RAIRO-Theor. Inf. Appl. 56 (2022) 7
https://doi.org/10.1051/ita/2022006

RAIRO - Theoretical Informatics and Applications
www.rairo-ita.org

AUTOMATIC SEQUENCES OF RANK TWO*

Jason P. Bell[1,**] and Jeffrey Shallit[2]

Abstract. Given a right-infinite word $\mathbf{x}$ over a finite alphabet A, the *rank* of $\mathbf{x}$ is the size of the smallest set S of words over A such that $\mathbf{x}$ can be realized as an infinite concatenation of words in S. We show that the property of having rank two is decidable for the class of k-automatic words for each integer $k \geq 2$.

Mathematics Subject Classification. 68R15, 11B85.

Received August 13, 2021. Accepted February 15, 2022.

1. Introduction

Let $k \geq 2$ be an integer. In this paper we study k-automatic sequences, which are those sequences (or infinite words) $(a_n)_{n\geq 0}$ over a finite alphabet, generated by a deterministic finite automaton with output (DFAO) taking, as input, the base-k representation of n and outputting a_n. (We call such a DFAO a k-DFAO.) Many interesting classical examples of sequences, including the Thue-Morse sequence, the Rudin-Shapiro sequence, and the paper-folding sequence are in this class. For more information about this well-studied class of sequences, see, for example, [1]. We mention that there is another well-known characterization of k-automatic sequences as the image, under a coding, of the fixed point of a k-uniform morphism [9]. Here a morphism is called k-uniform if the length of the image of every letter is k, and a coding is a 1-uniform morphism.

Let x be a finite nonempty word. We define x^ω to be the one-sided infinite word $xxx\cdots$. We say that an infinite word $\mathbf{z}$ is *ultimately periodic* if there exist finite words y, x, with x nonempty, such that $\mathbf{z} = yx^\omega$. Honkala proved [16] that the following problem is decidable: given a DFAO generating a k-automatic sequence $\mathbf{x}$, is $\mathbf{x}$ ultimately periodic? In fact, later results showed that this is actually *efficiently* decidable; see Leroux [25] and Marsault and Sakarovitch [28]. For other related work, see [3, 6, 11, 15, 24, 26, 31].

Let L be a language. We define L^ω to be the set of infinite words

$$\{x_1x_2\cdots : x_i \in L \setminus \{\epsilon\}\}.$$

If $\mathbf{x} \in L^\omega$ for some finite language L consisting of t nonempty words, then we say that $\mathbf{x}$ is of *rank* t. In particular, deciding the ultimate periodicity of $\mathbf{x}$ is the same as deciding if some suffix of $\mathbf{x}$ is of rank one. Words of rank one are periodic and this class of words is generally regarded as being well-behaved. For example, the subword

* Jason Bell is supported by NSERC grant 2016-03632. Jeffrey Shallit is supported by NSERC grant 2018-04118.

Keywords and phrases: Combinatorics on words, automatic sequence, primitive word, rank two.

[1] Department of Pure Mathematics, University of Waterloo, Waterloo, ON N2L 3G1, Canada.
[2] School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

** Corresponding author: shallit@uwaterloo.ca

complexity functions of periodic and more generally of ultimately periodic words are uniformly bounded. Thus rank can be seen as providing a measurement of how far a word is from being periodic, and in this sense we view the rank as giving some measure of the complexity of a word.

While the rank gives some insight into the structure of the word and how it can be constructed from subwords, it can nevertheless be difficult to determine its precise value. A large part of this difficulty lies in the fact that there is some relationship with undecidable "tiling" problems, in the sense that if L is a finite set of words then we can view these words as "tiles" and the question of whether a right-infinite word is in L^ω is then asking whether there exists a tiling of the word from L. For example, the Post correspondence problem [32], which is undecidable, asks whether, given two finite sets of words of the same size, $\{a_1, \ldots, a_m\}$ and $\{b_1, \ldots, b_m\}$, over a common alphabet, there exist $k \geq 1$ and $i_1, \ldots, i_k \leq m$ such that $a_{i_1} \cdots a_{i_k} = b_{i_1} \cdots b_{i_k}$. Thus it is possible that a right-infinite word $\mathbf{w}$ has rank m but one can have difficulties identifying the language L of size m for which $\mathbf{w} \in L^\omega$ by looking at prefixes.

In general, it is the fact that words cannot always be tiled unambiguously that complicates decision procedures involving tilings, and such problems typically become more complex as the number of tiles involved increases. For example, it is known that the Post correspondence problem is solvable if the lists consist only of one or two words [12], while it is undecidable for lists of seven words or more [29]. Similarly, for Wang tiles, there has been some interest in finding the smallest number N such that the tiling problem using N tiles is undecidable [10, 18, 19].

Our main result is to show that the property of being of rank two is decidable for automatic words.

Theorem 1.1. *Let $k \geq 2$ be an integer and let $\mathbf{x}$ be a k-automatic sequence. Then there is an algorithm to decide whether $\mathbf{x}$ is of rank two.*

This algorithm is considerably more involved than the corresponding algorithm used for determining whether a word has rank one; moreover, we do not currently know how to extend our method to arbitrary suffixes of $\mathbf{x}$, nor to sequences of higher rank.

A key component in the procedure given in Theorem 1.1 is the following result, which shows that there is a striking dichotomy in the possible powers of words that can appear in a k-automatic sequence. Recall that we say that a finite word x is a *factor* of a (possibly infinite) word y if x appears as a contiguous block inside y.

Theorem 1.2. *Let $k \geq 2$ be an integer and let $\mathbf{x}$ be a k-automatic sequence. Then there is a computable bound B such that, for each finite word y, if y^B occurs as a factor of $\mathbf{x}$ then y occurs with unbounded exponent in $\mathbf{x}$.*

The outline of this paper is as follows. In Section 2 we give the notation that will be used throughout the paper. In Section 3, we provide the necessary background on repetitive words. In Section 4, we recall a key result in first-order logic and use it to deduce Theorem 1.2 along with a key technical result that will be used in the proof of Theorem 1.1. Then in Section 5, we prove the key combinatorial lemmas that will be used in giving the decision procedure in Theorem 1.1. Finally, in Section 6, we give the proof of Theorem 1.1.

2. Notation and definitions

Throughout this paper we will make use of the following notation and definitions.

If $w = xyz$ for words w, x, y, z with w and z possibly infinite, we say that y is a *factor* of w, x is a *prefix* of w, and z is a *suffix* of w.

Let $\mathbf{x} = a_0 a_1 a_2 \cdots$ be an infinite word. By $\mathbf{x}[i..i+n-1]$ we mean the length-n word $a_i \cdots a_{i+n-1}$. By $\mathrm{Fac}(\mathbf{x})$ we mean $\{\mathbf{x}[i..i+n-1] : i, n \geq 0\}$, the set of all finite factors of $\mathbf{x}$.

