Delft University of Technology

Hidden automatic sequences

Allouche, Jean Paul; Dekking, Michel; Queffélec, Martine

Authors
Allouche, Jean-Paul
Dekking, Michel
Queffélec, Martine

HIDDEN AUTOMATIC SEQUENCES

Jean-Paul Allouche[1], Michel Dekking[2], and Martine Queffélec[3]

[1]*CNRS, IMJ-PRG, UPMC, 4 Place Jussieu, F-75252 Paris Cedex 05, France*
jean-paul.allouche@imj-prg.fr
[2]*Delft University of Technology, Faculty EEMCS, P.O. Box 5031, 2600 GA Delft, The Netherlands*
f.m.dekking@math.tudelft.nl
[3]*Université Lille 1, UMR 8524, F-59655 Villeneuve d'Ascq Cedex, France*
martine@math.univ-lille1.fr

Abstract. An automatic sequence is a letter-to-letter coding of a fixed point of a uniform morphism. More generally, morphic sequences are letter-to-letter codings of fixed points of arbitrary morphisms. There are many examples where an, a priori, morphic sequence with a *non-uniform* morphism happens to be an automatic sequence. An example is the Lysënok morphism $a \to aca$, $b \to d$, $c \to b$, $d \to c$, the fixed point of which is also a 2-automatic sequence. Such an identification is useful for describing the dynamical systems generated by the fixed point. We give several ways to uncover such hidden automatic sequences, and present many examples. We focus in particular on morphisms associated with Grigorchuk groups.

Keywords. Morphic sequences, automatic sequences, Grigorchuk groups

Mathematics Subject Classifications. 11B85, 68R15, 37B10

1. Introduction

The purpose of this paper is to describe how to detect whether a sequence given as a fixed point of a *substitution* on some *alphabet* is *automatic* (the words in italics will be defined a few lines below). Our starting point is the *substitution* $a \to aca$, $b \to d$, $c \to b$, $d \to c$ used by Lysënok [40] to provide a presentation by generators and (infinitely many) defining relations of the first Grigorchuk group. More recently Vorobets [48] proved several properties of the fixed point of this substitution. In an unpublished 2011 note the first and third authors proved among other things that the fixed point of this substitution is also the fixed point of the 2-substitution $a \to ac$, $b \to ad$, $c \to ab$, $d \to ac$, and so that this fixed point is 2-automatic [3]. This result was obtained again more recently in [27, 28], also see [7].

But, let us first quickly recall some definitions from combinatorics on words that are used above or in the sequel (for more details, see [4, 14, 23, 31, 37, 38, 39, 45]).

- An *alphabet* is a finite set of elements that are called *letters*. A finite sequence of letters $(a_1, a_2, \ldots, a_r)$ on the alphabet A is called a *word of length r on A* and denoted $a_1 a_2 \ldots a_r$: we write $|a_1 a_2 \ldots a_r| := r$. If the sequence is empty, it is called *the empty word* and its length is 0. The set of all words, including the empty word, is denoted A^*. The *concatenation* of the words $a_1 a_2 \ldots a_r$ and $b_1 b_2 \ldots b_s$ on the same alphabet is the word $a_1 a_2 \ldots a_r b_1 b_2 \ldots b_s$ of length $r + s$. The set A^* equipped with concatenation is a (free) monoid.

- Given two alphabets A and B, a *morphism* (also called *substitution*) φ from A^* to B^* is a map from $A^* \to B^*$ that preserves concatenation, i.e., such that, for all $(v, w) \in A^* \times A^*$, one has $\varphi(vw) = \varphi(v)\varphi(w)$. A morphism from A^* to itself is called an *(endo)morphism of A^** (or by abuse of notation a *morphism on A*). If the images of all letters by a morphism φ have the same length (q), the morphism is called *(q-)uniform* or *of constant length (q)*.

