* of the directive sequence (Thm. 30). In the case of Sturmian words, this has a nice geometric interpretation: an automaton can only resolve the slope and intercept associated with a Sturmian word up to a “pre-determined” finite precision.

## 2 Preliminaries

An alphabet  $\Sigma$  is a finite set of symbols. Throughout this paper we assume that  $|\Sigma| > 1$ , i.e. we have at least two distinct letters. We write  $\varepsilon$  for the empty word. For a word  $\alpha$ ,  $\alpha(j)$  denotes the letter at the  $j$ th position of  $\alpha$ ,  $\alpha[i, j)$  denotes the finite word  $\alpha(i) \cdots \alpha(j-1)$ , and  $\alpha[j, \infty)$  denotes the infinite word  $\alpha(j)\alpha(j+1) \cdots$ . A finite word  $u$  is a factor of a word  $v$  if there exist indices  $i, j$  such that  $v[i, j) = u$ . When we say that an object is computable, we mean that a representation in a scheme (that will be clear from the context) is computable.

A substitution  $\sigma$  is a non-erasing morphism from  $\Sigma^*$  to  $\Sigma^*$ , i.e.,  $\sigma(v) = \varepsilon$  if and only if  $v = \varepsilon$ . We denote the set of all such substitutions by  $S(\Sigma)$ . For substitutions  $\mu, \sigma$ , we write  $\mu\sigma$  for  $\mu \circ \sigma$ . A substitution is positive if every  $b \in \Sigma$  appears in  $\sigma(a)$  for all  $a \in \Sigma$ , and left-proper if the images of all letters by  $\sigma$  begin with the same letter. For more on the subject of substitutions, see e.g., [13].

### 2.1 Topology of finite and infinite words

We equip  $\Sigma^\infty := \Sigma^+ \cup \Sigma^\omega$  with the product topology, and define the distance between distinct words  $u, v$  to be  $2^{-n}$ , where  $n$  is the first position in which they differ. E.g., distinct  $a, b \in \Sigma$  are a distance of  $2^0 = 1$  apart. A notion of convergence of sequences of words follows naturally.

► **Definition 1.** Let  $(u_n)_{n \in \mathbb{N}}$  be a sequence of finite non-empty words. We define  $\alpha = \lim_{n \rightarrow \infty} u_n \in \Sigma^\infty \cup \{\perp\}$  as follows.

- If there exists  $v \in \Sigma^*$  and  $N$  such that for all  $n \geq N$ ,  $u_n = v$ , then  $\alpha = v$ .
- If there exists  $\beta \in \Sigma^\omega$  such that for all  $j$ ,  $u_n[0, j) = \beta[0, j)$  for all sufficiently large  $n$ , then  $\alpha = \beta$ .
- Otherwise,  $\alpha = \perp$ , which denotes lack of convergence in  $\Sigma^\infty$ .

Under this topology,  $\Sigma^\infty$  is compact. The cylinder sets defined by fixing finitely many letters are both closed and open.

## 2.2 Automata and semigroups for infinite words

We consider infinite words and  $\omega$ -regular languages from the algebraic and combinatorial perspectives. A language  $L \subseteq \Sigma^\omega$  is  $\omega$ -regular if and only if it can be recognised by a (nondeterministic) Büchi automaton  $\mathcal{A} = (\Sigma, Q, I, \Delta, F)$ , where  $\Sigma$  is the alphabet,  $Q$  is the finite set of states,  $I \subseteq Q$  is the set of initial states,  $\Delta \subseteq Q \times \Sigma \times Q$  is the transition relation, and  $F$  is the set of accepting states. A run  $r \in Q^\omega$  of the automaton on an input word  $\alpha$  satisfies  $r(0) \in I$ , and for all  $n$ ,  $(r(n), \alpha(n), r(n+1)) \in \Delta$ . A word  $\alpha$  is accepted if it has a run  $r$  such that  $r(n) \in F$  for infinitely many  $n$ .

Some of our technical tools, e.g., Semënov's theorem, require the automaton to be *deterministic*, i.e., there must be a single initial state, and the transition relation must induce a function  $\delta: Q \times \Sigma \rightarrow Q$ . For this reason, we also use deterministic parity automata, which are further equipped with a function  $\text{index}: Q \rightarrow \mathbb{N}$ . In the case of deterministic parity automata, a word  $\alpha$  has a single run  $r$ , and is accepted if  $\limsup_{n \rightarrow \infty} \text{index}(r(n))$  is even. Deterministic parity automata recognise precisely the class of  $\omega$ -regular languages. For such  $\mathcal{A}$  and  $\alpha \in \Sigma^\omega$ , we write  $\mathcal{A}(\alpha)$  for the sequence of states visited when  $\mathcal{A}$  reads  $\alpha$ .

We now recall  $\omega$ -semigroups (see [21, Sec. 7]) as an equivalent way to recognise  $\omega$ -regular languages. Formally, an  $\omega$ -semigroup  $M = (M_f, M_\omega)$  is a two-sorted algebra equipped with the following operations:

1. A binary operation defined on  $M_f$  and denoted multiplicatively,
2. A mapping  $M_f \times M_\omega \rightarrow M_\omega$ , called the *mixed product*, also denoted multiplicatively,
3. An infinite product  $\pi$  that maps infinite sequences over  $M_f$  to an element of  $M_\omega$ .

These operations must satisfy the following associativity properties:

1.  $M_f$ , equipped with the binary operation, is a semigroup.
2. For every  $m_1, m_2 \in M_f$  and  $m_3 \in M_\omega$ , we have that  $(m_1 m_2) m_3 = m_1 (m_2 m_3)$ .
3. For every sequence  $(m_n)_{n \in \mathbb{N}}$  over  $M_f$ , and every strictly increasing sequence  $(k_n)_{n \in \mathbb{N}}$  of indices, we have that  $\pi(m_0, m_1, m_2, \dots) = \pi(m_0 \cdots m_{k_0}, m_{k_0+1} \cdots m_{k_1}, \dots)$ .
4. For every  $m \in M_f$  and every sequence  $(m_n)_{n \in \mathbb{N}}$  over  $M_f$ , we have that  $m \cdot \pi(m_0, m_1, \dots) = \pi(m, m_0, m_1, \dots)$ .

The last two conditions show that the infinite product  $\pi(m_0, m_1, \dots)$  can also be denoted multiplicatively as  $m_0 m_1 \cdots$ . Indeed, an  $\omega$ -semigroup can be intuited as a semigroup where infinite products are defined. An immediate example is  $\Sigma^\infty$  with word concatenation, where  $M_f = \Sigma^+$ , and  $M_\omega = \Sigma^\omega$ .

Given  $\omega$ -semigroups  $M_1, M_2$  a morphism  $h$  of  $\omega$ -semigroups is a pair  $h_f, h_\omega$  such that  $h_f$  is a semigroup morphism from  $M_{1,f}$  into  $M_{2,f}$ , and  $h_\omega$  is a map from  $M_{1,\omega}$  to  $M_{2,\omega}$  preserving the mixed product and infinite product, i.e., for every  $m_1 \in M_{1,f}$  and  $m_2 \in M_{1,\omega}$ ,  $h_f(m_1) h_\omega(m_2) = h_\omega(m_1 m_2)$ , and for every sequence  $(m_n)_{n \in \mathbb{N}}$  over  $M_{1,f}$ , we have  $h_f(m_0) h_f(m_1) \cdots = h_\omega(m_0 m_1 \cdots)$ . We shall thus omit subscripts and denote the application of a morphism by simply  $h$ . As an immediate example, a non-erasing substitution  $\sigma$  defines a morphism from  $\Sigma^\infty$  to  $\Sigma^\infty$ .

A set  $L \subseteq \Sigma^\infty$  is recognised by a morphism  $h$  from  $\Sigma^\infty$  into an  $\omega$ -semigroup  $M$  if there exists a subset  $H \subseteq M_f \cup M_\omega$  such that  $L = h^{-1}(H)$ . A language  $L \subseteq \Sigma^\omega$  is  $\omega$ -regular if and only if it is recognisable by a finite  $\omega$ -semigroup [21, Thm. 7.6]. See [21, Sections 8,9] respectively for effective translations from Büchi automata to  $\omega$ -semigroups, and vice versa. Furthermore, analogous to the finite-word case, there is a notion of the (finite) syntactic  $\omega$ -semigroup of an  $\omega$ -regular language [21, Sec. 11].

A few remarks addressing presentation concerns are in order. A result of Wilke [28] showed that for finite  $\omega$ -semigroups  $M$ , the infinite product is fully determined by the function  $M_f \rightarrow M_\omega$  that maps  $m$  to  $m^\omega$  [21, Thm. 7.1]. This combinatorial result is proven through Ramsey's theorem, and gives a method to construct  $\omega$ -semigroups through Wilke algebras. We refer the reader to [21, Sec. 7] for details of how the semigroup structure of  $M_f$  determines an extension into an  $\omega$ -semigroup through *linked pairs*.

We use the equivalence of finite  $\omega$ -semigroups and automata through the following lemma.

► **Lemma 2.** *Let  $M = (M_f, M_\omega)$  be a finite  $\omega$ -semigroup, and  $x \in M_f \cup M_\omega$ . Extend  $M_f$  into a monoid  $M'_f$  by adjoining a distinguished neutral element  $1_M$ : for all  $m \in M'_f$ ,  $1_M \cdot m = m \cdot 1_M = m$ . We can construct an automaton  $\mathcal{A}_x$  over the alphabet  $M'_f$  that accepts  $m_0 m_1 \cdots \in (M'_f)^\omega$  if and only if the infinite product  $m_0 m_1 \cdots$  (defined as the possibly finite product of the subsequence  $m_{i_0} m_{i_1} \cdots$  obtained by discarding all  $1_M$  terms) equals  $x$ .*

For example, consider the  $\omega$ -regular language  $(a^*b)^\omega$  over the alphabet  $\{a, b\}$ , i.e., the language of infinite words containing infinitely many occurrences of the letter  $b$ . We recognise it algebraically as follows. Let  $M_f = \{e, m\}$ , where  $e \cdot e = e, e \cdot m = m \cdot e = m \cdot m = m$ . Let  $M_\omega = \{\hat{e}, \hat{m}\}$ , where  $e^\omega = \hat{e}$  and  $m^\omega = \hat{m}$ . We define the morphism  $h$  to map  $a$  to  $e$ , and  $b$  to  $m$ . Observe that a word  $\alpha$  is in the language if and only if  $h(\alpha) = \hat{m}$ .

### 2.3 Uniformly recurrent words

For a word  $\alpha \in \Sigma^\omega$ , the set  $\mathcal{L}(\alpha) = \{u : u \text{ is a factor of } \alpha\}$  is called the (*factor*) *language* of  $\alpha$ . The factor complexity function  $p_\alpha$  computes the number of factors of  $\alpha$  of a given length  $n$ . For example, if  $\alpha$  is Sturmian, then  $p_\alpha(n) = n + 1$  for all  $n$  (see e.g., [13]).