Given a finite word $w = a_1 a_2 \cdots a_n$ we say that w is of *period* p if $a_i = a_{i+p}$ for $1 \leq i \leq n-p$. A word can have multiple periods; for example, the French word `entente` has periods 3, 6, and 7. We refer to the smallest positive period as *the* period of w, and denote it by $\mathrm{per}(w)$. The *exponent* of a finite word w is defined to be $\exp(w) = |w|/\mathrm{per}(w)$.

A finite nonempty word w is *primitive* if it is a non-power, that is, if it cannot be written as $w = x^e$ for some $e \geq 2$. If w appears in $\mathbf{x}$ to arbitrarily large powers, we say that w is *of unbounded exponent* in $\mathbf{x}$. If $\mathbf{x}$ has only finitely many primitive factors of unbounded exponent, we say it is *discrete*.

We assume the reader has a basic background in formal languages and finite automata theory. For the needed concepts, see, for example, [17].

3. Repetitive words

An infinite word $\mathbf{x}$ is called *repetitive* if for all n there exists a finite nonempty word w such that $w^n \in \mathrm{Fac}(\mathbf{x})$. It is called *strongly repetitive* if there exists a finite nonempty word w such that $w^n \in \mathrm{Fac}(\mathbf{x})$ for all n.

Suppose h is a morphism and a is a letter such that $h(a) = az$ for some z for which $h^i(z) \neq \epsilon$ for all i. Then we say that h is *prolongable* on a. In this case $h^\omega(a) := a\, z\, h(z) \ldots$ is an infinite word that is a fixed point of h, and we say that $h^\omega(a)$ is a *pure morphic* word. If $\mathbf{x}$ is the image, under a coding, of a pure morphic word, we say that $\mathbf{x}$ is *morphic*.

Several writers have investigated the repetitive and strongly repetitive properties of pure morphic words. Ehrenfeucht and Rozenberg [13] showed that a pure morphic word is repetitive if and only if it is strongly repetitive, and also showed that these conditions are decidable. Mignosi and Séébold [30] proved that for every morphic word $\mathbf{x}$ there exists a constant M such that $w^M \in \mathrm{Fac}(\mathbf{x})$ if and only if $w^n \in \mathrm{Fac}(\mathbf{x})$ for all n; also see [21]. Kobayashi and Otto [23] gave an efficient algorithm to test the repetitivity of a pure morphic word. Klouda and Starosta [20] showed further that every pure morphic word $\mathbf{x}$ is discrete, while the authors [4] proved that words of linear factor complexity are discrete.

Only the last of these results applies to the case that concerns us in this paper (where $\mathbf{x}$ is k-automatic) because automatic words need not be pure morphic. For example, it is not hard to show that the Rudin-Shapiro sequence is not pure morphic.

Recently, in a thus-far unpublished manuscript, Klouda and Starosta [22] showed that morphic words (and hence k-automatic words) are discrete.

4. First-order logic

Let $\mathbf{x}$ be an infinite word. We will work with first-order logic (see, for example, [2]). Recall that $\langle \mathbb{N}, +, n \to \mathbf{x}[n] \rangle$ is the set of all first-order logical formulas consisting of variables (with domain $\mathbb{N} = \{0, 1, 2, \ldots\}$, the natural numbers), quantifiers $\exists$ and $\forall$, addition, logical operations, comparisons of integers, and indexing into $\mathbf{x}$. If a variable is associated with a quantifier, it is said to be *bound*; otherwise it is *unbound*.

Certain key parts of the decision procedure in Theorem 1.1 rely on the following result, which is essentially a consequence of the results of Bruyère *et al.* [5]; also see [7].

Theorem 4.1. *Let $k \geq 2$ be an integer, let $\mathbf{x}$ be a k-automatic sequence, and let φ be a logical formula, expressible in the first-order structure $\langle \mathbb{N}, +, n \to \mathbf{x}[n] \rangle$. Then*

(a) If φ has no unbound variables, then the truth of φ is decidable.
(b) If φ has unbound variables, we can computably determine a DFA that recognizes precisely the base-k representations of those natural number values of the unbound variables that make φ true.

As an illustration of these ideas, we consider four basic properties of factors of sequences.

- FactorEq(i, j, n) asserts that $\mathbf{x}[i..i + n - 1] = \mathbf{x}[j..j + n - 1]$;
- Pref(i, j, m, n) asserts that $\mathbf{x}[i..i + j - 1]$ is a prefix of $\mathbf{x}[m..m + n - 1]$;
- Suff(i, j, m, n) asserts that $\mathbf{x}[i..i + j - 1]$ is a suffix of $\mathbf{x}[m..m + n - 1]$;
- Per(i, n, p) asserts that $\mathbf{x}[i..i + n - 1]$ has period p;
- Match(i, j, n, p) asserts that $\mathbf{x}[j..j + n - 1]$ has period p and furthermore that $\mathbf{x}[i..i + p - 1] = \mathbf{x}[j..j + p - 1]$.

All four can be expressed in first-order logic, as follows:

- FactorEq(i, j, n): $\forall t\ (t < n) \implies \mathbf{x}[i+t] = \mathbf{x}[j+t]$.
- Pref(i, j, m, n): $(j \le n) \wedge$ FactorEq(i, m, j).
- Suff(i, j, m, n): $(j \le n) \wedge$ FactorEq$(i, m+n-j, j)$.
- Per(i, n, p): $(p \ge 1) \wedge (p \le n) \wedge$ FactorEq$(i, i+p, n-p)$.
- Match(i, j, n, p): Per$(j, n, p) \wedge$ FactorEq(i, j, p).

We use these as building blocks in what follows.

For automatic sequences, we can now get easy proofs of the decidability of the repetitive and strongly repetitive properties.

Theorem 4.2. *Let $k \ge 2$ be an integer, let $\mathbf{x}$ be a k-automatic sequence, and let z be a given nonempty word. Then the following problems are decidable:*

(a) Do arbitrarily large powers of z appear in $\mathbf{x}$?
(b) If the answer to (a) is no, what is the largest exponent e such that z^e appears in $\mathbf{x}$?

Proof.

(a) Let $z = a_1 a_2 \cdots a_r$. We can write a first-order statement asserting that z is a factor of $\mathbf{x}$ as follows:

$$\exists i\ \mathbf{x}[i] = a_1 \wedge \mathbf{x}[i+1] = a_2 \wedge \cdots \wedge \mathbf{x}[i+r-1] = a_r, \tag{4.1}$$

and so it is decidable if this is the case. If z does indeed appear in $\mathbf{x}$, by creating the automaton for the expression

$$\mathbf{x}[i] = a_1 \wedge \mathbf{x}[i+1] = a_2 \wedge \cdots \wedge \mathbf{x}[i+r-1] = a_r$$

and using breadth-first search to identify an accepting path we can find a specific $i = i_0$ for which equation (4.1) holds. Then arbitrarily large powers of z appear in $\mathbf{x}$ if and only if

$$\forall m\ \exists n > m\ \exists j\ \ \text{Match}(i_0, j, n, r).$$

We emphasize that here i_0 and r are constants depending on z, and not unbound variables.