- Let φ be a morphism of A^*, with $A := \{a_0, a_1, \ldots, a_{r-1}\}$. Its *incidence matrix* is the matrix $M := (m_{i,j})$, where $m_{i,j}$ is the number of letters a_i in the word $\varphi(a_j)$, for all i and j in $\{0, 1, \ldots, r-1\}$. The *length vector* of φ of A^* is the vector $L := (L_0, L_1, \ldots, L_{r-1})$, where $L_j := |\varphi(a_j)|$, i.e., the sum of the entries of the column indexed by j of the incidence matrix of φ.

- A morphism φ of A^* can be extended to $A^{\mathbb{N}}$, the set of infinite sequences on A, by setting $\varphi(u_0 u_1 \ldots u_r \ldots) := \varphi(u_0)\varphi(u_1) \ldots \varphi(u_r) \ldots$. If φ is a morphism of A^* to itself, and if there exist a letter a in A and a word v in A^* such that $\varphi(a) = av$, and $|\varphi^k(a)| \to \infty$ as $k \to \infty$, then there exists a unique infinite sequence admitting all the $\varphi^k(a)$ as prefixes. This sequence is a fixed point of (the extension to infinite sequences of) φ. It is called an *(iterative) fixed point*[1] of φ. If a sequence is a fixed point of some morphism it is called *purely morphic*.

- If A and B are two alphabets and π is a map from A to B, this map can be viewed also as a uniform morphism (of length 1) from A^* to B^*. If φ is a morphism from A^* to itself, and $(u_n)_{n \geqslant 0}$ is a fixed point of φ, the sequence $(\pi(u_n))_{n \geqslant 0}$ is said to be *morphic*. If furthermore the morphism φ has constant length q, the sequence $(\pi(u_n))_{n \geqslant 0}$ is said to be *uniformly morphic* or *q-automatic*.

Coming back to the phenomenon described above for the fixed point of the Lysënok morphism, one can ask whether it often happens that a sequence defined as a fixed point of a non-uniform morphism is also automatic. Actually such a phenomenon is rare, but was already encountered. One example is the proof by Berstel [13] that the Istrail squarefree sequence [35], defined as the unique fixed point of the morphism σ_{IS}, given by

$$\sigma_{\mathrm{IS}}(0) := 12, \ \sigma_{\mathrm{IS}}(1) := 102, \ \sigma_{\mathrm{IS}}(2) := 0,$$

---

[1] From now on, we will write *fixed point* instead of *iterative fixed point*. But we will keep in mind that the definition, implying iterating a morphism, rules out "trivial fixed points" not obtained by iteration, e.g., any sequence with values 0 and 1 is a "trivial fixed point" of the morphism $0 \to 0$, $1 \to 1$, $2 \to 11012$.

can be obtained also as the letter-to-letter image by the reduction modulo 3 of the fixed point beginning with 1 of the uniform morphism $0 \to 12$, $1 \to 13$, $2 \to 20$, $3 \to 21$.

This phenomenon is also interesting since substitutions of constant length d are "simpler" than general substitutions in particular, because they are related to d-ary expansions of the indexes of their terms.

In view of what precedes, a natural question arises: *how to recognize that a fixed point of a non-uniform morphism is an automatic sequence?*

Note.   To our knowledge, whether a fixed point of a non-uniform morphism is an automatic sequence is not known to be decidable. This question is very probably undecidable.

Of course not every fixed point of a non-uniform morphism is q-automatic for some q: a property of automatic sequences is that, if the frequency of a letter exists, then it must be a rational number [17]. Thus the Fibonacci binary sequence (i.e., the fixed point of the morphism $0 \to 01$, $1 \to 0$) is not q-automatic for any $q \geqslant 2$, since the frequencies of its letters are not rational. However, it is true that any q-automatic sequence ($q \geqslant 2$) can be obtained as a non-uniformly morphic sequence, i.e., as the letter-to-letter image of a fixed point of a non-uniform morphism [5, Theorem 11].