For an infinite word  $\alpha$  and  $l \geq 0$ , denote by  $R_\alpha(l)$  the smallest  $r \in \mathbb{N} \cup \{\infty\}$  such that every factor of  $\alpha$  of length  $l$  is a factor of every factor of  $\alpha$  of length  $r$ . We call  $R_\alpha$  the *recurrence function* of  $\alpha$ . A word  $\alpha \in \Sigma^\omega$  is said to be *uniformly recurrent* if  $R_\alpha(l) \in \mathbb{N}$  for every  $l \in \mathbb{N}$ . That is, every  $u \in \Sigma^l$  either does not occur in  $\alpha$ , or occurs infinitely often with bounded gaps. It is clear from the definition that for uniformly recurrent  $\alpha$ , the value of  $R_\alpha(l)$  only depends on  $\mathcal{L}(\alpha)$ . We record the following, which follows by brute enumeration.

► **Lemma 3.** *Let  $\alpha \in \Sigma^\omega$  be uniformly recurrent. Suppose we have access to an oracle that, given  $u \in \Sigma^*$ , checks whether  $u \in \mathcal{L}(\alpha)$ . Then we can effectively compute  $R_\alpha(l)$ .*

Semënov [24] gave an algorithm<sup>3</sup> for determining whether a given deterministic automaton  $\mathcal{A}$  accepts a given uniformly recurrent word  $\alpha$ , which is represented by (A) an oracle computing  $\alpha(n)$  on input  $n$  and (B) an oracle computing an upper bound  $\bar{R}_\alpha(l)$  on  $R_\alpha(l)$  given  $l$ .

► **Theorem 4 (Semënov).** *Let  $\mathcal{A}$  be a deterministic automaton and  $\alpha$  be a uniformly recurrent word represented by the oracles (A-B). We can effectively compute  $M \in \mathbb{N}$  such that a state  $q$  of  $\mathcal{A}$  appears infinitely often in  $\mathcal{A}(\alpha)$  if and only if it appears in  $\mathcal{A}(\alpha)[M, 2M]$ .*

In other words, to check whether  $\mathcal{A}$  accepts  $\alpha$ , we simply need to run  $\mathcal{A}$  on  $\alpha$  for  $2M$  steps, and observe the states that are visited. See [18, Chap. 3.1] for effective bounds on  $M$ , from which we can deduce the following.

► **Proposition 5.** *Let  $\mathcal{A}$  be a deterministic automaton and  $\alpha$  be a uniformly recurrent word represented by the oracles (A-B). We can compute  $l, M$  such that for any uniformly recurrent  $\beta$  with  $\beta[0, M) = \alpha[0, M)$  and  $R_\beta(n) = R_\alpha(n)$  for all  $n \leq l$ , we have that  $\mathcal{A}$  accepts  $\alpha$  if and only if it accepts  $\beta$ .*

<sup>3</sup> Semënov's result applies to the more general family of *effectively almost-periodic words* [20].

## 2.4 Ostrowski numeration system

We recall properties of a numeration system which proves to be particularly convenient to describe any Sturmian word in  $S$ -adic terms [13]. Let  $\eta \in (0, 1) \setminus \mathbb{Q}$ . The *continued fraction expansion* of  $\eta$  is the unique sequence  $(a_n)_{n \geq 1}$  of positive integers such that

$$\eta = \frac{1}{a_1 + \frac{1}{a_2 + \dots}}$$

We write  $\eta = [0; a_1, a_2, \dots]$ . The *convergents*  $(p_n/q_n)_{n \in \mathbb{N}}$  of  $\eta$  are obtained by truncating the expansion at the  $n$ -th level. The numerators and denominators satisfy the recurrences  $p_0 = 0$ ,  $q_0 = 1$ ,  $p_1 = 1$ ,  $q_1 = a_1$ , and  $(p_{n+2}, q_{n+2}) = a_{n+2} \cdot (p_{n+1}, q_{n+1}) + (p_n, q_n)$  for all  $n \geq 0$ . The convergents are the *locally best approximants* of  $\eta$ : for every  $n \in \mathbb{N}$ ,  $p \in \mathbb{Z}$ , and  $0 < q < q_n$ ,

$$|q_n \eta - p_n| < \min_{p \in \mathbb{N}} |q \eta - p|$$

which implies that  $|\eta - \frac{p_n}{q_n}| < \min_{p \in \mathbb{N}} |\eta - \frac{p}{q}|$ . The *Ostrowski numeration system in base  $\eta$*  is based on the sequence  $\theta_n = q_n \eta - p_n$ . For any  $\chi \in [-\eta, 1 - \eta]$ , there exists a sequence  $(b_n)_{n \geq 1}$  over  $\mathbb{N}$  such that (i)  $0 \leq b_1 < a_1$ , (ii)  $0 \leq b_n \leq a_n$  for all  $n \geq 2$ , (iii) for all  $n$ ,  $b_n = 0$  if  $b_{n+1} = a_{n+1}$ , and

$$\chi = \sum_{n=1}^{\infty} b_n \theta_{n-1}.$$

We refer to  $(b_n)_{n \geq 1}$  as an *Ostrowski expansion of  $\chi$  in base  $\eta$* . Conversely, every  $(b_n)_{n \in \mathbb{N}}$  satisfying (i-iii) is an Ostrowski expansion of some  $\chi$  in base  $\eta$ , i.e., the infinite sum converges to a value in  $[-\eta, 1 - \eta]$ .<sup>4</sup> If  $\chi \notin \mathbb{Z} + \eta\mathbb{Z}$  or if  $\chi \in \mathbb{Z}_{\geq 1} + \eta\mathbb{Z}$ , then  $\chi$  has a unique Ostrowski expansion in base  $\eta$ . Otherwise,  $\chi$  can have two expansions in base  $\eta$ . For more on the subject, see e.g., [9].

### 3 $S$ -adic words

We now establish basic definitions and facts about  $S$ -adicity; see e.g., [4] for more on the subject, with examples being provided in Sec. 3.2.

#### 3.1 Directive and generating sequences

Let  $\Sigma$  be an alphabet and  $S \subseteq S(\Sigma)$  be a possibly infinite set of non-erasing substitutions. We refer to  $\alpha \in \Sigma^* \cup \Sigma^\omega$  as  *$S$ -directed* if there exists a sequence  $(\sigma_n)_{n \in \mathbb{N}}$  over  $S$  and a sequence  $(\alpha^{(n)})_{n \in \mathbb{N}}$  of words over  $\Sigma$  such that  $\alpha^{(0)} = \alpha$  and  $\sigma_n(\alpha^{(n+1)}) = \alpha^{(n)}$  for all  $n \in \mathbb{N}$ ; note that each word in the sequence is the image of the subsequent one. We say that  $(\sigma_n)_{n \in \mathbb{N}}$  *directs*  $\alpha$ . A word  $\alpha \in \Sigma^* \cup \Sigma^\omega$  is called  *$S$ -generated* if there exists a sequence  $(\sigma_n, a_n)_{n \in \mathbb{N}}$  over  $S \times \Sigma$  such that

$$\alpha = \lim_{n \rightarrow \infty} \sigma_0 \cdots \sigma_n(a_n). \tag{1}$$

<sup>4</sup> To check this, observe that  $\theta_n$  alternates between positive and negative;  $\theta_0 = \eta$ ,  $\theta_1 = a_1 \eta - 1$ ;  $\theta_{n+2} = a_{n+2} \theta_{n+1} + \theta_n$ .

We refer to  $(\sigma_n, a_n)_{n \in \mathbb{N}}$  as an  $S$ -adic expansion of  $\alpha$ . For finite or infinite  $\alpha$ , whenever (1) holds we say that  $(\sigma_n, a_n)_{n \in \mathbb{N}}$  generates  $\alpha$ . In both the  $S$ -directed and the  $S$ -generated settings (to which we collectively refer as  $S$ -adic),  $(\sigma_n)_{n \in \mathbb{N}}$  is called a *directive sequence*.

► **Lemma 6.** *If  $\alpha \in \Sigma^* \cup \Sigma^\omega$  is generated by  $s$  over  $S(\Sigma)$ , then it is also directed by  $s$ . Furthermore, every  $s$  over  $S(\Sigma)$  directs at least one non-empty word  $\beta \in \Sigma^+ \cup \Sigma^\omega$ .*

**Proof.** We will first prove a slightly more general version of the first statement. Let  $(a_n)_{n \in \mathbb{N}}$  be a sequence of letters and  $U_m = \{\sigma_m \cdots \sigma_n(a_n) : n \geq m\}$  for all  $m \in \mathbb{N}$ . Suppose  $\alpha$  is an accumulation point of  $U_0$ , which subsumes the case of  $\alpha$  being generated by  $(\sigma_n, a_n)_{n \in \mathbb{N}}$ . We will inductively prove the existence of  $(\alpha^{(m)})_{m \in \mathbb{N}}$  such that  $\alpha^{(0)} = \alpha$ ,  $\alpha^{(m)} = \sigma_m(\alpha^{(m+1)})$ , and  $\alpha^{(m)}$  is an accumulation point of  $U_m$  for all  $m$ . The base case is immediate.

For the inductive step, suppose we have constructed  $\alpha^{(0)}, \dots, \alpha^{(m)}$  with the properties above. Write  $u_{m,n} = \sigma_m \cdots \sigma_n(a_n)$ , and observe that by the induction hypothesis,  $\alpha^{(m)}$  is the limit of some sequence  $(\sigma_m(u_{m+1, n_i}))_{i \in \mathbb{N}}$ . By compactness, the sequence  $(u_{m+1, n_i})_{i \in \mathbb{N}}$  itself has an infinite subsequence  $(u_{m+1, k_j})_{j \in \mathbb{N}}$  that converges. We choose  $\alpha^{(m+1)}$  to be the limit, which is an accumulation point of  $U_{m+1}$ . By the continuity of  $\sigma_m : \Sigma^* \cup \Sigma^\omega \rightarrow \Sigma^* \cup \Sigma^\omega$ ,

$$\lim_{j \rightarrow \infty} \sigma_m(u_{m+1, k_j}) = \sigma_m \left( \lim_{j \rightarrow \infty} u_{m+1, k_j} \right)$$

which implies that  $\alpha^{(m)} = \sigma_m(\alpha^{(m+1)})$ .

To prove the second claim, choose an arbitrary sequence  $(a_n)_{n \in \mathbb{N}}$  of letters and let  $\beta$  be an accumulation point of  $\{\sigma_0 \cdots \sigma_n(a_n) : n \in \mathbb{N}\}$ . Apply the preceding argument. ◀

The converse of the lemma above, however, is not true: take  $s$  to be the sequence of identity morphisms. We next study various special classes of  $S$ -adic words with which we will work.

► **Definition 7.** *A sequence  $(\sigma_n)_{n \in \mathbb{N}}$  over  $S(\Sigma)$  is weakly primitive if for every  $n$  there exists  $m \geq n$  such that  $\sigma_n \cdots \sigma_m$  is positive, i.e., for every  $b, c \in \Sigma$ , the letter  $b$  appears in  $\sigma_n \cdots \sigma_m(c)$ .*

If  $(\sigma_n)_{n \in \mathbb{N}}$  is weakly primitive, then we can compute a sequence  $(k_n)_{n \in \mathbb{N}}$  of increasing integers with  $k_0 = 0$  such that  $\sigma_{k_n} \cdots \sigma_{k_{n+1}-1}$  is positive for all  $n$ . Consequently, for any sequence of letters  $(a_n)_{n \in \mathbb{N}}$  we have that  $\lim_{n \rightarrow \infty} |\sigma_0 \cdots \sigma_n(a_n)| = \infty$ ; this is known as being *everywhere growing*. A word directed by a weakly primitive sequence is uniformly recurrent [11].

The directive sequences with which we will work generate (and, as we will show in Sec. 4, direct) a unique word due to left-properness.