(b) If z occurs in $\mathbf{x}$, but not with arbitrarily large powers, then we can determine the largest (fractional) power z^e occurring in $\mathbf{x}$ as follows: create the DFA corresponding to the logical formula

$$\exists m\ (\exists j\ \ \text{Match}(i_0, j, m, r) \wedge \neg \exists j'\ \ \text{Match}(i_0, j', m-1, r)).$$

Again, i_0 and r are constants and not unbound variables. This DFA will accept the base-k representation of exactly one m, and then $e = m/r$.

□

As a consequence we can prove Theorem 1.2.

Proof of Theorem 1.2. Let M be a k-DFAO generating the sequence $\mathbf{x}$. As is well-known ([14], Thm. 14) for every automatic sequence $\mathbf{x}$ there is a computable constant C such that if y appears as a factor of $\mathbf{x}$, it must appear starting at a position that is $\le C|y|$.

Now consider the following first-order formula:

$$\varphi(i, n, p) : \exists j\, \text{Match}(i, j, n, p).$$

Theorem 4.1 implies that from a DFA for $\mathbf{x}$ we can computably determine a DFA M' accepting those triples $(i,n,p)_k$ in parallel for which $\varphi(i,n,p)$ is true. Suppose M' has r states. We now claim that M' accepts some word $(i,n,p)_k$ with $n/p > k^r C$ if and only if $\mathbf{x}$ contains arbitrarily large powers of $\mathbf{x}[i..i+n-1]$.

One direction is trivial. For the other direction, note that if there exists at least one i for which $\varphi(i,n,p)$ is true, then by the discussion above, there must be an $i' \leq Cp$ for which $\varphi(i',n,p)$ is true. Consider the base-k representation of the triple (i',n,p). Since $n > k^r Cp$ we have $n > k^r i'$, and hence the base-k representation of (i',n,p) starts with at least r zeros in the components corresponding to both i' and p, and a nonzero digit in the n component. We may now apply the pumping lemma to $z = (i',n,p)_k$ to get longer and longer strings with the same value of i' and p, but arbitrarily large n. From the definition of M' this means that there is an infinite sequence of increasing n for which there exists a j with $\mathbf{x}[j..j+n-1] = \mathbf{x}[i'..i'+p-1]^{n/p}$.

We may now take $B = k^r C$ to prove the result. □

We can also prove that k-automatic words are discrete, and even something more general. The following result appears in earlier work of the authors ([4], Thm. 1.2) but we give a simpler and more self-contained proof here. We note, however, that the following proof does not recover the upper bound on the number of primitive factors that can occur with unbounded exponent (up to cyclic equivalence), which was given in [4].

Theorem 4.3. *Let $\mathbf{x}$ be an infinite word with linear subword complexity. Then $\mathbf{x}$ is discrete.*

Proof. Suppose that the subword complexity of $\mathbf{x}$ is bounded by cn for some constant c and $n \geq 1$. Further suppose, contrary to what we want to prove, that $\mathbf{x}$ has infinitely many primitive factors $x_1, x_2, \ldots, x_{2c+1}, \ldots$ for which arbitrarily large powers appear in $\mathbf{x}$.

Choose $2c+1$ of them that are strictly increasing in size, $1 \leq |x_1| < |x_2| < \cdots < |x_{2c+1}|$. We will use these factors to derive a contradiction by showing that $\mathbf{x}$ must have more than cn words of length n for some c. Replace each x_i by y_i, an appropriate power of the x_i so that all the y_i are the same length $d > c$. (The y_i are no longer primitive, but it doesn't matter.)

Pick $n = 3d+1$. For each word y_i find an occurrence of the third power of y_i in $\mathbf{x}$. Note that $\mathbf{x}$ cannot be eventually periodic since by assumption it has infinitely many distinct primitive factors, and so $\mathbf{x}$ has a factor of the form $y_i^3 y$ with $|y| \leq d$ and y not a prefix of y_i. We let z denote the largest common prefix of y and y_i. Then $y_i = zz_i$ and so we see $\mathbf{x}$ has a factor of the form $(zz_i)^3 za$ in which a is a letter that is different from the first letter of z_i. Thus after replacing y_i by $z_i z$, we may assume that $\mathbf{x}$ has an occurrence of y_i^3 that is followed by a letter a that is different from the first letter of y_i.

For each i, we take a suffix $\mathbf{x}_i$ of $\mathbf{x}$ that has $y_i^3 a_i$ as a prefix for some letter a_i that is not the first letter of y_i. Now consider the list L_i of the $2d+1$ length-n factors of $\mathbf{x}$ that start at position p of $\mathbf{x}_i$ for $p = 0, \ldots, 2d$. For each $i \in \{1, \ldots, 2c+1\}$ there are $2d+1$ such words. We first claim for $i = 1, \ldots, 2c+1\}$ these $2d+1$ words are pairwise distinct. To see this, since a_i is not the first letter of y_i and since $p \leq 2d$ we that the length-n factor of $\mathbf{x}$ that starts at the p-th position of $\mathbf{x}_i$ has a prefix of the form $(y_i')^{3-p/d} a_i$ for some cyclic shift y_i' of y_i; moreover, $(y_i')^{3-p/d} a_i \neq (y_i')^{3-(p+1)/d}$ since a_i is not the first letter of y_i. Since $p \leq 2$, we then see that we can recover y_i' and p from the length-n prefix $\mathbf{x}_i$ that begins at position i.

Similarly, if we compare a word z of L_i and a word z' of L_j for $i \neq j$, then since $n = 3d+1$, z and z' share a common length-d prefix. By construction, this prefix is a cyclic shift of both y_i and y_j, which is impossible since the original x_i were distinct primitive words.

So we have constructed at least $(2d+1)(2c+1)$ distinct length-n subwords of $\mathbf{x}$. But by assumption, $\mathbf{x}$ has at most $cn = c(3d+1)$ words of length n, and since $c(3d+1) < (2d+1)(2c+1)$, we obtain a contradiction. □

Corollary 4.4. *If $\mathbf{x}$ is k-automatic, we can explicitly determine the (finitely many) primitive factors w such that w is of unbounded exponent in $\mathbf{x}$.*

Proof. We can easily write down a first-order formula asserting that $w = \mathbf{x}[i..i+p-1]$ is primitive, that it is the first occurrence of this factor in $\mathbf{x}$, and that unboundedly large powers of w appear in $\mathbf{x}$, as follows:

$$\forall m \; \exists j, n \; (n > m) \; \wedge \; \textsc{Primitive}(i,p) \; \wedge \; \textsc{EarliestFac}(i,j,p) \; \wedge \; \textsc{Per}(j,n,p),$$

where

$$\begin{aligned}\textsc{EarliestFac}(i,j,n) &= \textsc{FactorEq}(i,j,n) \ \wedge \ (\forall t \ \textsc{FactorEq}(t,j,n) \implies t \geq i);\\ \textsc{Primitive}(i,n) &:= \neg(\exists j \ (j>0) \ \wedge \ (j<n) \ \wedge \ \textsc{FactorEq}(i,i+j,n-j) \wedge \\ &\qquad \textsc{FactorEq}(i,(i+n)-j,j)).\end{aligned}$$

So by Theorem 4.1 we can compute an automaton recognizing the finitely many pairs i, p. □

Finally, we prove two technical results, which together will play a key role in the proof of Theorem 1.1.