In Section 2 we will revisit a 1978 theorem of the second author to give a sufficient condition for a fixed point of a non-uniform morphism to be automatic. This is Theorem 2.2 below. Section 3 will show an interplay between this Theorem 2.2 and the result [5, Theorem 11] cited above. In Section 4, we will apply Theorem 2.2 (more precisely a particular case, the "Anagram Theorem") to prove that several examples of sequences in the OEIS [43] defined as fixed points of non-uniform morphisms, are actually automatic. We will recall the 2-automaticity of the fixed point of the Lysënok morphism in Section 5, and give several examples of sequences related to Grigorchuk groups and similar groups. Finally an Appendix by the second author will give a simpler proof of the main result in [5].

2. A general theorem revisited

Note that the vector of lengths of the Istrail morphism σ_{IS}, which is $(2, 3, 1)$, happens also to be a left eigenvector of the incidence matrix of the morphism. So Berstel's result also follows from [18, Section V, Theorem 1], as noted as an example in the same paper [18, Section IV, Example 8][2]. Since this theorem is stated in [18] in the context of dynamical systems, we will give an equivalent reformulation in Theorem 2.2 below. Before stating the theorem, we need a lemma on nonnegative matrices, which does not use any result à la Perron–Frobenius: see, e.g., the proof in [34, Corollary 8.1.30, p. 522].

Lemma 2.1. *Let M be a matrix whose entries are all nonnegative. If v is an eigenvector of M with positive coordinates associated with a real eigenvalue λ, then λ is equal to the spectral radius of M.*

---

[2]The matrix there is the transpose of what is now considered to be the incidence matrix of a morphism; see our definition above.

We invite the reader to verify that the statement in this lemma is not true if v is supposed only to have nonnegative coordinates.

Theorem 2.2. ([18]) *Let σ be a morphism on $\{0, \ldots, r-1\}$ for some integer $r > 1$. Let $L := (|\sigma(0)|, \ldots, |\sigma(r-1)|)$ be the length vector of σ. Suppose that σ is non-erasing (i.e., for all i in $\{0, \ldots, r-1\}$, one has $|\sigma(i)| \geqslant 1$). Let x be a fixed point of σ, and let M be the incidence matrix of σ. If L is a left eigenvector of M, then x is q-automatic, where q is the spectral radius of M.*

We sketch a proof of this result, which will be useful in the sequel. Let $L_i := |\sigma(i)|$ be the length of $\sigma(i)$ for $i \in \{0, \ldots, r-1\}$. The idea is to define a morphism τ on an alphabet of $L_0 + \cdots + L_{r-1}$ symbols (i, j), $0 \leqslant i < r$, $1 \leqslant j \leqslant L_i$ by setting

$$\tau(i, j) := (i^*, 1) \ldots (i^*, L_{i^*}) \quad \text{if } \sigma(i)_j = i^*.$$

If σ is non-uniform, then τ is still non-uniform, but the rigid way in which the symbols (i, j) occur permits one to 'reshuffle' τ to a morphism τ' which is uniform, and the eigenvector criterion ensures that this can be done consistently. Here 'reshuffling' means that we write the concatenation $\tau(i, 1) \ldots \tau(i, L_i)$ as a concatenation of L_i words of length q, so that $\tau'(i, j)$ is the j-th such word. Rather than going into the details, we illustrate the argument with the Istrail morphism σ_{IS}. Here the alphabet is $\{(0, 1), (0, 2), (1, 1), (1, 2), (1, 3), (2, 1)\}$. We obtain

$$\begin{aligned}
&\tau_{\text{IS}}(0, 1) = (1, 1)(1, 2)(1, 3), \ \tau_{\text{IS}}(0, 2) = (2, 1) \\
&\tau_{\text{IS}}(1, 1) = (1, 1)(1, 2)(1, 3), \ \tau_{\text{IS}}((1, 2)) = (0, 1)(0, 2), \ \tau_{\text{IS}}(1, 3) = (2, 1) \\
&\tau_{\text{IS}}(2, 1) = (0, 1)(0, 2).
\end{aligned}$$