► **Lemma 8.** *Let  $(\sigma_n)_{n \in \mathbb{N}} \in S(\Sigma)^\omega$  be weakly primitive with infinitely many left-proper terms. Then there exists a unique word  $\alpha \in \Sigma^\omega$  such that for any  $(a_n)_{n \in \mathbb{N}} \in \Sigma^\omega$ ,  $(\sigma_n, a_n)_{n \in \mathbb{N}}$  generates  $\alpha$ .*

**Proof.** Let  $(k_n)_{n \in \mathbb{N}}$  be an increasing sequence such that  $k_0 = 0$  and  $\sigma_{k_n}$  is left-proper for  $n > 0$ . Further, let  $b_n \in \Sigma$  be such that all images of  $\sigma_{k_n}$  begin with  $b_n$ . Then  $\alpha = \lim_{n \rightarrow \infty} \sigma_0 \cdots \sigma_{k_n-1}(b_n)$ . This limit exists and is infinite because the terms of the sequence are strict prefixes of one another. Now take  $(\sigma_n, a_n)_{n \in \mathbb{N}}$ , where  $(a_n)_{n \in \mathbb{N}}$  is arbitrary. Then  $\sigma_0 \cdots \sigma_n(a_n)$  must agree with  $\alpha$  on the first  $|\sigma_0 \cdots \sigma_{k_m-1}(a_m)|$  letters, where  $k_m$  is maximal with the property that  $k_m \leq n$ . Since  $k_m$  becomes arbitrarily large as  $n \rightarrow \infty$ , we have that  $\lim_{n \rightarrow \infty} \sigma_0 \cdots \sigma_n(a_n) = \alpha$ . ◀

### 3.2 Sturmian and Arnoux-Rauzy words

We now illustrate the concepts above through examples. Let  $\eta \in (0, 1) \setminus \mathbb{Q}$ . The *characteristic* Sturmian word with slope  $\eta$  is defined by

$$\alpha_\eta(n) = \lfloor (n + 2)\eta \rfloor - \lfloor (n + 1)\eta \rfloor \in \{0, 1\} \tag{2}$$

for  $n \in \mathbb{N}$ . For instance, the Fibonacci word has  $\eta = 1/\phi^2$ , where  $\phi$  is the golden ratio. Characteristic Sturmian words are  $S$ -directed for  $S = \{\lambda_0, \lambda_1\}$  where  $\lambda_i$  maps  $i$  to  $i$  and the other letter  $j$  to  $ij$ , i.e., it inserts  $i$  to the left.

The word  $\alpha_\eta$  is intimately connected to the continued fraction expansion of  $\eta$ . Suppose  $\eta = [0; 1 + a_1, a_2, \dots]$ . We then have that  $\alpha_\eta$  is the unique word directed by the sequence

$$\underbrace{\lambda_0, \dots, \lambda_0}_{a_1 \text{ times}}, \underbrace{\lambda_1, \dots, \lambda_1}_{a_2 \text{ times}}, \underbrace{\lambda_0, \dots, \lambda_0}_{a_3 \text{ times}}, \underbrace{\lambda_1, \dots, \lambda_1}_{a_4 \text{ times}}, \dots$$

which we denote by  $s_\eta$ . For example,  $1/\phi^2$  has the continued fraction expansion  $[0; 2, 1, 1, \dots]$ , and hence the Fibonacci word is directed by  $(\lambda_0\lambda_1)^\omega$ . Observe that  $\lambda_0, \lambda_1$  are left-proper. Moreover, for every  $k, m > 0$ , we have that  $\lambda_0^k\lambda_1^m$  is positive and hence  $s_\eta$  is weakly primitive.

A (general) Sturmian word  $\alpha$  of slope  $\eta \in (0, 1) \setminus \mathbb{Q}$  and intercept  $\chi \in [-\eta, 1 - \eta]$  is given by one of the following:

$$\alpha(n) = \lfloor (n + 2)\eta + \chi \rfloor - \lfloor (n + 1)\eta + \chi \rfloor, \text{ for all } n \tag{3}$$

$$\alpha(n) = \lceil (n + 2)\eta + \chi \rceil - \lceil (n + 1)\eta + \chi \rceil, \text{ for all } n. \tag{4}$$

Sturmian words are uniformly recurrent, and are equivalently characterised by their factor complexity  $p(n) = n + 1$ , which is the lowest among non-periodic words. A Sturmian word  $\alpha$  with slope  $\eta$  and intercept  $\chi$  satisfies  $\mathcal{L}(\alpha) = \mathcal{L}(\alpha_\eta)$ . That is, the language of a Sturmian word only depends on its slope. See for instance [13] for more on the subject.

Sturmian words are not necessarily  $S$ -adic for  $S$  defined above. However, they are  $S$ -adic for  $S = \{\lambda_0, \lambda_1, \rho_0, \rho_1\}$  where the substitution  $\rho_i$  inserts  $i$  to the right, i.e., maps  $i$  to  $i$  and  $j$  to  $ji$ . Let  $\alpha$  be a Sturmian word with slope  $\eta = [0; a_1 + 1, a_2, \dots]$  and intercept  $\chi$ . Then there exists (by [5, Prop. 2.7, also see remark after Thm. 2.10]) an Ostrowski expansion  $(b_n)_{n \in \mathbb{N}}$  of  $\chi$  in base  $\eta$  such that  $\alpha$  is directed by a sequence  $(\tau_n)_{n \in \mathbb{N}}$  where

$$\tau_n = \rho_0^{b_{2n+1}} \lambda_0^{a_{2n+1} - b_{2n+1}} \rho_1^{b_{2n+2}} \lambda_1^{a_{2n+2} - b_{2n+2}}. \tag{5}$$

Conversely, the rules of Ostrowski expansion guarantee that each  $(\tau_n)_{n \in \mathbb{N}}$  obtained from the expansions  $(a_n)_{n \in \mathbb{N}}$  and  $(b_n)_{n \in \mathbb{N}}$  as above is weakly primitive and directs a Sturmian word. Observe that we can unpack every  $\tau_n$  to obtain a *bona fide* directive sequence over  $S = \{\lambda_0, \lambda_1, \rho_0, \rho_1\}$ . We also observe that characteristic Sturmian words, with  $\chi = 0$ , are have directive sequences that use only  $\lambda_0, \lambda_1$ .

In summary, Sturmian words can be characterised as the set of all  $\alpha$  generated by some weakly primitive  $(\sigma_n)_{n \in \mathbb{N}} \in \{\lambda_0, \lambda_1, \rho_0, \rho_1\}$  containing infinitely many left-proper terms. Arnoux-Rauzy words generalise Sturmian words to larger alphabets. Let  $\Sigma = \{0, \dots, d - 1\}$ . For distinct  $i, j \in \Sigma$ , define  $\lambda_i(j) = ij$  and  $\rho_i(j) = ji$ , and for  $i = j \in \Sigma$ , let  $\lambda_i(j) = \rho_i(j) = i$ . Observe that each  $\lambda_i$  is left-proper. Then a word  $\alpha \in \Sigma^\omega$  is Arnoux-Rauzy if and only if it is generated by a weakly primitive  $(\sigma_n)_{n \in \mathbb{N}} \in \{\lambda_0, \dots, \lambda_{d-1}, \rho_0, \dots, \rho_{d-1}\}$  containing infinitely many left-proper terms (see, e.g., [16, Sec. 2.3, Thm. 4.12, Sec. 5]). Arnoux-Rauzy words have factor complexity  $p(n) = (n - 1)d + 1$ , but this is not a characterisation for  $d > 2$ . Observe that Sturmian words are precisely the Arnoux-Rauzy words over a two-letter alphabet. Other  $S$ -adic generalisations of Sturmian words include episturmian words [16] and dendric shifts [3, 14].

#### 4 Congenial expansions for $S$ -adic words

Let  $s$  be a directive sequence over  $S(\Sigma)$  for an alphabet  $\Sigma$ . In this section we will show that any word directed by  $s$  can be written as a product of *congenial* words generated by  $s$ . Congenial expansions are one of the main insights of this paper, and form the combinatorial cornerstone of our analysis of the automaton acceptance problem for  $S$ -adic words.

► **Definition 9.** Let  $\Sigma$  be an alphabet. A sequence  $((\sigma_n, a_n))_{n \in \mathbb{N}}$  over  $S(\Sigma) \times \Sigma$  is *congenial* if  $\sigma_{n+1}(a_{n+1})$  begins with  $a_n$  for all  $n$ . A word  $\alpha \in \Sigma^+ \cup \Sigma^\omega$  is  *$s$ -congenial* for a directive sequence  $s$  if  $\alpha = \lim_{n \rightarrow \infty} \sigma_0 \cdots \sigma_n(a_n)$  for a congenial sequence  $((\sigma_n, a_n))_{n \in \mathbb{N}}$  that augments  $s$ .

The most desirable property of a congenial sequence  $(\sigma_n, a_n)_{n \in \mathbb{N}}$  is that  $\lim_{n \rightarrow \infty} \sigma_0 \cdots \sigma_n(a_n)$  is guaranteed to exist, and has every  $\sigma_0 \cdots \sigma_n(a_n)$  as a prefix. For example, the sequence  $((\lambda_0, 1)(\lambda_1, 0))^\omega$ , where  $\lambda_0, \lambda_1$  are Sturmian substitutions, is congenial. This sequence generates the Fibonacci word.

The following lemma captures this property, and is proven via a straightforward induction.

► **Lemma 10.** Let  $((\sigma_n, a_n))_{n \in \mathbb{N}}$  be congenial, and for  $n \geq 1$ , let  $v_n \in \Sigma^*$  be such that  $\sigma_n(a_n) = a_{n-1}v_n$ . For all  $n$ ,

$$\sigma_0 \cdots \sigma_n(a_n) = \sigma_0(a_0) \cdot \sigma_0(v_1) \cdot \sigma_0\sigma_1(v_2) \cdots (\sigma_0 \cdots \sigma_{n-1}(v_n)).$$

For a directive sequence  $s$ , denote by  $\text{congenials}_s$  the set of all  $s$ -congenial  $\alpha \in \Sigma^\omega$ . We next show that this set is finite.

► **Lemma 11.** Let  $\Sigma$  be an alphabet and  $s = (\sigma_n)_{n \in \mathbb{N}}$  be a directive sequence over  $S(\Sigma)$ . There exist at least one and at most  $|\Sigma|$  congenial expansions of the form  $(\sigma_n, a_n)_{n \in \mathbb{N}}$ , and hence  $1 \leq |\text{congenials}_s| \leq |\Sigma|$ .

**Proof.** By Lem. 6, there exists  $(\alpha^{(n)})_{n \in \mathbb{N}}$  over  $\Sigma^\omega$  such that  $\sigma_0 \cdots \sigma_n(\alpha^{(n+1)}) = \alpha^{(0)}$  for all  $n$ . Let  $a_n$  be the first letter of  $\alpha^{(n)}$ . We have that  $(\sigma_n, a_{n+1})$  is congenial, and hence  $|\text{congenials}_s| \geq 1$ .

Now suppose there exist  $m \geq |\Sigma| + 1$  congenial sequences  $(\sigma_n, a_n^{(i)})_{n \in \mathbb{N}}$ . By a pigeonhole argument, there must exist  $i \neq j$  such that  $a_n^{(i)} = a_n^{(j)}$  for infinitely many  $n$ . From congeniality it follows that  $a_n^{(i)} = a_n^{(j)}$  for all  $n$ . ◀

Congenial words constitute the building blocks of directed words. The “if” part of the following lemma follows by definition; the “only if” part holds because a directed word can naturally be factorised into a congenial prefix and a directed suffix (if the former is finite).