Proposition 4.5. *Given a k-automatic sequence $\mathbf{x}$, and a nonempty factor u, and fixed natural numbers L, N, one can decide if there exist a word v of length $\geq N$, such that u is neither a prefix or suffix of v, and natural numbers $p_1, p_2, \ldots, p_L$ such that $vu^{p_1}vu^{p_2}v\cdots vu^{p_L}$ is a prefix of $\mathbf{x}$.*

Proof. Fix i and d such that $u = \mathbf{x}[i..i+d-1]$. Consider the logical formula ρ defined below:

$$\begin{aligned}\rho(r) : r \geq N \ \wedge \ \neg\,\textsc{Pref}(i,d,0,r) \ \wedge \ \neg\,\textsc{Suff}(i,d,0,r) \ \wedge \ \exists p_1,p_2,\ldots,p_L \\ \textsc{FactorEq}(0,r+p_1d,r) \ \wedge \ \textsc{FactorEq}(0,2r+(p_1+p_2)d,r) \ \wedge \cdots \wedge \\ \textsc{FactorEq}(0,(L-1)r,(p_1+p_2+\cdots+p_{L-1})d) \wedge \\ \textsc{Per}(r,p_1d,d) \ \wedge \ \textsc{Per}(2r+p_1d,p_2d,d) \ \wedge \cdots \wedge \ \textsc{Per}(Lr+(p_1+p_2+\cdots+p_{L-1})d,p_Ld,d).\end{aligned}$$

The formula $\gamma : \exists r \ \rho(r)$ is true if and only if the desired $v = \mathbf{x}[0..r-1]$ and the $p_1, p_2, \ldots, p_L$ exist. If γ is true, then from the automaton for ρ we can easily find the smallest r for which $\rho(r)$ evaluates to true, just as we did above in the proof of Corollary 4.4. □

Proposition 4.6. *Let $k, m \geq 2$ be integers. Given a k-automatic sequence $\mathbf{x}$, and a binary word $i_0\cdots i_{m-1} \in \{0,1\}^m$, we can decide whether there exist nonempty factors u_0 and u_1 of $\mathbf{x}$ such that u_0 is neither a prefix nor suffix of u_1; u_1 is neither a prefix nor suffix of u_0; and such that $u_{i_0}\cdots u_{i_{m-1}}$ is a prefix of $\mathbf{x}$.*

Proof. We can construct a first-order formula encoding these assertions. The idea is that $u_0 = \mathbf{x}[i..i+r-1]$ and $u_1 = \mathbf{x}[j..j+s-1]$ for some i, j, r, s. If $u_{i_0}\cdots u_{i_{m-1}}$ is a prefix of $\mathbf{x}$, then there exist starting positions $p_0 = 0, p_1, \ldots, p_{m-1}$ and lengths $q_0, q_1, \ldots, q_{m-1}$ corresponding to each of the occurrences of the u_{i_t}. We then assert that the starting positions obey the rule that $p_{t+1} = p_t + q_t$ for $0 \leq t < m-1$, and that each occurrence u_{i_t} matches $\mathbf{x}[i..i+r-1]$ or $\mathbf{x}[j..i+s-1]$, according to whether $i_t = 0$ or $i_t = 1$, respectively. This gives us the following first-order formula:

$$\begin{aligned}&\exists i,j,r,s,p_0,p_1,\ldots,p_{m-1},q_0,q_1,\ldots,q_{m-1} \ r>0 \ \wedge \ s>0 \wedge \\ &\neg\,\textsc{Pref}(i,r,j,s) \ \wedge \ \neg\,\textsc{Suff}(i,r,j,s) \ \wedge \ \neg\,\textsc{Pref}(j,s,i,r) \ \wedge \ \neg\,\textsc{Suff}(j,s,i,r) \wedge \\ &(\forall t<m \ ((i_t=0) \implies q_t = r \ \wedge \ \textsc{FactorEq}(i,p_t,q_t)) \wedge \\ &((i_t=1) \implies q_t = s \ \wedge \ \textsc{FactorEq}(j,p_t,q_t))) \wedge \\ &p_0 = 0 \ \wedge \ (\forall t < m-1 \ p_{t+1} = p_t + q_t).\end{aligned}$$

□

5. Combinatorial lemmas

In this section, we prove results about semigroup equations, which will again play a key role in our decision procedure for testing rank two.

Given a finite alphabet Σ, we write $a \leq b$ for $a, b \in \Sigma^*$ if a is a prefix of b. We note that if a and b have the property that $a \not\leq b$ and $b \not\leq a$ then a and b generate a free semigroup and every word in Σ^* has a unique largest prefix in $\{a, b\}^*$.

Lemma 5.1. *Let Σ be a finite alphabet and let r and s be elements in Σ^* that do not commute. Then there exist $a, b \in \Sigma^*$ such that $a \not\leq b$ and $b \not\leq a$ and such that $r, s \in \{a, b\}^*$.*

Proof. Suppose towards a contradiction that the conclusion to the statement of the lemma does not hold. Then among all counterexamples $(r, s) \in (\Sigma^*)^2$, we pick (r, s) with $|r| + |s|$ minimal. Then either $r \leq s$ or $s \leq r$, or else we can take $a = r$ and $b = s$ and we obtain a contradiction. Thus we may assume without loss of generality that $r \leq s$ and so we write $s = ry$. Then $r, s \in \{r, y\}^*$. Then either $|r| + |y| < |r| + |s|$ or $r = \epsilon$. But we cannot have $r = \epsilon$ since r and s do not commute. Since s and r do not commute, r and y do not commute and by minimality of $|r| + |s|$, we have $\{r, y\}^* \subseteq \{a, b\}^*$ with $a \not\leq b$ and $b \not\leq a$. But this is a contradiction, since $r, s \in \{r, y\}^* \subseteq \{a, b\}^*$. The result follows. □

Lemma 5.2. *Let Σ be a finite alphabet, let $u, v \in \Sigma^+$ be words in Σ^* such that there do not exist $a, b \in \Sigma^*$ with $|a| + |b| < |u| + |v|$ and with $u, v \in \{a, b\}^*$. Let x, y be single letters, and suppose that $\sigma : \{x, y\}^* \to \{u, v\}^*$ is a morphism defined by $\sigma(x) = u$ and $\sigma(y) = v$. Suppose further that there exist words $d, d' \in \Sigma^*$, and words $w, w' \in \{x, y\}^*$ such that the following hold:*

1. *$d\sigma(w) = \sigma(w')d'$;*
2. *w, w' both have x as a prefix and at least one occurrence of y;*
3. *d does not contain u as a suffix and $|d| < \max(|u|, |v|)$.*

Then $d = \epsilon$.