Coding $a := (0, 1), b := (0, 2), c := (1, 1), d := (1, 2), e := (1, 3), f := (2, 1)$, the reshuffled τ'_{IS} is given by

$$\tau'_{\text{IS}}(a) = cd, \ \tau'_{\text{IS}}(b) = ef, \ \tau'_{\text{IS}}(c) = cd, \ \tau'_{\text{IS}}(d) = ea, \ \tau'_{\text{IS}}(e) = bf, \ \tau'_{\text{IS}}(f) = ab.$$

The letter-to-letter projection π is given by $a \to 1, b \to 2, c \to 1, d \to 0, e \to 2, f \to 0$. This gives the Istrail sequence as a 2-automatic sequence by projection of a fixed point of the uniform morphism τ'_{IS} on a six-letter alphabet. But, since $\tau'_{\text{IS}}(a) = \tau'_{\text{IS}}(c)$, and $\pi(a) = \pi(c)$, we can merge a and c. Finally, since $\pi(b) = \pi(e)$, and the first letters of $\tau'_{\text{IS}}(b)$ and $\tau'_{\text{IS}}(e)$ are b and e, and the second letters are equal, also b and e can be merged. After a recoding, this gives Berstel's morphism above.

Let q be the constant length of the morphism τ'. We show in general why q is equal to the spectral radius of M. Namely, let λ be the eigenvalue associated with the left eigenvector $(L_0, L_1, \ldots, L_{r-1})$, and let $r' := L_0 + \cdots + L_{r-1}$. Then

$$qr' = \sum_i \sum_j |\tau'(i, j)| = \sum_i \sum_j |\tau(i, j)| = \sum_i (LM)_i = \lambda \sum_i L_i = \lambda r'.$$

This implies that q has to be equal to λ. But, from Lemma 2.1 above, λ must be equal to the spectral radius of M. $\square$

Remark 2.3. The condition about the length vector given in Theorem 2.2 is not necessary. The Lysënok morphism $a \to aca$, $b \to d$, $c \to b$, $d \to c$ is an example, as follows directly from Proposition 2.4 below. Another example is given in Corollary 5.2.

Proposition 2.4. *Let μ be a non-uniform morphism with incidence matrix M. If M is invertible, then the length vector of μ cannot be a left eigenvector of M.*

Proof. Let us write $\mathbf{1} := (1, 1, \ldots, 1)$. If the length vector L of μ is a left eigenvector of M, one has $LM = \lambda L$ for some eigenvalue λ. But, by definition of L, $L = \mathbf{1} \cdot M$. Hence $(\lambda \mathbf{1} - L)M = 0$. Since M is invertible, this implies $\lambda \mathbf{1} = L$, hence $L_i = \lambda$ for all i, but the morphism was supposed to be non-uniform. □

Remark 2.5. One might ask whether the fact that the spectral radius of the incidence matrix in Theorem 2.2 is an integer is *a priori* clear: it is a well-known application of the Gauss lemma that *a root of a monic polynomial with integer coefficients is either irrational or it is an integer* (see, e.g., [32, Theorem 45, p. 41] for a classical proof; also see [24] for an interesting less classical proof).

3. From q-automatic to non-uniformly morphic and back

The paper [5] gives an algorithm to represent any q-automatic sequence with associated morphism γ as a morphic sequence, where the morphism γ' is *non*-uniform. We call this algorithm the CUP-algorithm, standing for Create Unique Pair. The question arises: if we are given this non-uniform representation, how do we find the uniform representation? The answer lies, once more, directly in the left eigenvector criterion.

In the following we use the short version of the CUP-algorithm as given in Section 7.