► **Lemma 12.** Let  $\Sigma$  be an alphabet and  $s = (\sigma_n)_{n \in \mathbb{N}}$  be a directive sequence over  $S(\Sigma)$ . A word  $\alpha \in \Sigma^+ \cup \Sigma^\omega$  is directed by  $s$  if and only if it can be expressed as a (possibly infinite) concatenation  $u_0u_1 \cdots$  of  $s$ -congenial words.

**Proof.** Suppose  $\alpha = u_0u_1 \cdots$ , where  $u_i \in \text{congenials}_s$  for all  $i$ . Let  $(\sigma_n, a_n^{(i)})_{n \in \mathbb{N}}$  be a congenial sequence generating  $u_i$ , and  $u_i^{(n)} = \lim_{k \rightarrow \infty} \sigma_n \cdots \sigma_k(a_k^{(i)})$  for all  $i$ . We have that  $u_n^{(i)} = \sigma_n(u_{n+1}^{(i)})$  for all  $n$ . It remains to define  $\alpha^{(n)} = u_0^{(n)}u_1^{(n)} \cdots$ . Then  $\alpha^{(0)} = \alpha$  and  $\alpha_n^{(i)} = \sigma_n(\alpha_{n+1}^{(i)})$  for all  $n$ .

Now suppose  $\alpha$  is  $s$ -directed, and let  $(\alpha^{(n)})_{n \in \mathbb{N}}$  be the witnessing sequence of words with  $\alpha^{(0)} = \alpha$ . Write  $a_n$  for the first letter of  $\alpha^{(n)}$ . We construct the desired factorisation inductively. Let  $v$  be the word defined by the congenial sequence  $(\sigma_n, a_{n+1})_{n \in \mathbb{N}}$ . By the choice of  $(a_n)_{n \in \mathbb{N}}$ ,  $v$  is a prefix of  $\alpha$ . If  $v = \alpha$ , then we are done. Otherwise,  $v$  must be finite.

Let  $(v_n)_{n \in \mathbb{N}}$  be the unique sequence of finite words such that  $v_0 = v$ ,  $v_n$  is a prefix of  $\alpha^{(n)}$  for all  $n$ , and  $\sigma_n(v_{n+1}) = v_n$  for all  $n$ . Write  $\alpha^{(n)} = v_n \gamma^{(n)}$  for all  $n$ . Because  $\sigma_n(\alpha^{(n+1)}) = \alpha^{(n)}$  for all  $n$  and  $\sigma_n(v_{n+1}) = v_n$ , we have that  $\sigma_n(\gamma^{(n+1)}) = \gamma^{(n)}$ . That is,  $\gamma = \gamma^{(0)}$  is a suffix of  $\alpha$  directed by  $s$ . Set  $u_0 = v$ , and repeat the process on  $\gamma$ . ◀

If  $s$  is weakly primitive, then we have the following strengthening of Lem. 12.

► **Lemma 13.** *Suppose  $s = (\sigma_n)_{n \in \mathbb{N}}$  is weakly primitive and directs  $\alpha \in \Sigma^\omega$ . Then  $\alpha$  has a congenial expansion  $(\sigma_n, a_n)_{n \in \mathbb{N}}$ .*

**Proof.** Let  $(\alpha^{(n)})_{n \in \mathbb{N}}$  be such that  $\alpha^{(0)} = \alpha$  and  $\sigma_n(\alpha^{(n+1)}) = \alpha^{(n)}$ , and  $a_n$  be the first letter of  $\alpha^{(n+1)}$ . Then  $(\sigma_n, a_n)_{n \in \mathbb{N}}$  is congenial,  $\sigma_0 \cdots \sigma_n(a_n)$  is a prefix of  $\alpha$  for all  $n$ , and  $\lim_{n \rightarrow \infty} |\sigma_0 \cdots \sigma_n| = \infty$  by the growth assumption. Therefore,  $(\sigma_n, a_n)_{n \in \mathbb{N}}$  generates  $\alpha$ . ◀

Combining lemmas 6 and 13 we obtain the following.

► **Lemma 14.** *Let  $s$  be a weakly primitive directive sequence. A word  $\alpha$  is directed by  $s$  if and only if it is  $s$ -congenial.*

## 5 Equivalence of substitutions modulo a semigroup

The motivation behind the semigroup-based approach to the language-membership problem is that even though there are infinitely many substitutions in  $S(\Sigma)$ , from the perspective of a finite  $\omega$ -semigroup  $M_L$  recognising an  $\omega$ -regular language  $L \subseteq \Sigma^\omega$ , they can be divided into finitely many equivalence classes.

Let  $L \subseteq \Sigma^\omega$  be an  $\omega$ -regular language, and let it be recognised by a morphism  $h_L$  into a finite  $\omega$ -semigroup  $M_L = (M_{L,f}, M_{L,\omega})$ , e.g., the morphism into the syntactic  $\omega$ -semigroup as defined in [21, Sec. 11]. Observe that an  $\omega$ -semigroup morphism from  $\Sigma^\infty$  into  $M$  is completely determined by the images of each letter in  $M_{L,f}$ . Since  $M_{L,f}$  is finite, there are only finitely many possible morphisms. We denote the set of these morphisms by  $\text{morphisms}_L$ . We make a small technical adaptation, and interpret these as monoid morphisms, i.e., we adjoin a fresh neutral element  $1_{M_L}$  to  $M_L$ , and assign  $h(\varepsilon) = 1_{M_L}$  for each  $h$ .

We define an equivalence relation on the set of non-erasing substitutions  $\sigma: \Sigma^* \rightarrow \Sigma^*$ . Let  $\text{segments}_\sigma$  be the function that takes a letter  $a \in \Sigma$  and  $h \in \text{morphisms}_L$ , and returns a finite sequence of pairs from  $\Sigma \times M_L$ , determined as follows. Write

$$\sigma(a) = b_1 v_1 \cdots b_d v_d$$

where  $b_1, \dots, b_d$  are distinct letters and  $v_i \in \{b_1, \dots, b_i\}^*$  for all  $1 \leq i \leq d$ . I.e., we consider the factorisation of  $\sigma(a)$  into segments based on the first occurrence of each letter. Then

$$\text{segments}_\sigma(a, h) = \langle (b_1, h(v_1)), \dots, (b_d, h(v_d)) \rangle. \quad (6)$$

Note that there are only finitely many possibilities for  $\text{segments}_\sigma$ . For  $\sigma, \mu \in S(\Sigma)$ , define

$$\sigma \equiv_L \mu \Leftrightarrow \text{segments}_\sigma = \text{segments}_\mu.$$

We denote the class of  $\sigma$  by  $[\sigma]_L$ , and the finite set of the equivalence classes by  $\Xi_L$ . We next show how to effectively provide representatives for the equivalence classes.

► **Lemma 15.** *Given an  $\omega$ -regular language  $L$ , we can compute morphisms  $\sigma_1, \dots, \sigma_m$  such that  $[\sigma_i]_L \neq [\sigma_j]_L$  for all  $i \neq j$  and  $\Xi_L = \{[\sigma_i]_L : 1 \leq i \leq m\}$ .*

**Proof.** We simply iterate over each syntactic possibility for  $\text{segments}_\sigma$  (as in Eq. (6)), and check if it is realised by a non-erasing substitution. In order to do so, for each letter  $a$ , we will find a word  $w_a$  such that assigning  $\sigma(a) = w_a$  is consistent with  $\text{segments}_\sigma(a, h)$  for all  $h$ . For  $a \in \Sigma$  and  $h \in \text{morphisms}_L$  under consideration, let the purported  $\text{segments}_\sigma(a, h) = \langle (b_1, x_1), \dots, (b_k, x_k) \rangle$ , with  $k \geq 1$  such that  $b_i \neq b_j$  for all  $i \neq j$ , and  $x_i \in M_L$ ,  $b_i \in \Sigma$  for all  $i$ . We can compute regular languages  $L_1, \dots, L_k \subseteq \Sigma^*$  such that for all  $i$  and  $w \in \Sigma^*$ ,  $w \in L_i$  if and only if  $h(w_i) = x_i$  and  $w_i \in \{b_1, \dots, b_i\}^*$ . Let  $L_{a,h}$  denote  $b_1 L_1 \cdots b_k L_k$ .

We can effectively check whether  $L_a = \bigcap_{h \in \text{morphisms}_L} L_{a,h}$  is non-empty, and if yes, effectively compute  $w_a \in L_a$ . Such a word can be computed as an image for every letter (if and) only if the purported  $\text{segments}_\sigma$  is indeed realisable by assigning each  $\sigma(a)$  to the corresponding  $w_a$ . ◀

For technical convenience, we define the following auxiliary functions that can be derived from  $\text{segments}_\sigma$ ; the first four of them are independent of  $L$ .

- (a)  $\text{expanding}_\sigma$  records for each letter  $a$  whether  $|\sigma(a)| > 1$ . It evaluates to false if and only if  $\text{segments}_\sigma(a, h) = \langle (b, 1_{M_L}) \rangle$  for all  $h$ .
- (b)  $\text{introduces}_\sigma$  maps each letter to a finite sequence of pairs of letters with Boolean flags: if  $\sigma(a) = b_1 v_1 \cdots b_d v_d$  when factorised as in the definition of  $\text{segments}_\sigma$ , then  $\text{introduces}_\sigma(a)$  is  $\langle (b_1, f_1), \dots, (b_d, f_d) \rangle$  where  $f_i = 1$  if and only if  $v_i \neq \varepsilon$ .
- (c)  $\text{head}_\sigma$  maps each letter  $a$  to the first letter in  $\sigma(a)$ .
- (d)  $\text{tail}_\sigma$  takes as input a letter  $a$  and a morphism  $h$ . Write  $\sigma(a) = \text{head}_\sigma(a) \cdot v$ . We define  $\text{tail}_\sigma(a, h) = h(v)$ .
- (e)  $\text{compose}_\sigma : \text{morphisms}_L \rightarrow \text{morphisms}_L$  takes  $h$  and returns  $h \circ \sigma$ .