Proof. Let $z = d\sigma(w) = \sigma(w')d'$. By considering the prefix of z of length $|du|$, we see $du = uc$ for some c and hence by the Lyndon-Schützenberger theorem [27] there exist words r, s such that $u = (rs)^\alpha r$, $d = rs$, $c = sr$. Notice that if r and s commute, then again by the Lyndon-Schützenberger theorem, they are powers of some word t, which then gives $u = t^i$ and $d = t^j$. But now $\{u, v\} \subseteq \{t, v\}^*$ and so $|t| + |v| = |u| + |v|$ by hypothesis and so $u = t$. But then since u is not a suffix of d, we must have $d = \epsilon$. Hence we may assume that r and s do not commute. Now by Lemma 5.1, there are words a and b such that $a \not\leq b$ and $b \not\leq a$ such that r and s are in the free monoid generated by a and b. Then $d, u \in \{a, b\}^*$. We claim that $v \notin \{a, b\}^*$. To see this, suppose that this is not the case. Then $u, v \in \{a, b\}^*$ and so by assumption, we must then have $u = a$ and $v = b$ after relabelling. But now d is such that $|d| < \max(|a|, |b|)$ and since $d \in \{a, b\}^*$, we see that d is either a power of a or a power of b. But d cannot be a^i with $i \geq 1$ since d does not have $u = a$ as a suffix and hence d must be b^j. But now the equation $du = uc$ gives $b^j a = ac$, which is impossible since $b \not\leq a$ and $a \not\leq b$. Thus $v \notin \{a, b\}^*$.

We write $v = \gamma v'$ such that $\gamma \in \{a, b\}^*$ and v' does not have a or b as a prefix. Then v' is nonempty since $v \notin \{a, b\}^*$.

By assumption there is some $i \geq 1$ such that $x^i y$ is a prefix of w. Then $du^i v$ is a prefix of z and so the longest prefix of z in $\{a, b\}^*$ is $du^i\gamma$. On the other hand, there is some $j \geq 1$ such that w' has $x^j y$ as a prefix and so the longest prefix of z in $\{a, b\}^*$ is $u^j\gamma$. Then we must have $du^i\gamma = u^j\gamma$. It follows that $j \geq i$ since u is nonempty. Cancelling $u^i\gamma$ on the right gives $d = u^{j-i}$. Since d does not have u as a suffix, we then see that $j = i$ and so $d = \epsilon$, as required. □

Notation 5.1. *For the remainder of this section, we adopt the following notation and assumptions:*

- *u and v are words with the property there do not exist $a, b \in \Sigma^*$ with $|a| + |b| < |u| + |v|$ and with $u, v \in \{a, b\}^*$;*
- *we take x and y to be symbols and we let $\sigma : \{x, y\}^* \to \{u, v\}^*$ denote the homomorphism from the free monoid on the set $\{x, y\}$ to the monoid generated by u and v given by $\sigma(x) = u$ and $\sigma(y) = v$.*

Proposition 5.1. *Let $w \in \{x,y\}^*$ be a word that contains at least five occurrences of xy and let z be a word in Σ^* with $|z| = \max(|u|,|v|)$ and such that z does not contains u or v as a prefix. Then $\sigma(w)z$ is not a factor of a word in $\{u,v\}^\omega$.*

Proof. Since w has at least five occurrences of xy, we can write w in the form w_0xyw_1 with w_0 and w_1 both having at least two occurrences of xy.

Suppose towards a contradiction that $\sigma(w)z$ is a factor of a word in $\{u,v\}$. Then there is some word $w' = \xi_1 \cdots \xi_s \in \{x,y\}^*$, with $\xi_i \in \{x,y\}$ for $i = 1,\ldots,s$, such that $\sigma(w)z$ is a factor of $\sigma(w')$, and so we can write $\sigma(w') = a\sigma(w_0)uv\sigma(w_1)zb$ for some words a and b. Then there must be some largest i such that $w'' := \sigma(\xi_1)\cdots\sigma(\xi_i)$ is a prefix of $a\sigma(w_0)u$. We now argue *via* cases.

Case I: $\xi_{i+1} = y$.

In this case, $w''v$ is not a prefix of $a\sigma(w_0)u$, but it is a prefix of $a\sigma(w_0)uv$. In particular, we can factor $u = u_0d$ such that $\sigma(w'') = a\sigma(w_0)u_0$ and we have

$$dv\sigma(w_1)zb = v\sigma(\xi_{i+2})\cdots\sigma(\xi_s).$$

Notice $|d| \le |u| \le \max(|u|,|v|)$, and we cannot have $|d| = |u|$, or else u and v would share a prefix, which cannot happen by hypothesis. Thus $|d| < \max(|u|,|v|)$ and d cannot have v as a suffix, since otherwise v would be a suffix of u and so we could write $u = av$ and then $u, v \in \{c,v\}^*$ with $|c|+|v| < |u|+|v|$, which is a contradiction. Also, we cannot have $d = \epsilon$, since $v \not\leq u$ and $u \not\leq v$, and this would imply that w_1 is a prefix of $\xi_{i+2}\cdots\xi_s$, and so zb would have to be a prefix of some word of the form $\sigma(\xi_j)\cdots\sigma(\xi_s)$. But this would then say that either u or v is a prefix of z, since $|z| = \max(|u|,|v|)$ and this is a contradiction.

Now there is some smallest j such that $w'' := v\sigma(\xi_{i+2})\cdots\sigma(\xi_j)$ contains dv as a prefix. Then if $\xi_{i+2} = \cdots = \xi_j = y$, then v is a factor of $\sigma(\xi_{j-1})\sigma(\xi_j) = vv$ and since v is primitive, this forces $dv = v\sigma(\xi_{i+2})\cdots\sigma(\xi_j)$ (see, *e.g.*, [8, p. 336]). In particular, this means d would be a power of v, contradicting the fact that it cannot contain v as a suffix.

By assumption w_1 has at least one occurrence of xy and so we may write $w_1 = y^qxw_2$ for some $q \ge 0$ and w_2 containing at least one copy of x and y. Then there is some smallest k such that $v\sigma(\xi_{i+2})\cdots\sigma(\xi_k)$ contains $dv^{q+1}u$ as a prefix, and since w_2 contains a copy of at least one x and y, we see that $v\sigma(\xi_{i+2})\cdots\sigma(\xi_k)$ is a prefix of $dv^{q+1}u\sigma(w_2)$. Hence we can write

$$dv^{q+1}u\sigma(w_2) = v\sigma(\xi_{i+2})\cdots\sigma(\xi_k)d'.$$

Since $j \le k$, we see that $\xi_\ell = x$ for some $\ell \in \{i+2,\ldots,k\}$. Then Lemma 5.2 now gives that $d = \epsilon$, which we have ruled out. Thus we have completed the proof in this case.

Case II: $\xi_{i+1} = x$.