We first give an example. We start with a famous 2-automatic sequence: the Thue–Morse sequence. Since Theorem 7.1 requires $q \geqslant 3$, we take for γ the square of the Thue–Morse morphism:

$$\gamma(0) = 0110, \quad \gamma(1) = 1001.$$

The Thue–Morse sequence is the iterative fixed point of γ starting with $a = 0$, and the letters $b = 1$, and $c = 1$ are the two letters which give two extra letters b' and c' in the CUP algorithm. We define morphisms γ' on an extended alphabet $\{0, 1, b', c'\}$, where b' and c' will be projected on 1. Define the two non-empty words z and t as any two words of which the concatenation gives

$$zt = \gamma(01) = 01101001,$$

for example $z = 0, t = 1101001$. Then define γ' on $\{0, 1, b', c'\}$ by

$$\gamma'(0) := 0b'c'0, \ \gamma'(1) := \gamma(1), \ \gamma'(b') := z, \ \gamma'(c') := t.$$

As in the proof of Theorem 7.1 it is easy to see that the infinite fixed point of γ' starting with 0 maps to the Thue–Morse sequence under the projection D given by $D(0) := 0, D(1) := 1$, $D(b') := 1, D(c') := 1$.

The incidence matrix of these morphisms is

$$M' := \begin{pmatrix} 2 & 2 & m_0 & 4 - m_0 \\ 0 & 2 & m_1 & 4 - m_1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

where m_0 is the number of 0's in z, and m_1 is the number of 1's in z. Let $L' = (4, 4, |z|, |t|)$ be the length vector of γ'. Then the following holds for any choice of z and t:

$$L'M' = 4L'.$$

This is exactly the left eigenvector criterion of Theorem 2.2. The general result is the following theorem.

Theorem 3.1. *Let x be a q-automatic sequence, and let γ' be the non-uniform morphism turning x into a (non-uniformly) morphic sequence in the short version of the CUP algorithm. Then the incidence matrix of γ' satisfies the left eigenvector criterion.*

Proof. Let γ be the uniform morphism of length q on the alphabet $\{0, 1, \ldots, r-1\}$ such that x is a letter-to-letter projection of a fixed point y of γ. It is easy to see that, as in the proof of Theorem 7.1, we may suppose that $y = x$. We may also suppose that $q \geqslant 3$. Let $L = (q, q, \ldots, q)$ be the length vector of γ, and let M be the incidence matrix of γ. Note that M satisfies the eigenvector criterion: $LM = qL$. Let b and c be the second and third letter of $\gamma(0)$, and let $b' = r$ and $c' = r + 1$ be the two extra letters in the CUP algorithm. Let $|z|$ and $|t|$ be the lengths of z and t in the CUP splitting $\gamma(bc) = zt$. Then $\gamma'(r) = z$ and $\gamma'(r + 1) = t$.

The length vector L' of γ' is equal to

$$L' = (q, q, \ldots, q, |z|, |t|).$$

The first column of M' is equal to

$$(m_{00} + b + c - 2, m_{10} - b - c, m_{20}, m_{30}, \ldots, m_{r-1,0}, 1, 1)^{\top}.$$

The inner product of the length vector L' with this first column is equal to

$$q(m_{00} + b + c - 2) + q(m_{10} - b - c) + qm_{20} + \cdots + qm_{r-1,0} + |z| + |t| =$$
$$q(m_{00} + m_{10} + m_{20} + \cdots + m_{r-1,0}) - 2q + 2q = q^2.$$

Obviously the inner product of L' with the second through r^{th} column is also equal to q^2. The inner product of L' with the $(r + 1)^{\text{th}}$ column is equal to $q|z|$, and the inner product of L' with the $(r + 2)^{\text{th}}$ column is equal to $q|t|$. This finishes the checking of the left eigenvector criterion $L'M' = qL'$. $\qquad\square$

4. First examples of hidden automatic sequences

We start this section with the following *Rank 1 Theorem*, which is a linear algebra result for matrices with nonnegative entries. The interest of this theorem is that it implies a special case of Theorem 2.2, the *Anagram Theorem* (Theorem 4.2 below), that permits one to prove that some fixed points of non-uniform morphisms are automatic in a purely "visual" (but rigorous) way.