We next argue that composition of morphisms can be defined on equivalence classes. Let  $\sigma, \mu : \Sigma^* \rightarrow \Sigma^*$  be non-erasing. We show how to determine  $\text{segments}_{\sigma \circ \mu}(a, h)$  using only  $\text{segments}_\mu$  and  $\text{segments}_\sigma$ . Suppose  $\mu(a) = b_1 v_1 \cdots b_m v_m$ , where  $b_i \in \Sigma$  and  $v_i \in \{b_1, \dots, b_i\}^*$  for all  $i$ . Then

$$\sigma(\mu(a)) = \sigma(b_1)\sigma(v_1) \cdots \sigma(b_m)\sigma(v_m)$$

and each letter of  $\sigma(v_i)$  will have already appeared in one of  $\sigma(b_1), \dots, \sigma(b_i)$ . Write  $\text{segments}_\sigma(b_i, h) = \langle (c_{i,1}, h(w_{i,1})), \dots, (c_{i,k_i}, h(w_{i,k_i})) \rangle$ . Then  $\sigma(\mu(a))$  is equal to

$$c_{1,1} w_{1,1} \cdots c_{1,k_1} w_{1,k_1} \sigma(v_1) \cdots c_{m,1} w_{m,1} \cdots c_{m,k_m} w_{m,k_m} \sigma(v_m).$$

Write  $t_i = w_{i,k_i} \sigma(v_i)$ . Observe that each letter of  $w_{i,j}$  appears as a factor  $c_{e,l}$  before  $w_{i,j}$  in the factorisation above; the same applies to every  $t_i$ . To compute  $\text{segments}_{\sigma \circ \mu}(a, h)$  we begin with the finite sequence

$$(c_{1,1}, h(w_{1,1})), \dots, (c_{1,k_1-1}, h(w_{1,k_1-1})), (c_{1,k_1}, h(t_1)) \cdots \\ (c_{m,1}, h(w_{m,1})), \dots, (c_{m,k_m-1}, h(w_{m,k_m-1})), (c_{m,k_m}, h(t_m)).$$

Note that the above can be effectively computed from  $\text{segments}_\sigma$  and  $\text{segments}_\mu$  as  $h(t_i) = h(w_{i,k_i}) h(\sigma(v_i))$  and the two factors can be gleaned from, respectively,  $\text{segments}_\sigma(b_i, h)$  and  $\text{segments}_\mu(a, \text{compose}_\sigma(h))$ . Rename the indices in the sequence above to obtain  $\langle (c_1, h(w_1)), \dots, (c_M, h(w_M)) \rangle$  where  $c_i \in \Sigma$  and  $w_i \in \Sigma^*$  for all  $i$ . Recall that  $w_i \in \{c_1, \dots, c_i\}^*$  for all  $i$ . But it is possible that  $c_i = c_j$  for some  $i, j$ . To eliminate these, we repeat the following process for as long as possible. Find the smallest  $j$  such that  $c_i = c_j$  for some  $i < j$ . Replace the two consecutive terms  $(c_{j-1}, h(w_{j-1})), (c_j, h(w_j))$  with  $(c_{j-1}, h(w_{j-1}) h(c_j) h(w_j))$ . In the end we are left with  $\text{segments}_{\sigma \circ \mu}(a, h)$ .

Observe that  $\text{segments}_\sigma$  only depends on  $\xi := [\sigma]_L$ ; we can thus index the auxiliary functions (a-e) above by the equivalence class  $\xi$ . To summarise, we have the following.

► **Proposition 16.** *The set  $\Xi_L$  is a finite monoid with the binary operation  $[\sigma]_L \cdot [\mu]_L = [\sigma \circ \mu]_L$  and the identity element  $[\text{id}]_L$ , where  $\text{id}(w) = w$  for all  $w \in \Sigma^*$ .*

**Proof.** The set of substitutions is a monoid with composition being the binary operation. The map from substitutions to their equivalence classes respects the binary operation, by construction. Using this fact, it is straightforward to check that the binary operation on equivalence classes is associative, and that  $[\text{id}]_L$  is indeed the identity element. ◀

Let  $(\sigma_n)_{n \in I}$  be a sequence over  $S(\Sigma)$ ,  $\xi_n = [\sigma_n]_L$  for all  $n$ , and  $(a_n)_{n \in I}$  be a sequence of letters from  $\Sigma$ , where  $I$  can be finite or infinite. We define  $\text{trace}_L((\sigma_n)_{n \in I}) = (\xi_n)_{n \in I}$  and  $\text{trace}_L((\sigma_n, a_n)_{n \in I}) = (\xi_n, a_n)_{n \in I}$ . We extend the definition of  $[\cdot]_L$  to finite sequences in the natural way: if  $I$  is finite, then  $[(\sigma_n)_{n \in I}]_L$  is the product  $\prod_{n \in I} \xi_n \in \Xi_L$ . The next result shows that for any infinite  $\alpha$  and substitution  $\sigma$ , whether  $\sigma(\alpha) \in L$  can be determined from  $[\sigma]_L$ .

► **Lemma 17.** *Let  $L \subseteq \Sigma^\omega$  be an  $\omega$ -regular language and  $\alpha \in \Sigma^\omega$ . There exists  $\Phi \subseteq \Xi_L$  such that for any  $\sigma: \Sigma^* \rightarrow \Sigma^*$ ,  $\sigma(\alpha) \in L$  if and only if  $[\sigma]_L \in \Phi$ . Furthermore,  $\Phi$  can be effectively computed if we can compute  $h(\alpha)$  for all  $h \in \text{morphisms}_L$ .*

**Proof.** Recall [21, Thm. 7.6] that we have a set  $G \subseteq M_{L,\omega}$  such that for all  $\alpha \in \Sigma^\omega$ ,  $\alpha \in L$  if and only if  $h_L(\alpha) \in G$ . We now view the substitution  $\sigma$  as an  $\omega$ -semigroup morphism:

$$h_L(\sigma(\alpha)) = h_L \circ \sigma(\alpha) = \text{compose}_\sigma(h_L)(\alpha).$$

We thus define  $\Phi = \{\xi \in \Xi_L : \text{compose}_\xi(h_L)(\alpha) \in G\}$ . The effectiveness claim follows from the fact that  $\text{compose}_\xi(h_L) \in \text{morphisms}_L$  for all  $\xi$ . ◀

## 6 Morphic words in a given $\omega$ -regular language

We shall now use the machinery developed in Sec. 5 to characterise the set of morphic words in a given  $\omega$ -regular language  $L \subseteq \Sigma^\omega$ , thus generalising the main result of [10]. A word  $\alpha \in \Sigma^\omega$  is *substitutive* if it is a fixed point of a non-trivial substitution, i.e., if there exists  $\sigma \in S(\Sigma)$  not the identity such that  $\sigma(\alpha) = \alpha$ . A word  $\alpha \in \Sigma^\omega$  is *morphic* if it is of the form  $\pi(\beta)$  for  $\pi \in S(\Sigma)$  and  $\beta$  a substitutive word. For  $u \in \Sigma^*$  and substitutions  $\sigma, \pi$ , define  $\sigma^\omega(u) = \lim_{n \rightarrow \infty} \sigma^n(u)$  and  $\pi \circ \sigma^\omega(u) = \pi(\sigma^\omega(u))$ . We will show that, for an  $\omega$ -regular language  $L$  and substitutions  $\sigma, \pi$ , the respective equivalence classes  $\xi, \zeta \in \Xi_L$  of substitutions  $\sigma, \pi$  determine:

- Whether the substitution  $\sigma$  has a fixed point in  $L$  (Thm. 22).
- A regular language  $\mathcal{L}(\xi, \zeta) \subseteq \Sigma^+$  such that  $\pi \circ \sigma^\omega(u)$  is in  $L$  if and only if  $u \in \mathcal{L}(\xi, \zeta)$  (Thm. 21).

Before we proceed to state our key properties, we remark that we work with  $\Sigma^\infty \cup \{\perp\}$ . For the sake of brevity, we (syntactically) allow infinite words to be concatenated, i.e., for  $\alpha \in \Sigma^\omega$  and  $\beta \in \Sigma^\infty$ , the syntactic concatenation  $\alpha\beta$  equals  $\alpha$ . We furthermore allow  $\perp$  to be concatenated: for any  $u$ ,  $\perp u = \perp$ ,  $u\perp = \perp$  if  $u$  is finite, and  $u\perp = u$  if  $u$  is infinite.

► **Lemma 18.** *Let  $\Sigma$  be an alphabet and  $\sigma$  be a substitution. The following properties hold.*

**Saturation** For  $u \in \Sigma^*$ , we have that  $\sigma^\omega(u) = \sigma^\omega(\sigma(u))$ .

**Distributivity** For  $u, v \in \Sigma^*$ ,  $\sigma^\omega(uv) = \sigma^\omega(u) \cdot \sigma^\omega(v)$ .

**Left-Expansion** For  $a \in \Sigma, u \in \Sigma^+, v \in \Sigma^*$ ,  
if  $\sigma^\omega(a) = u \cdot \sigma^\omega(a) \cdot \sigma^\omega(v) \neq \perp$  then  $\sigma^\omega(a) = u^\omega$ .

**Right-Expansion** For  $a \in \Sigma, u \in \Sigma^*$ , if  $\sigma(a) = au$ , then  
 $\sigma^\omega(a) = a \cdot u \cdot \sigma(u) \cdots \sigma^n(u) \cdots$ .

**Cycle of Contradiction** Let  $a_0, \dots, a_{p-1}, p > 1$  be distinct letters such that  $\text{head}_{\sigma^p}(a_0) = a_0$ ,  
and for any  $r \in \{1, \dots, p-1\}$ ,  $\text{head}_{\sigma^r}(a_0) = a_r$ . We have that  
 $\sigma^\omega(a_0) = \cdots = \sigma^\omega(a_{p-1}) = \perp$ .

**Terminal Letters** If, for a letter  $a$ ,  $\sigma^\omega(a) = u \in \Sigma^+$ , then for all  $n \geq |\Sigma|$ ,  $\sigma^n(a) = u$ .

**Proof. Saturation.** The sequences  $(\sigma^n(u))_n$  and  $(\sigma^{n+1}(u))_{n+1}$  have the same limit.

**Distributivity.** We have that for all  $n$ ,  $\sigma^n(uv) = \sigma^n(u)\sigma^n(v)$ . If  $\sigma^\omega(u) = \mu \in \Sigma^*$ , then for all large  $n$ ,  $\sigma^n(uv) = \mu \cdot \sigma^n(v)$ . Taking the limit, we get  $\sigma^\omega(uv) = \mu \cdot \sigma^\omega(v)$ . If  $\sigma^\omega(u) \in \Sigma^\omega$ , then for every position  $j$ , there exists  $N$  such that for all  $n \geq N$ ,  $|\sigma^n(u)| > j$ . In other words, every position  $j$  is eventually part of  $\sigma^n(u)$ . Thus,  $\sigma^\omega(uv) = \sigma^\omega(u)$ . However, if  $\beta \in \Sigma^\omega$ , then  $\beta\mu = \beta$  for all  $\mu \in \Sigma^\infty$ . If  $\sigma^\omega(u) = \perp$ , there is some position  $j$  such that the letter  $\sigma^n(u)(j)$  fluctuates with  $n$ . This means that the limit of  $\sigma^n(u)\sigma^n(v)$  must also be mapped to  $\perp$ .

**Left-Expansion.** Follows by repeatedly unrolling the equality.

**Right-Expansion.** Follows by repeatedly applying  $\sigma$ .

**Cycle of Contradiction.** The limit must be  $\perp$  as the first letter of  $\sigma^n(a_r)$  keeps alternating between  $a_0, \dots, a_{p-1}$ .

**Terminal Letters.** We find the set  $A$  of letters  $a$  such that  $\sigma^\omega(a) = u \in \Sigma^+$  by saturation. The key idea is that if  $\sigma^\omega(a)$  converges within  $n$  iterations, then  $\sigma^\omega(\sigma(a))$  must converge within  $n-1$  iterations, i.e., for every letter  $b$  in  $\sigma(a)$ ,  $\sigma^\omega(b)$  must converge within  $n-1$  iterations.