In this case there is some word d with $|d| \le |u|$ such that $\sigma(\xi_1)\cdots\sigma(\xi_i)ud = a\sigma(w_0)u$. Arguing as in Case I, we can show that some letter in w_0 is equal to y and that some ξ_j with $j \le i$ is equal to y. Then applying Lemma 5.2 in the opposite monoid now gives $d = \epsilon$. But we obtain a contradiction in this case as in Case I. This completes the proof. □

Definition 5.3. Let p be a positive integer. We say that a word in $\{u,v\}^\omega$ is *p-syndetic* if it has no factors of the form uv^ju or vu^jv with $j \ge p$.

Lemma 5.4. *Let $k \ge 2$ be an integer and let $\mathbf{x}$ be a k-automatic word. Then there is a positive integer D such that whenever $u, v \in \Sigma^*$ are such that*

1. *the assumptions of Notation 5.1 hold,*

2. $\mathrm{Fac}(u^\omega) \not\subseteq \mathrm{Fac}(\mathbf{x})$,
3. $\mathrm{Fac}(v^\omega) \not\subseteq \mathrm{Fac}(\mathbf{x})$, *and*
4. $\mathbf{x}$ *has a prefix in* $\{u,v\}^D$,

we necessarily have $\mathbf{x}$ *is in* $\{u,v\}^\omega$.

Proof. By Theorem 1.2, there is some computable $p = p(\mathbf{x})$ such that u^p and v^p are not factors of $\mathbf{x}$. By a result of Cobham [9], there is a computable number $\kappa = \kappa(\mathbf{x})$ such that every factor of $\mathbf{x}$ of length L occurs in the prefix of $\mathbf{x}$ of length κL. We now take $D = 10p^2\kappa + p$ and suppose that $\mathbf{x}$ has a prefix in $\{u,v\}^D$ but that it is not in $\{u,v\}^\omega$. Then there is some largest $d \geq D$ such that $\mathbf{x}$ has a prefix in $\{u,v\}^d$. Then $\mathbf{x}$ has a prefix of the form az with $a \in \{u,v\}^d$ and $|z| = \max(|u|,|v|)$ and z does not have u or v as a prefix. Now a is necessarily p-syndetic since u^p and v^p are not factors of w. It follows that every factor of w in $\{u,v\}^{2p-1}$ must have at least one occurrence of uv, and so we can write $a = bc$ where $b \in \{u,v\}^{d-5(2p-1)}$ and $c \in \{u,v\}^{10p-5}$. Then c has at least five occurrences of uv and so cz is not a factor of an element of $\{u,v\}^\omega$ by Proposition 5.1. In particular, cz is not a factor of a. Since cz is a factor of w of length at most

$$(10p-5)\max(|u|,|v|) + |z| \leq 10p\max(|u|,|v|),$$

we see that cz must occur in a prefix of w of length $10p\kappa\max(|u|,|v|)$. Thus since cz is not a factor of a, we must have that

$$|a| < 10p\kappa\max(|u|,|v|).$$

But $a \in \{u,v\}^d$ has no occurrences of u^p or v^p, and hence each factor of length d of a (when viewed as a word over $\{u,v\}$) must contain at least one occurrence of u and at least one occurrence of v. Then a has a prefix of the form $a_1 \cdots a_{\lfloor d/p \rfloor}$ with each $a_i \in \{u,v\}^p$ and so a must have at least $\lfloor d/p \rfloor$ copies of u and at least $\lfloor d/p \rfloor$ copies of v. Hence

$$|a| \geq (d/p - 1)\max(|u|,|v|) \geq (D-p)\max(|u|,|v|)/p.$$

In particular, $10p\kappa\max(|u|,|v|) > (D-p)\max(|u|,|v|)/p$, and so $D < 10p^2\kappa + p$, a contradiction. □

6. Proof of Theorem 1.1

We now give the decision procedure that makes up the content of Theorem 1.1; namely, we show how to decide whether there exist finite words u and v such that $\mathbf{x} \in \{u,v\}^\omega$, when $\mathbf{x}$ is a k-automatic sequence. The procedure is divided into two cases, which depend upon whether one of the words u or v has arbitrarily large powers occurring as factors of $\mathbf{x}$. The former case is dealt with *via* using the following lemma and proposition.

Lemma 6.1. *Let* $k \geq 2$ *be an integer and let* $\mathbf{x}$ *be a* k*-automatic sequence. Then there is a computable number* L *such that whenever* u *is a nontrivial factor of* $\mathbf{x}$ *with the property that there exists a prefix* v *of* $\mathbf{x}$ *with* $|v| \geq |u|$ *such that*

1. u *is not a prefix nor suffix of* v,
2. v *does not occur in* $\mathbf{x}$ *with unbounded exponent,*
3. v *is not a factor of* u^ω,
4. *there exist* $p_1, \ldots, p_L \geq 0$ *with the property that* $vu^{p_1}vu^{p_2}v\cdots vu^{p_L}$ *is a prefix of* $\mathbf{x}$,

we necessarily have $\mathbf{x} \in \{u,v\}^\omega$.

Proof. We recall that by a result of Cobham [9] there is a computable number $\kappa = \kappa(\mathbf{x})$ such that every factor of $\mathbf{x}$ of length N occurs in the prefix of $\mathbf{x}$ of length κN. Theorem 1.2 gives that there is a computable number

$p = p(\mathbf{x})$ such that every factor of $\mathbf{x}$ has the property that it either occurs in $\mathbf{x}$ with exponent at most p or it occurs with unbounded exponent.

We take $L = (15p+4)\kappa$ and we claim that if there exist factors u and v of $\mathbf{x}$ with $|v| \geq |u|$ such that v is a prefix of $\mathbf{x}$ and hypotheses 1–4 hold then $\mathbf{x} \in \{u, v\}^\omega$.

To see this, suppose towards a contradiction that $\mathbf{x} \notin \{u, v\}^\omega$. Then after possibly enlarging L, we may assume that $vu^{p_1}vu^{p_2}v \cdots vu^{p_L}$ is a prefix of $\mathbf{x}$ but neither $vu^{p_1}vu^{p_2}v \cdots vu^{p_L+1}$ nor $vu^{p_1}vu^{p_2}v \cdots vu^{p_L}v$ are prefixes of $\mathbf{x}$. Then there is some word z with $|z| = |v|$ such that $vu^{p_1}vu^{p_2}v \cdots vu^{p_L}z$ is a prefix of $\mathbf{x}$ and neither u nor v is a prefix of z.

To complete the proof, we now look at cases.

Case I. For each $i \in \{L - 5p, \ldots, L\}$, we have $|p_i| \cdot |u| \leq |2v|$.