Theorem 4.1. ["Rank 1 Theorem"] *Let M be a matrix whose entries are all nonnegative. Suppose that the vector L of the column sums of M is positive. If M has rank 1, then L is a left eigenvector of M, associated with an eigenvalue that is equal to the spectral radius of M.*

Proof. If M has rank 1, then there exist vectors k and r such that $M = k^T r$. Since M is nonnegative, k and r can also be supposed to be non-negative. Let $t := rk^T$, then t is a positive real number. Let $L = \mathbf{1}M$ be the vector of the column sums. Then

$$LM = \mathbf{1}Mk^T r = \mathbf{1}k^T rk^T r = \mathbf{1}k^T tr = t\mathbf{1}k^T r = tL.$$

Now Lemma 2.1 implies that t is the spectral radius of M. $\qquad\qquad\square$

A combination of Theorem 4.1 and Theorem 2.2 yields the following result.

Theorem 4.2. ["Anagram Theorem"] *Let $\mathcal{A}$ be a finite set. Let W be a set of anagrams on $\mathcal{A}$ (the words in W are also said to be abelian equivalent; they have the same Parikh vector). Let ψ be a morphism on $\mathcal{A}$ admitting a fixed point, such that the image of each letter is a concatenation of words in W. Then any fixed point of ψ is q-automatic, where q is the quotient of the length of $\psi(w)$ by the length of w, which is the same for all $w \in W$.*

Proof. The conditions imply that the incidence matrix M of ψ has rank 1, and so any fixed point of ψ is q-automatic, for some integer q. To determine q, let k and r be such that $M = k^T r$. The natural choices for k and r are the frequency vector of any word from W, and the vector that counts the number of words from W used to form the ψ words. Now note that

$$|\psi(w)| = \mathbf{1}Mk^T = \mathbf{1}k^T rk^T = |w|rk^T = |w|t,$$

where $t := rk^T$ is the spectral radius of M, as shown in the proof of Theorem 4.1. Moreover, Theorem 4.1 yields that $L = \mathbf{1}M$, is a left eigenvector of M, and so by Theorem 2.2, any fixed point of ψ is q-automatic, where $q = t$. $\qquad\qquad\square$

Example 4.3. Let ψ be the morphism on a three-letter alphabet given by

$$\psi(a) = aabc, \ \psi(b) = bacaaabc, \ \psi(c) = bacabacabaca.$$

By taking the set $W = \{aabc, baca\}$, we see immediately from Theorem 4.2 that the fixed points of ψ are 7-automatic.

Example 4.4. The sequence A285249 from [43] is called the 0-limiting word of the morphism f which maps $0 \to 10$, $1 \to 0101$ on $\{0, 1\}^*$, i.e., A285249 is the fixed point of f^2 starting with 0, where f^2 is given by $f^2(0) = 010110$, $f^2(1) = 100101100101$. The images of 0 and of 1 by f^2 can be respectively written www' and $w'www'ww$, with $w = 01$ and $w' = 10$. Again, Theorem 4.2 gives that the fixed points of f^2 are 9-automatic, which is equivalent to being 3-automatic.

More examples like sequence A285249 are collected in the following corollary to Theorem 4.2.