We start with the set  $A_0$  of letters  $a_0$  such that  $\sigma(a_0) = a_0$ . We construct  $A_{n+1}$  as the union of  $A_n$  with the set of letters  $a_{n+1}$  such that  $\sigma(a_{n+1})$  only contains letters from  $A_n$ . This construction will saturate within  $|\Sigma|$  steps. The invariant is that  $A_j$  is the set of  $a$  such that  $\sigma^\omega(a)$  converges within  $j$  iterations. We conclude that since for  $n \geq |\Sigma|$ ,  $A_n = A_{|\Sigma|}$ , if  $\sigma^\omega(a)$  converges in  $n$  steps then it must have already converged within  $|\Sigma|$  steps. ◀

The following result, along with distributivity, implies that whether  $\sigma^\omega(u)$  for a finite word  $u$  is an infinite word accepted by  $\mathcal{A}$  is determined by the equivalence class  $[\sigma]_L$ ; this is the main technical novelty of this section.

► **Theorem 19.** Let  $L \subseteq \Sigma^\omega$  be an  $\omega$ -regular language recognised by an  $\omega$ -semigroup  $M_L = (M_{L,f}, M_{L,\omega})$ . For any  $h \in \text{morphisms}_L$ , substitutions  $\sigma, \tau \in S(\Sigma)$  with  $\sigma \equiv_L \tau$ , and  $a \in \Sigma$ , we have that  $h \circ \sigma^\omega(a) = h \circ \tau^\omega(a)$ . Moreover,  $h \circ \sigma^\omega(a)$  can be computed given only the equivalence class  $\xi \in \Xi_L$  of  $\sigma, \tau$  along with the values of  $h, a$ .

**Proof.** By Lem. 18, we have a dynamic programming algorithm to compute  $h \circ \sigma^\omega(a)$  in a “depth-first” manner. We shall show that this algorithm only uses  $[\sigma]_L = \xi$ .

Consider  $b \in \Sigma$  and write  $\sigma(b) = c_1 v_1 \cdots c_k v_k$ , where the factorisation is based on the first occurrence of each letter: we get  $c_1, \dots, c_k$  from  $\text{introduces}_\xi$ . By Saturation, we have that

$$h \circ \sigma^\omega(b) = h \circ \sigma^\omega(c_1) \cdot h \circ \sigma^\omega(v_1) \cdots h \circ \sigma^\omega(c_k) \cdot h \circ \sigma^\omega(v_k).$$

We observe that the terms  $h \circ \sigma^\omega(v_i)$  will affect the result only if all  $h \circ \sigma^\omega(c_j)$  for  $j \leq i$  are elements of  $M_{L,f}$ . In this case,  $\sigma^\omega(v_i)$  will also be a finite word, and by the Terminal Letters

property, be  $\sigma^{|\Sigma|}(v_i)$ . We denote  $h \circ \sigma^{|\Sigma|}$  by  $g$ ; thus in this case, we have  $h \circ \sigma^\omega(v_i) = g(v_i)$ . We can rewrite

$$h \circ \sigma^\omega(b) = h \circ \sigma^\omega(c_1) \cdot g(v_1) \cdots h \circ \sigma^\omega(c_k) \cdot g(v_k). \quad (7)$$

The elements  $g(v_1), \dots, g(v_k)$  can be obtained from  $\text{segments}_\xi(b, g)$ . If  $c_1 = b$ , we are in the simpler case of Right-Expansion where  $\sigma(b) = bu$ , and have that

$$h \circ \sigma^\omega(b) = h(b) \cdot h(u) \cdot h \circ \sigma(u) \cdot h \circ \sigma^2(u) \cdots$$

The first two terms are easily obtained through  $\text{tail}_\xi$ , and  $h \circ \sigma^n(u) = \text{tail}_\xi(b, \text{compose}_{\xi^n}(h))$ . Since the sequence of monoid elements  $\xi^n$  is effectively ultimately periodic, so is the sequence of factors in the above infinite product, allowing us to compute it.

If  $c_1 \neq b$ , we first check that there is no  $r$  such that  $\sigma^r(c_1) = b$ : this can be done with access to  $\text{head}_{\xi^r}$  for  $r \leq |\Sigma|$ . If this check fails, we have a Cycle of Contradiction, and have that  $h \circ \sigma^\omega(b) = \perp$ .

We now evaluate expansion (7). Write  $m_0 = 1_{M_L}$ , and

$$m_i = h \circ \sigma^\omega(c_1) \cdot g(v_1) \cdots h \circ \sigma^\omega(c_k) \cdot g(v_i).$$

Clearly,  $m_{i+1} = m_i \cdot h \circ \sigma^\omega(c_{i+1}) \cdot g(v_{i+1})$ , and

$$h \circ \sigma^\omega(b) = m_i \cdot h \circ \sigma^\omega(c_{i+1}) \cdot g(v_{i+1}) \cdots h \circ \sigma^\omega(c_k) \cdot g(v_k).$$

In particular when  $m_i \notin M_{L,f}$  for some  $i$ , then  $h \circ \sigma^\omega(b) = m_i$ . For each  $i$ , if  $c_i \neq b$ , we evaluate  $h \circ \sigma^\omega(c_i)$  (we make a recursive call if it has not been evaluated before, otherwise we look up the memoized value). Otherwise,  $c_i = b$  and we are in the case of Left-Expansion, and get that  $h \circ \sigma^\omega(b) = m_{i-1}^\omega$ . In any case, we will eventually compute  $m_k = h \circ \sigma^\omega(b)$ .

The result follows by applying the above depth-first routine to compute  $h \circ \sigma^\omega(a)$ . ◀

As a corollary, we obtain that whether  $\pi(\sigma^\omega(u))$  belongs to an  $\omega$ -regular language  $L$  is completely determined by the equivalence classes  $[\sigma]_L$  and  $[\pi]_L$ : indeed, acceptance only depends on  $h_L \circ \pi \circ \sigma^\omega(u)$ , which is the same as  $h := \text{compose}_{[\pi]_L}(h_L)$  applied to  $\sigma^\omega(u)$ .

► **Corollary 20.** *Let  $L \subseteq \Sigma^\omega$  be an  $\omega$ -regular language. For any letter  $a$ ,  $ah \in \text{morphisms}_L$ , and substitutions  $\sigma, \pi$ , we can compute  $h \circ \pi \circ \sigma^\omega(a)$  given only  $h, a, [\sigma]_L, [\pi]_L$ .*

We thus arrive at the following.

► **Theorem 21.** *Let  $L \subseteq \Sigma^\omega$  be an  $\omega$ -regular language and  $\sigma, \pi$  be substitutions with respective equivalence classes  $\xi, \zeta \in \Xi_L$ . We can compute a regular language  $\mathcal{L}(\xi, \zeta) \subseteq \Sigma^+$  such that  $\pi \circ \sigma^\omega(u) \in L$  if and only if  $u \in \mathcal{L}(\xi, \zeta)$ .*

**Proof.** We construct a deterministic finite-word automaton recognising  $\mathcal{L}(\xi, \zeta)$ . The set of states is the set of elements of  $M_{L,f} \cup M_{L,\omega}$ , the initial state is  $1_{M_L}$ , the set of accepting states is  $G$  (the semigroup elements whose preimages comprise  $L$ ), and the transition function maps  $(m, a)$  to  $m \cdot (h_L \circ \pi \circ \sigma^\omega)(a)$ . The latter is effective by Cor. 20. ◀

Finally, whether  $\sigma$  has a fixed point in  $L$  is effectively determined by  $[\sigma]_L$ .

► **Theorem 22.** *Let  $L \subseteq \Sigma^\omega$  be an  $\omega$ -regular language and  $\sigma$  be a substitution. Given  $[\sigma]_L$  and  $h \in \text{morphisms}_L$ , we can compute the set  $\{h(\alpha) : \sigma(\alpha) = \alpha\} \subseteq M_L$ .*

**Proof.** We observe that any fixed point of  $\sigma$  must be a concatenation of words  $\sigma^\omega(a)$ , where  $\sigma(a) = au$  for  $u \in \Sigma^*$ . Write  $A$  for the set of all such letters  $a$ , which can be extracted from  $\text{introduces}_\sigma$ . The required set is then the sub- $\omega$ -semigroup of  $M_L$  generated by  $\{h \circ \sigma^\omega(a) : a \in A\}$ , whose elements in turn can be computed using Thm. 19. ◀

## 7

 $S$ -adic words in a given  $\omega$ -regular language

We now present our main results, i.e., solutions to the language membership problem for  $S$ -adic words. In this section, let  $L \subseteq \Sigma^\omega$  be an  $\omega$ -regular language recognised by a morphism  $h_L$  into a finite  $\omega$ -semigroup  $M_L = (M_{L,f}, M_{L,\omega})$ , and  $H \subseteq M_{L,\omega}$  be such that  $h_L^{-1}(H) = L$ . For convenience, further let  $\mathcal{A}$  be a deterministic automaton accepting  $L$ . Recall the set  $\Xi_L$  of equivalence classes of substitutions modulo  $M_L$ , defined in Sec. 5. We will first prove that for congenial  $\hat{s}$  over  $S(\Sigma) \times \Sigma$ , defining a word accepted by  $\mathcal{A}$  is a property of  $\text{trace}_L(\hat{s})$ , and the set of all such traces is  $\omega$ -regular.

► **Theorem 23** (Main Result for Generated Words). *Let  $L \subseteq \Sigma^\omega$  be an  $\omega$ -regular language. We can compute an automaton  $\mathcal{B}$  over  $\Xi_L \times \Sigma$  such that for all infinite sequences  $\hat{s}$  over  $S(\Sigma) \times \Sigma$ ,  $\mathcal{B}$  accepts  $\text{trace}_L(\hat{s})$  if and only if  $\hat{s}$  is congenial and generates a word in  $L$ .*

**Proof.** The automaton  $\mathcal{B}$  accepts  $\hat{s} \in (\Xi_L \times \Sigma)^\omega$  if and only if  $\hat{s}$  is accepted by  $\mathcal{B}_x$  (constructed in Lem. 24 below) for some  $x \in H$ . Intuitively,  $\mathcal{B}_x$ , upon reading  $\hat{s}$ , checks that the expansion is congenial. If yes, then the generated word can be factorised as  $u_0 u_1 \cdots$  as described in Lem. 12. The automaton then evaluates the product  $h_L(u_0 u_1 \cdots) = h_L(u_0) h_L(u_1) \cdots$ . ◀

► **Lemma 24.** *Let  $x \in M_{L,f} \cup M_{L,\omega}$ . We can construct an automaton  $\mathcal{B}_x$  over  $\Xi_L \times \Sigma$  such that for all infinite sequences  $\hat{s}$  over  $S(\Sigma) \times \Sigma$ ,  $\mathcal{B}_x$  accepts  $\text{trace}_L(\hat{s})$  if and only if  $\hat{s}$  is congenial and  $h_L(\alpha) = x$ , where  $\alpha$  is the word generated by  $\hat{s}$ .*

**Proof.** Recall that a sequence  $s = (\sigma_n, a_n)_{n \in \mathbb{N}}$  over  $\Xi_L \times \Sigma$  is congenial if and only if  $\text{head}_{\xi_n}(a_n) = a_{n-1}$  for all  $n \geq 1$ , where  $\xi_n = [\sigma_n]_L$ . This property depends only on  $\text{trace}_L(s)$ . The automaton  $\mathcal{B}_x$ , first and foremost, checks the condition above for all  $n \geq 1$ , and permanently transitions into a rejecting state if it observes a violating triple  $\xi_n, a_n, a_{n-1}$ .