In this case, by hypothesis 2, v^p is not a factor of $\mathbf{x}$. It follows that for $j = 1, \ldots, 5$ there is at least one $i \in \{L - jp, \ldots, L - (j-1)p + 1\}$ such that $p_i \neq 0$ since otherwise

$$vu^{p_{L-5p}}v \cdots vu^{p_L}$$

would have v^p as a factor. In particular,

$$vu^{p_{L-5p}}v \cdots vu^{p_L}$$

has at least five occurrences of vu, and so by Proposition 5.1, $vu^{p_{L-5p}}v \cdots vu^{p_L}z$ is not a factor of a word in $\{u, v\}^\omega$. Notice that the length of

$$y := vu^{p_{L-5p}}v \cdots vu^{p_L}z$$

is at most $|v|(15p+4)$, since each u^{p_j} factor has length at most $2|v|$ and $|z| = |v|$. Then by Cobham's result [9] this word y must occur in a prefix of $\mathbf{x}$ of length $\kappa|v|(15p+4)$. But since $vu^{p_1}vu^{p_2}v \cdots vu^{p_L}$ has length at least $L|v| = \kappa|v|(15p+4)$, we see that this cannot be the case.

Case II. There is some $i \in \{L - 5p, \ldots, L\}$ such that $|p_i| \cdot |u| > |2v|$.

In this case, there is some maximal i in this interval with this property, and we let j denote this index. Then there is some $q \leq p_j$ such that $q|u| > 2|v| \geq (q-1)|u|$, and so $|u|^q \leq 2|v| + |u| \leq 3|v|$. Then we consider the suffix $y' := u^q vu^{p_{j+1}}v \cdots vu^{p_L}z$ of $vu^{p_1}vu^{p_2}v \cdots vu^{p_L}z$. Notice that $|y'| \leq 3|v|(1 + L - j) + |v| \leq 3|v|(5p+1)$. Since $vu^{p_1}vu^{p_2}v \cdots vu^{p_L}$ has length at least $L|v| > \kappa|v|(15p+3)$, we see by Cobham's theorem [9] that y' must be a factor of $vu^{p_1}vu^{p_2}v \cdots vu^{p_L+1}$. That is, there are words a and b such that $ay'b = \xi_1 \cdots \xi_t$ with each $\xi_i \in \{u, v\}$. Since $u \not\leq v$ and $v \not\leq u$, we may assume that neither a nor b is trivial, neither u nor v is a prefix of a, and neither u nor v is a suffix of b and we may assume that a is shorter than the length of ξ_1 and that b is shorter than the length of ξ_t. Then since $|u|^q > 2|v| \geq 2|u|$, we must have that ξ_2 is a factor of u^q. By assumption v is not a factor of u^ω and so ξ_2 must be u. But now $\xi_2 = u$ and so u is in fact a factor of u^2. Since u is primitive, we know (see, *e.g.*, [8, p. 336]) that if $a'ub' = u^2$, then either $a' = \epsilon$ or $b' = \epsilon$, and so we see that $\xi_1 = au^i$ for some $i \geq 0$. But now if $\xi_1 = v$, then u is a prefix of v, which is not allowed, and if $\xi_1 = u$ either $a = \epsilon$ or $a = u$, neither of which is allowed.

This completes the proof. □

Lemma 6.2. *The following problem is decidable: given an integer $k \geq 2$, a k-automatic word $\mathbf{x}$, and two finite words $r, s \in \Sigma^+$, determine whether $\mathbf{x} \in \{r, s\}^\omega$.*

Proof. Given words r, s we can easily find a finite automaton $A = (Q, \Sigma, q_0, \delta, F)$ recognizing the language $\{r, s\}^*$. We now regard A as a (uniform) finite-state transducer T where each transition $\delta(p, a) = q$ has the output q associated with it. Hence on input $a_1 a_2 \cdots a_t$ the transducer T outputs the sequence of states

$$\delta(q_0, a_1)\delta(q_0, a_1 a_2) \cdots \delta(q_0, a_1 a_2 \cdots a_t).$$

By a theorem of Cobham ([9] or [1], Thm. 6.9.2), automatic sequences are (computably) closed under uniform transductions, so we can compute a k-DFAO M' that computes the k-automatic sequence $\mathbf{y} := T(\mathbf{x})$. Then $\mathbf{x} \in \{r, s\}^*$ if and only if $\mathbf{y}$ contains infinitely many occurrences of states of F. This is first-order expressible, and hence decidable. □

Proposition 6.1. *Let $k \geq 2$ be an integer, let $\mathbf{x}$ be a k-automatic word, and let u be a primitive factor of $\mathbf{x}$ with the property that $\mathrm{Fac}(u^\omega) \subseteq \mathrm{Fac}(\mathbf{x})$. Then there is a decision procedure that determines whether there is a word v such that $\mathbf{x} \in \{u, v\}^\omega$.*

Proof. We recall that by Corollary 4.4 there is a finite set $\{w_1, \ldots, w_r\}$ of primitive factors of $\mathbf{x}$ that occur with unbounded exponents, which we can explicitly determine. Then by assumption $u \in \{w_1, \ldots, w_r\}$.

By Lemma 6.2, we may decide in the case that $v = w_j$ for some j and when $|v| \leq |u|$.

Hence it suffices to deal with the case when $v \notin \{w_1, \ldots, w_r\}$ and $|v| > |u|$. Moreover, by removing a prefix of $\mathbf{x}$ of the form u^i, we may assume without loss of generality that u is not a prefix of $\mathbf{x}$; and we may assume without loss of generality that u is neither a prefix nor a suffix of v and that v is primitive.

Since v is a prefix of $\mathbf{x}$ and since there is a unique longest prefix of $\mathbf{x}$ that is in $\mathrm{Fac}(u^\omega)$, we can decide whether there exists $v \in \mathrm{Fac}(u^\omega)$ such that $\mathbf{x} \in \{u, v\}^\omega$.

Thus we may assume, in addition to the other assumptions given, that v is not a factor of u^ω. It follows from Lemma 6.1 that there is a computable number $L = L(\mathbf{x})$ such that if there exist $p_1, \ldots, p_L \geq 0$ with the property that $vu^{p_1}vu^{p_2} \cdots vu^{p_L}$ is a prefix of $\mathbf{x}$, then $\mathbf{x} \in \{u, v\}^\omega$. By Proposition 4.5, it is decidable whether $\mathbf{x}$ has a prefix of the form $vu^{p_1}vu^{p_2} \cdots vu^{p_L}$ for some v having the desired constraints and some choice of $p_1, \ldots, p_L$, and so we are done. □

Proof of Theorem 1.1. We give the steps in the algorithm, which determines whether the rank of $\mathbf{x}$ is two. Since the property of being periodic is decidable, we may assume that the rank of $\mathbf{x}$ is at least two. We note that if $\mathbf{x}$ is of rank two, then there exist words u and v such that $\mathbf{x} \in \{u, v\}^\omega$; then by picking such (u, v) with $|u| + |v|$ minimal, we may assume without loss of generality that the assumptions from Notation 5.1 hold.

Step 1. Using Theorem 1.2, compute $p = p(\mathbf{x})$ such that for every u with the property that u^p is a factor of $\mathbf{x}$ we have $\mathrm{Fac}(u^\omega) \subseteq \mathrm{Fac}(\mathbf{x})$.

Step 2. By Corollary 4.4, there is a finite computable set of primitive words $\{w_1, \ldots, w_r\}$ such that $\mathrm{Fac}(w_i^\omega) \subseteq \mathrm{Fac}(\mathbf{x})$ for $i = 1, \ldots, r$.