Corollary 4.5. *The following automaticity properties for sequences in the OEIS hold.*

- *The four sequences* A284878, A284905, A285305*, and* A284912 *are generated by morphisms f, where $f(0)$ and $f(1)$ can be written as concatenations of one, respectively two of the two words $w = 01$ and $w' = 10$. So Theorem 4.2 immediately implies that they are all 3-automatic.*

- *The sequences* A285252, A285255 *and* A285258*, are fixed points of squares of such morphisms, and so they are 9-automatic (hence 3-automatic).*

- *Finally the fact that* A284878 *is 3-automatic easily implies that* A284881 *is 3-automatic.*

Remark 4.6. Other sequences in the OEIS that do not satisfy the hypotheses of Theorem 4.2 can be proved automatic because they satisfy the hypotheses of Theorem 4.1: for example the sequences A285159 and A285162 (replace the morphism given in the OEIS by its square to obtain these two sequences as fixed points of morphisms), A285345, A284775 and A284935 are 3-automatic.

5. Hidden automatic sequences and self-similar groups

The substitution τ defined by $\tau(a) = aca$, $\tau(b) = d$, $\tau(c) = b$, $\tau(d) = c$ was used by Lysënok to provide a presentation by generators and (infinitely many) defining relations of the first Grigorchuk group. Note that this substitution does not satisfy the "left eigenvector criterion". The proof given in [3] relied on the morphism ψ defined by

$$\psi(a) := ac, \ \ \psi(b) := ad, \ \ \psi(c) := ab, \ \ \psi(d) := ac,$$

which satisfies the relation $\tau \circ \psi = \psi \circ \psi$. This relation easily implies that τ and ψ have the same fixed point beginning with a. A similar proof was given in [27].

Another proof (essentially the one in [28] and [7]) uses a non-overlapping-2-block morphism. A *non-overlapping-2-block-morphism* is a morphism that, starting from a sequence $u_0, u_1, u_2, u_3 \ldots$, yields the sequence $(u_0 u_1)(u_2 u_3) \ldots$, on the new "letters" $u_0 u_1$, $u_2 u_3$, $\ldots$. This proof introduces the morphism (coding $ab = 1, ac = 2, ad = 3$)

$$1 \to 23, \ 2 \to 21, \ 3 \to 22$$

(one can see that $1 = ab \rightarrow acad = 23$, etc.). The fixed point $21232122\ldots$ of this new morphism is also the image of the fixed point of τ by a 2-block morphism. Looking at the even-indexed terms and at the odd-indexed terms of the fixed point of τ gives that the Lysënok sequence is also a fixed point of a morphism of constant length 2.

We may ask whether this second approach works in other "similar" situations, i.e., for morphisms related to the Grigorchuk group or Grigorchuk-like groups. Before we address this question, it is worthwhile to give a general result on automatic sequences in terms of "non-overlapping-k-block morphisms".

Theorem 5.1. *Let $q \geqslant 2$ and let $\mathbf{u} = (u(n))_{n \geqslant 0}$ be a sequence with values in $\mathcal{A}$. Then, $\mathbf{u}$ is q-automatic if and only if there exist a positive integer r and a q-uniform morphism μ on $\mathcal{A}^{q^r}$ such that the sequence of q^r-blocks $(u(q^r n), u(q^r n + 1), \ldots, u(q^r n + q^r - 1))_{n \geqslant 0}$, obtained by grouping in $\mathbf{u}$ the terms q^r at a time, is a fixed point of μ.*

Proof. This is essentially Theorem 1 in [17]. $\qquad \square$

Theorem 5.1 is indeed illustrated by the Lysënok fixed point, and by the following example (which, contrary to the Lysënok morphism, is primitive).

Corollary 5.2. *Let σ be the morphism defined by*

$$\sigma: \quad a \rightarrow acaba, \ b \rightarrow bac, \ c \rightarrow cab.$$

Then the fixed point of σ beginning with a is 4-automatic (hence 2-automatic).