Now suppose  $s = (\sigma_n, a_n)_{n \in \mathbb{N}}$  is congenial, and define  $\xi_n$  as above. As shown in Lem. 10,

$$\sigma_0 \cdots \sigma_n(a_n) = \sigma_0(a_0) \cdot \sigma_0(v_1) \cdot \sigma_0 \sigma_1(v_2) \cdots (\sigma_0 \cdots \sigma_{n-1}(v_n))$$

for all  $n$ , where  $v_n$  satisfies  $\sigma_n(a_n) = a_{n-1} v_n$ . Let  $u_0 = \sigma_0(a_0)$  and  $u_n = \sigma_0 \cdots \sigma_{n-1}(v_n)$  for  $n \geq 1$ . By the properties of infinite products in  $M_L$  (see Sec. 2.2) we have that  $h_L(\alpha) = x$  if and only if  $\prod_{n=0}^{\infty} h_L(u_n) = x$ . The automaton  $\mathcal{B}_x$  simulates the run of the automaton  $\mathcal{A}_x$  of Lem. 2 on the sequence  $(h_L(u_n))_{n \in \mathbb{N}}$ . It remains to show how the automaton keeps track of  $h_L(u_n)$  as it reads  $(\xi_n, a_n)_{n \in \mathbb{N}}$ . We have that

$$\begin{aligned} h_L(u_0) &= (\text{compose}_{\xi_0}(h_L))(a_0), \\ h_L(u_n) &= \text{tail}_{\xi_n}(a_n, \text{compose}_{\xi_0 \cdots \xi_{n-1}}(h_L)) \end{aligned}$$

for  $n \geq 1$ . The automaton  $\mathcal{B}_x$  keeps track of one piece of information  $\xi \in \Xi_L$ , in addition to the state required for simulating a run of  $\mathcal{A}_x$  on  $(h_L(u_n))_{n \in \mathbb{N}}$ . Before reading  $(\xi_n, a_n)$  the value of  $\xi$  is  $\xi_0 \cdots \xi_{n-1}$ , where the empty product (corresponding to the initial value of  $\xi$ ) is the identity element of  $\Xi_L$ . Upon reading  $(\xi_n, a_n)$ , the automaton  $\mathcal{B}_x$  first computes  $h_L(u_n)$ , in which the value of  $\xi$  is used, then feeds the computed value to  $\mathcal{A}_x$ , and finally updates  $\xi$  to  $\xi \cdot \xi_n$ . Finally,  $\mathcal{B}_x$  accepts  $(\sigma_n, a_n)_{n \in \mathbb{N}}$  if and only if  $\mathcal{A}_x$  accepts  $(h_L(u_n))_{n \in \mathbb{N}}$ . ◀

When  $S$  is finite, for every individual  $\sigma \in S(\Sigma)$  we can compute the equivalence class  $[\sigma]_L$ , which yields the following.

► **Corollary 25.** *Let  $S \subseteq S(\Sigma)$  be finite. We can compute an automaton  $\mathcal{B}$  over  $S \times \Sigma$  such that  $\mathcal{B}$  accepts  $\hat{s}$  if and only if  $\hat{s}$  is congenial and generates a word in  $L$ .*

We next consider  $S$ -directed words, which are products of congenial words by Lem. 12.

► **Theorem 26** (Main Result for Directed Words). *Let  $L \subseteq \Sigma^\omega$  be an  $\omega$ -regular language. We can construct an automaton  $\mathcal{B}$  over  $\Xi_L$  such that for all  $s \in S(\Sigma)^\omega$ ,  $\mathcal{B}$  accepts  $\text{trace}_L(s)$  if and only if  $s$  directs some  $\alpha \in L$ .*

**Proof.** Denote the image of  $\text{congenials}_s$  under  $h_L$  by  $X_s \subset M_{L,f} \cup M_{L,\omega}$ . Recall from Lem. 12 that directed words are obtained by concatenating congenial words. To prove Thm. 26, we need to recognise the set of all  $\text{trace}_s$  for which the sub- $\omega$ -semigroup generated by  $X_s$  intersects the accepting set  $H$ . We can precompute a set  $\mathcal{X}$  of sets  $X$  that generate sub- $\omega$ -semigroups intersecting  $H$ . Our automaton  $\mathcal{B}$  needs to check that  $X_s$  contains at least one such set  $X$ . We denote by  $\mathcal{C}_x$  the projection of  $\mathcal{B}_x$  from Thm. 24 to  $\Xi_L$ , and observe:  $s \in \mathcal{L}(\mathcal{B})$  if and only if  $\bigvee_{X \in \mathcal{X}} \bigwedge_{x \in X} s \in \mathcal{L}(\mathcal{C}_x)$ . ◀

► **Corollary 27.** *Let  $S \subseteq S(\Sigma)$  be finite. We can compute an automaton  $\mathcal{B}$  over  $S$  such that  $\mathcal{B}$  accepts  $s$  if and only if  $s$  directs a word in  $L$ .*

The corollary above is proven in the same way as Cor. 25. We can apply the former to Arnoux-Rauzy words (which subsume Sturmian words) over the alphabet  $\Sigma = \{0, \dots, d-1\}$ . Let  $S = \{\lambda_0, \dots, \lambda_{d-1}, \rho_0, \dots, \rho_{d-1}\}$ . Recall that a word  $\alpha$  is Arnoux-Rauzy if and only if it is directed by a sequence  $s \in S^\omega$  which is (i) weakly primitive and (ii) contains infinitely many left-proper terms; by Lem. 8, such directive sequences direct a unique word, which is moreover infinite. By inspection,  $s \in S^\omega$  is weakly primitive if and only if for every  $0 \leq i < d$ , either  $\rho_i$  or  $\lambda_i$  occurs infinitely often. Hence the requirements (i-ii) can be checked by an automaton, and from Cor. 27 we obtain the following.

► **Theorem 28** (Main Result for Arnoux-Rauzy words). *Let  $\Sigma = \{0, \dots, d-1\}$ ,  $L \subseteq \Sigma^\omega$  be an  $\omega$ -regular language, and  $S$  be the finite set of morphisms generating the Arnoux-Rauzy words over  $\Sigma$  as described above. We can compute an automaton  $\mathcal{C}$  that accepts  $s \in S^\omega$  if and only if  $s$  is weakly primitive, contains infinitely many left-proper terms, and the unique Arnoux-Rauzy word directed by  $s$  is in  $L$ .*

## 8 Factor complexity and partial quotients

In this section, we refine our main results presented in Sec. 7 when extra information concerning the factor complexity is available. For characteristic Sturmian words, in particular, we will show that whether a deterministic automaton  $\mathcal{A}$  accepts the characteristic Sturmian word  $\alpha_\eta$  with slope  $\eta$  (as defined in (2)) only depends on the first  $N(\mathcal{A})$  terms in the continued fraction expansion of  $\eta$ , where  $N$  is independent of the slope  $\eta$ .

Fix an alphabet  $\Sigma$ , a (deterministic) automaton  $\mathcal{A}$  over  $\Sigma$  recognising an  $\omega$ -regular language  $L$ , a class of uniformly recurrent words  $\mathcal{W} \subseteq \Sigma^\omega$ , and  $S \subseteq S(\Sigma)$  such that

1.  $\alpha \in \mathcal{W}$  if and only if  $\alpha$  is directed (and hence generated) by some weakly primitive  $s \in S^\omega$ , and
2. there exists a computable function  $p$  such that  $p_\alpha(n) = p(n)$  for all  $\alpha \in \mathcal{W}$  and  $n \in \mathbb{N}$ , where  $p_\alpha(n)$  is the number of distinct factors of  $\alpha$  of length  $n$ .

By assumption (1), every  $\alpha \in \mathcal{W}$  is uniformly recurrent (see Sec. 3). Characteristic Sturmian, Sturmian, and  $d$ -letter Arnoux-Rauzy words are captured, respectively, with  $S = \{\lambda_0, \lambda_1\}$ ,  $p(n) = n + 1$ ;  $S = \{\lambda_0, \rho_0, \lambda_1, \rho_1\}$ ,  $p(n) = n + 1$ ; and  $S = \{\lambda_0, \dots, \lambda_{d-1}, \rho_0, \dots, \rho_{d-1}\}$ ,  $p(n) = n(d-1) + 1$  (see Sec. 3.2). We briefly mention that these techniques can also be applied to dendric words, whose factor-complexity function has a computable closed-form

expression. The directive sequences generating such words, however, are characterised by conditions that are more involved [15].

Our main idea is that every  $\alpha \in \mathcal{W}$  has a prefix  $u$  that is  $p$ -saturated with respect to  $\mathcal{A}$ , i.e., any uniformly recurrent extension  $\beta$  of  $u$  with factor complexity  $p$  agrees with  $\alpha$  upon acceptance by  $\mathcal{A}$  (Lem. 29), and hence also agrees upon membership in the recognised  $\omega$ -regular language  $L$ . The proof involves a careful consideration of Semënov's algorithm for deciding whether a given automaton accepts a given uniformly recurrent word.

► **Lemma 29.** *Let  $\alpha \in \mathcal{W}$  with effectively computable letters. Given an automaton  $\mathcal{A}$ , we can compute  $M \in \mathbb{N}$  such that any  $\beta \in \mathcal{W}$  with  $\beta[0, M) = \alpha[0, M)$  is accepted by  $\mathcal{A}$  if and only if  $\alpha$  is accepted by  $\mathcal{A}$ .*

**Proof.** Since we have access to the factor complexity function of  $\alpha$  and  $\beta$  by assumption (2), we can compute the recurrence function  $R_\alpha(n)$ : given  $n$ , enumerate prefixes  $\alpha[0, B)$  and integers  $K$  until we find that (i)  $\alpha[0, B)$  contains exactly  $p_\alpha(K)$  distinct factors of length  $K$ , and (ii) each of these  $p_\alpha(K)$  factors contains  $p_\alpha(n)$  distinct factors of length  $n$ . Then  $K = R_\alpha(n)$ .

Applying Prop. 5, we obtain  $M_1$  and  $\ell$  such that for any  $\beta \in \mathcal{W}$ , if  $\alpha[0, M_1) = \beta[0, M_1)$  and  $R_\alpha(n) = R_\beta(n)$  for  $n \leq \ell$ , then  $\alpha$  and  $\beta$  agree upon acceptance by  $\mathcal{A}$ . By the above paragraph, we can compute  $M_2$  such that if  $\alpha[0, M_2) = \beta[0, M_2)$ , then  $R_\alpha(n) = R_\beta(n)$  for  $n \leq \ell$ . It remains to take  $M = \max(M_1, M_2)$ . ◀

Recall Lem. 13. Given a weakly primitive and congenial expansion  $(\sigma_n, a_n)_{n \in \mathbb{N}}$  of  $\alpha$ , we can compute an increasing sequence  $(k_n)_{n \in \mathbb{N}}$  with  $k_0 = 0$  such that, writing  $l_n = k_{n+1} - 1$ ,  $\tau_n = \sigma_{k_n} \cdots \sigma_{l_n}$  and  $b_n = a_{l_n}$ , each  $\tau_n$  is positive, none of the strict prefixes of the composition  $\sigma_{k_n} \cdots \sigma_{l_n}$  is positive, and  $(\tau_n, b_n)_{n \in \mathbb{N}}$  is also a congenial expansion of  $\alpha$ . We refer to  $(\tau_n, b_n)_{n \in \mathbb{N}}$  as the sequence of *partial quotients* of  $(\sigma_n, a_n)_{n \in \mathbb{N}}$ . Since each  $\tau_n$  is positive, the sequence of partial quotients is, in particular, strongly primitive. The main result of this section is that, for the class  $\mathcal{W}$ , acceptance by  $\mathcal{A}$  is determined by a fixed number of initial partial quotients.