Step 3. Use the decision procedure from Proposition 6.1 to decide whether there exists a word u such that $\mathbf{x} \in \{w_i, u\}^\omega$ for some u and some $i \in \{1, \ldots, r\}$. If such a u exists, the algorithm halts and returns that $\mathbf{x}$ has rank two; if no such u exists, we go to the next step.

Step 4. It now suffices to decide whether there exist words u, v such that $\mathbf{x}$ is in $\{u, v\}^\omega$ and the assumptions of Notation 5.1 apply to u and v. By Step 3, we can also assume that $u, v \notin \{w_1, \ldots, w_r\}$, where $w_1, \ldots, w_r$ are as in Step 3. Thus u^p and v^p are not factors of $\mathbf{x}$. Then compute the integer D given in the statement of Lemma 5.4.

Step 5. For each of the 2^D binary words y of length D, use Proposition 4.6 to determine whether there exist u and v such that $y(u, v)$ is a prefix of $\mathbf{x}$, where $y(u, v)$ is the word in $\{u, v\}^*$, obtained by applying the coding sending 0 to u and 1 to v, to the binary word y; if there is some binary word for which this holds then $\mathbf{x}$ has rank two by Lemma 5.4 and we stop; if this does not hold for these words, then $\mathbf{x}$ has rank at least three and we stop.

□

Acknowledgements. We thank the anonymous referees for many helpful comments and suggestions.

References

[1] J.-P. Allouche and J. Shallit, Automatic Sequences: Theory, Applications, Generalizations. Cambridge University Press (2003).

[2] J. Barwise, An introduction to first-order logic. In *Handbook of Mathematical Logic*, edited by J. Barwise. North-Holland (1977), pp. 5–46.

[3] J. Bell, E. Charlier, A. Fraenkel and M. Rigo, A decision problem for ultimately periodic sets in non-standard numeration systems. *Internat. J. Algebra Comput.* **19** (2009) 809–839.

[4] J.P. Bell and J. Shallit, Lie complexity of words. Available at https://arxiv.org/abs/2102.03821 (2021).

[5] V. Bruyère, G. Hansel, C. Michaux and R. Villemaire, Logic and p-recognizable sets of integers. *Bull. Belgian Math. Soc.* **1** (1994) 191–238. Corrigendum, *Bull. Belg. Math. Soc.* **1** (1994) 577.

[6] E. Charlier, A. Massuir, M. Rigo and E. Rowland, Ultimate periodicity problem for linear numeration systems. Preprint at https://arxiv.org/abs/2007.08147 (2020).

[7] E. Charlier, N. Rampersad and J. Shallit, Enumeration and decidable properties of automatic sequences. *Internat. J. Found. Comp. Sci.* **23** (2012) 1035–1066.

[8] C. Choffrut and J. Karhumäki. Combinatorics of words. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, Vol. 1, pp. 329–438. Springer-Verlag, 1997.

[9] A. Cobham, Uniform tag sequences. *Math. Systems Theory* **6** (1972) 164–192.

[10] K. Culik II, An aperiodic set of 13 Wang tiles. *Discrete Math.* **160** (1996) 245–251.

[11] F. Durand, Decidability of the HD0L ultimate periodicity problem. *RAIRO: ITA* **47** (2013) 201–214.

[12] A. Ehrenfeucht, J. Karhumäki and G. Rozenberg, The (generalized) Post correspondence problem with lists consisting of two words is decidable. *Theoret. Comput. Sci.* **21** (1982) 119–144.

[13] A. Ehrenfeucht and G. Rozenberg, Repetition of subwords in D0L languages. *Inform. Comput.* **53** (1983) 13–35.

[14] D. Goc, L. Schaeffer and J. Shallit, The subword complexity of k-automatic sequences is k-synchronized. In *DLT 2013*, Vol. 7907 of *Lecture Notes in Computer Science*, edited by M.-P. Béal and O. Carton. Springer-Verlag (2013) 252–263.

[15] T. Harju and M. Linna, On the periodicity of morphisms on free monoids. *RAIRO: ITA* **20** (1986) 47–54.

[16] J. Honkala, A decision method for the recognizability of sets defined by number systems. *RAIRO: ITA* **20** (1986) 395–403.

[17] J.E. Hopcroft and J.D. Ullman, Introduction to Automata Theory, Languages, and Computation. Addison-Wesley (1979).

[18] E. Jeandel and M. Rao, An aperiodic set of 11 Wang tiles. Preprint https://arxiv.org/abs/1506.06492 (2015).

[19] J. Kari, A small aperiodic set of Wang tiles. *Discrete Math.* **160** (1996) 259–264.

[20] K. Klouda and Š. Starosta, An algorithm for enumerating all infinite repetitions in a D0L-system. *J. Discrete Algor.* **33** (2015) 130–138.

[21] K. Klouda and Š. Starosta, Characterization of circular D0L-systems. *Theoret. Comput. Sci.* **790** (2019) 131–137.

[22] K. Klouda and Š. Starosta, Repetitiveness of HD0L-systems. Unpublished manuscript (2021).

[23] Y. Kobayashi and F. Otto, Repetitiveness of languages generated by morphisms. *Theoret. Comput. Sci.* **240** (2000) 337–378.

[24] B. Lando, Periodicity and ultimate periodicity of D0L systems. *Theoret. Comput. Sci.* **82** (1991) 19–33.

[25] J. Leroux, A polynomial time Presburger criterion and synthesis for number decision diagrams. In *20th IEEE Symposium on Logic in Computer Science (LICS 2005)*. IEEE Press (2005) 147–156.

[26] M. Linna, On periodic ω-sequences obtained by iterating morphisms. *Ann. Univ. Turku. Ser. A I* **186** (1984) 64–71.

[27] R.C. Lyndon and M.P. Schützenberger, The equation $a^M = b^N c^P$ in a free group. *Michigan Math. J.* **9** (1962) 289–298.

[28] V. Marsault and J. Sakarovitch, Ultimate periodicity of b-recognisable sets: a quasilinear procedure. In M. P. Béal and O. Carton, editors, *Developments in Language Theory, 17th International Conference, DLT 2013*, Vol. 7907 of *Lecture Notes in Computer Science*. Springer-Verlag (2013) 362–373.

[29] Y. Matiyasevich and G. Sénizergues, Decision problems for semi-Thue systems with a few rules. *Theoret. Comput. Sci.* **330** (2005) 145–169.

[30] F. Mignosi and P. Séébold, If a D0L language is k-power free then it is circular. In A. Lingas, R. Karlsson and S. Carlsson, editors, *Proc. 20th Int'l Conf. on Automata, Languages, and Programming (ICALP)*, Vol. 700 of *Lecture Notes in Computer Science* (1993) 507–518.

[31] J.-J. Pansiot, Decidability of periodicity for infinite words. *RAIRO: ITA* **20** (1986) 43–46.
[32] E.L. Post, A variant of a recursively unsolvable problem. *Bull. Amer. Math. Soc.* **52** (1946) 264–268.