► **Theorem 30.** *We can compute  $N$  (that depends only on  $\mathcal{A}$  and  $p$ ) with the following property. Let  $\alpha, \alpha'$  be in  $\mathcal{W}$  with congenial  $S$ -adic expansions  $(\sigma_n, a_n)_{n \in \mathbb{N}}, (\sigma'_n, a'_n)_{n \in \mathbb{N}}$  and partial quotients  $(\tau_n, b_n)_{n \in \mathbb{N}}$  and  $(\tau'_n, b'_n)_{n \in \mathbb{N}}$ , respectively. If  $\tau_n \equiv_L \tau'_n$  and  $b_n = b'_n$  for all  $n \leq N$ , then  $\mathcal{A}$  accepts  $\alpha$  if and only if it accepts  $\alpha'$ .*

**Proof.** Let  $P$  be the set of all positive products  $\sigma_1 \cdots \sigma_r$  such that  $\sigma_i \in S$  for all  $i$  and  $\sigma_1 \cdots \sigma_i$  is not positive for all  $i < r$ . Construct finite  $\Pi \subseteq P$  such that  $\{[\pi]_L : \pi \in P\} = \{[\pi]_L : \pi \in \Pi\}$  and for all distinct  $\pi_1, \pi_2 \in \Pi$  we have that  $[\pi_1]_L \neq [\pi_2]_L$ . Let  $\Omega$  be the set of all congenial  $(\pi_n, b_n)_{n \in \mathbb{N}}$  over  $\Pi \times \Sigma$ . From the definition of congeniality it follows that  $\Omega \subseteq (\Pi \times \Sigma)^\omega$  is closed. Since  $(\Pi \times \Sigma)^\omega$  is compact, so is  $\Omega$ .

Next, consider  $\alpha$  with a congenial expansion  $\hat{s} = (\pi_n, b_n)_{n \in \mathbb{N}} \in \Omega$ . Such  $\alpha$  belongs to  $\mathcal{W}$ , but not every word in  $\mathcal{W}$  necessarily has such an expansion. By Lem. 29 there exists a prefix of  $\alpha$ , which is generated by  $(\pi_n, b_n)_{n=0}^M$  for some  $M$ , that is  $p$ -saturated. Write  $\Omega_M$  for the set of all  $\hat{s} \in \Omega$  whose first  $M$  terms generate a  $p$ -saturated finite word, observing that each  $\Omega_M$  is open. From Lem. 29 it follows that  $\{\Omega_M : M \in \mathbb{N}\}$  is an open cover of  $\Omega$ , which, by compactness, admits a finite sub-cover. That is, there exists  $N$  (which can be effectively computed by enumeration) such that for every  $(\pi_n, b_n)_{n \in \mathbb{N}} \in \Omega$ ,  $(\pi_n, b_n)_{n=0}^N$  generates a finite word that is  $p$ -saturated. That is, for expansions in  $\Omega$ , whether the generated word is accepted or not only depends on the first  $N$  terms.

Now consider (arbitrary)  $\alpha \in \mathcal{W}$  as in the statement of the theorem. Let  $(\pi_n)_{n \in \mathbb{N}}$  over  $\Pi$  be such that  $\tau_n \equiv_L \pi_n$  for all  $n$ , and observe that  $\hat{t} := (\pi_n, b_n)_{n \in \mathbb{N}}$  is also congenial. Let  $\beta$  be the word generated by  $\hat{t}$ . By Thm. 24,  $\mathcal{A}$  accepts  $\alpha$  if and only if it accepts  $\beta$ . By the earlier argument, whether  $\mathcal{A}$  accepts  $\beta$  only depends on  $(\pi_n, b_n)_{n=0}^N$ . It remains to observe that  $(\pi_n, b_n)_{n=0}^N$  is completely determined by  $(\tau_n, b_n)_{n=0}^N$ . ◀

Let us illustrate this result on Sturmian words. Given an automaton  $\mathcal{A}$ , apply Thm. 30 with  $S = \{\lambda_0, \lambda_1, \rho_0, \rho_1\}$  and  $p(n) = n + 1$  to compute  $N$ . Let  $\eta \in (0, 1) \setminus \mathbb{Q}$ ,  $\chi \in [-\eta, 1 - \eta]$ ,  $[0; a_1 + 1, a_2, \dots]$  be the continued fraction expansion of  $\eta$ ,  $(b_n)_{n \in \mathbb{N}}$  be an Ostrowski expansion of  $\chi$  in base  $\eta$ , and  $\alpha$  be the corresponding Sturmian word with slope  $\eta$  and intercept  $\chi$ . Recall that  $\alpha$  is directed (and uniquely generated; see Lem. 8) by the sequence

$$(\sigma_n)_{n \in \mathbb{N}} = \langle \lambda_0^{c_1}, \rho_0^{b_1}, \lambda_1^{c_2}, \rho_1^{b_2}, \lambda_0^{c_3}, \rho_0^{b_3}, \lambda_1^{c_4}, \rho_1^{b_4} \dots \rangle$$

where  $c_i = a_i - b_i$ . By the rules of Ostrowski expansion, at least one of  $c_n, c_{n+1}$  is non-zero for all  $n$ . Moreover, every composition of morphisms from  $\{\lambda_0, \lambda_1, \rho_0, \rho_1\}$  that includes two morphisms with differing indices is positive. Since at least one of  $b_n, c_n$  is non-zero for all  $n > 1$  (we could, however, have  $a_1 = b_1 = c_1 = 0$ ), we have that for every  $n$ ,  $\sigma_n \cdots \sigma_{n+5}$  is positive. Applying Thm. 30, whether  $\alpha$  is accepted by  $\mathcal{A}$  can be determined by looking at the first  $6N$  digits of the computable expansions of  $\eta$  and  $\chi$ , since these are guaranteed to generate at least the first  $N$  partial quotients of the unique  $S$ -adic expansion of  $\alpha$ .

---

## References

- 1 Jean-Paul Allouche and Jeffrey Shallit. *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press, 2003.
- 2 Pierre Béaur and Benjamin Hellouin de Menibus. Sturmian and infinitely desubstitutable words accepted by an  $\omega$ -automaton. In Anna Frid and Robert Mercas, editors, *Combinatorics on Words*, pages 104–116, Cham, 2023. Springer Nature Switzerland.
- 3 Valérie Berthé, Clelia De Felice, Francesco Dolce, Julien Leroy, Dominique Perrin, Christophe Reutenauer, and Giuseppina Rindone. Maximal bifix decoding. *Discrete Mathematics*, 338(5):725–742, 2015.
- 4 Valérie Berthé and Vincent Delecroix. Beyond substitutive dynamical systems:  $S$ -adic expansions. *RIMS Kôkyûroku Bessatsu*, B46:81–123, 2014.
- 5 Valérie Berthé, Charles Holton, and Luca Q. Zamboni. Initial powers of Sturmian sequences. *Acta Arithmetica*, 122(4):315–347, 2006.
- 6 Valérie Berthé, Toghrul Karimov, Joris Nieuwveld, Joël Ouaknine, Mihir Vahanwala, and James Worrell. On the decidability of monadic second-order logic with arithmetic predicates. In *Proceedings of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 1–14, 2024.
- 7 Valérie Berthé, Toghrul Karimov, Joris Nieuwveld, Joël Ouaknine, Mihir Vahanwala, and James Worrell. The monadic theory of toric words. *Theoretical Computer Science*, 1025:114959, 2025.
- 8 J. R. Büchi. On a decision method in restricted second order arithmetic. In *The Collected Works of J. Richard Büchi*, pages 425–435. Springer, 1990.
- 9 Yann Bugeaud and Michel Laurent. Combinatorial structure of Sturmian words and continued fraction expansion of Sturmian numbers. *Annales de l’Institut Fourier*, 73(5):2029–2078, 2023.
- 10 Olivier Carton and Wolfgang Thomas. The monadic theory of morphic infinite words and generalizations. *Information and Computation*, 176(1):51–65, 2002.
- 11 Fabien Durand. Corrigendum and addendum to “Linearly recurrent subshifts have a finite number of non-periodic factors”. *Ergodic Theory and Dynamical Systems*, 23(2):663–669, 2003.

- 12 Calvin C. Elgot and Michael O. Rabin. Decidability and undecidability of extensions of second (first) order theory of (generalized) successor. *The Journal of Symbolic Logic*, 31(2):169–181, 1966.
- 13 N. Pytheas Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*. Springer, 2002.
- 14 France Gheeraert, Marie Lejeune, and Julien Leroy.  $S$ -adic characterization of minimal ternary dendric shifts. *Ergodic Theory and Dynamical Systems*, 42(11):3393–3432, 2022.
- 15 France Gheeraert and Julien Leroy.  $S$ -adic characterization of minimal dendric shifts. *Discrete Mathematics & Theoretical Computer Science*, 27(2), 2025.
- 16 Amy Glen and Jacques Justin. Episturmian words: a survey. *RAIRO - Theoretical Informatics and Applications*, 43(3):403–442, 2009.
- 17 Philipp Hieronymi, Dun Ma, Reed Oei, Luke Schaeffer, Christian Schulz, and Jeffrey Shallit. Decidability for Sturmian words. *Logical Methods in Computer Science*, 20(3):12:1–12:38, 2024.
- 18 Toghrul Karimov. *Algorithmic verification of linear dynamical systems*. PhD thesis, Saarland University, 2024.
- 19 Toghrul Karimov, Edon Kelmendi, Joël Ouaknine, and James Worrell. What’s decidable about discrete linear dynamical systems? In *Principles of Systems Design: Essays Dedicated to Thomas A. Henzinger on the Occasion of His 60th Birthday*, pages 21–38. Springer, 2022.
- 20 A. Muchnik, A. Semenov, and M. Ushakov. Almost periodic sequences. *Theoretical Computer Science*, 304(1–3):1–33, 2003.
- 21 Dominique Perrin and Jean-Éric Pin. Semigroups and automata on infinite words. In *Semigroups, Formal Languages and Groups*, pages 49–72. Kluwer Academic Publishers, 1995.
- 22 Alexander Rabinovich. On decidability of monadic logic of order over the naturals extended by monadic predicates. *Information and Computation*, 205(6):870–889, 2007.
- 23 Raphael Robinson. Restricted set-theoretical definitions in arithmetic. *Proceedings of the American Mathematical Society*, 9(2):238–242, 1958.
- 24 A. L. Semenov. Logical theories of one-place functions on the set of natural numbers. *Mathematics of the USSR-Izvestiya*, 22(3):587–618, 1984.
- 25 Jeffrey Shallit. *The Logical Approach to Automatic Sequences: Exploring Combinatorics on Words with Walnut*. Cambridge University Press, 2022.
- 26 Wolfgang Thomas. A note on undecidable extensions of monadic second order successor arithmetic. *Archiv für mathematische Logik und Grundlagenforschung*, 17(1):43–44, 1975.
- 27 B. Trahtenbrot. Finite automata and the logic of one-place predicates. *Siberian Mathematical Journal*, 3:103–131, 1962.
- 28 Thomas Wilke. An Eilenberg theorem for  $\omega$ -languages. In *Automata, Languages and Programming*, pages 588–599, 1991.