Proof. There are only the 2-blocks ac, ab occurring at even positions in the fixed point $x := acabacab\ldots$ of σ. In fact σ induces the following morphism $\sigma^{[2]}$ on non-overlapping-2-blocks:

$$\sigma^{[2]}: \quad ab \rightarrow ac\,ab\,ab\,ac, \ ac \rightarrow ac\,ab\,ac\,ab.$$

The fact that $\sigma^{[2]}$ has constant length 4 implies that x is a 4-automatic sequence, hence a 2-automatic sequence. $\qquad \square$

Another general result will prove useful.

Proposition 5.3. *If the incidence matrix of a primitive non-uniform morphism has an irrational dominant eigenvalue, then a fixed point of this morphism cannot be automatic.*

Proof. Since the morphism is primitive, the frequency of each letter exists, and the vector of frequencies is the unique normalized eigenvector of the matrix for the dominant eigenvalue. If the sequence were automatic, all the frequencies of letters would be rational, which gives a contradiction with the irrationality of the eigenvalue and the fact that the entries of the matrix are integers. $\qquad \square$

We present a list giving the nature (i.e., whether they are automatic or not automatic) of fixed points of morphisms related to Grigorchuk-like groups.

1. The fixed point of the morphism $a \to aba$, $b \to d$, $c \to b$, $d \to c$ (see, e.g., [10, Proposition 5.6]) is 2-automatic (with the same proof as for the fixed point of the Lysënok morphism).

2. The fixed point of the morphism $a \to aca$, $b \to d$, $c \to aba$, $d \to c$ (see [11, Theorem 4.1]) is not automatic. (Namely the matrix of this morphism is primitive and its characteristic polynomial, which is equal to $x^4 - 2x^3 - 2x^2 - x + 2$, clearly has no rational root; the result follows from Proposition 5.3 above.)

3. The fixed point of the morphism $x \to xzy$, $y \to xx$, $z \to yy$ (see [9, Proof of Proposition 4.7]) is not automatic. (Again this is an application of Proposition 5.3 above, since the characteristic polynomial of the (primitive) incidence matrix is equal to $x^3 - x^2 - 2x - 4$ which has no rational root.)

4. The fixed point beginning with 2^* of the cube of the morphism $1 \to 2$, $1^* \to 2^*$, $2 \to 1^*2^*$, $2^* \to 21$ (see [42]) is not automatic. This follows from Proposition 5.3. A more detailed analysis is the following. Put $A := 2^*2$ and $B := 11^*$, then the sequence can be written $ABABAABAABA\ldots$ which is a fixed point of the morphism $A \to ABABA$, $B \to ABA$, that is easily seen to be Sturmian from the criterion [47, Proposition 1.2] since
$$AB \to ABABAABA = ABA(BA)ABA$$
while
$$BA \to ABAABABA = ABA(AB)ABA.$$
Actually a more precise result holds: this morphism is conjugate to f^3 where f is the Fibonacci morphism $A \to AB$, $B \to A$ (see the comments of the second author for the sequences A334413 and A006340 in [43], where the alphabet $\{1, 0\}$ corresponds to our $\{A, B\}$ here).

5. The fixed point of the morphism $a \to aba$, $b \to c$, $c \to b$ (see, e.g., [41, p. 40]) can also be generated by the morphism on the non-overlapping-2-blocks $0 = ab$ and $1 = ac$ defined by $0 \to 01$, $1 \to 00$, i.e., the "period-doubling" morphism, and so this fixed point is 2-automatic.

6. The morphism $a \to b$, $b \to c$, $c \to aba$ (see, [29, Theorem 3.1], also see [41, p. 40]) has the property that its cube has a fixed point. This fixed point is not automatic by an application of Proposition 5.3.

7. The morphism $a \to c$, $b \to aba$, $c \to b$ (see, e.g., [41, p. 46]) has the property that its cube has three fixed points. None of them is automatic. (Namely, the characteristic polynomial of the (primitive) incidence matrix is equal to $x^3 - x^2 - 2$, which has no rational root.)

Remark 5.4. There exist fixed points of morphisms which are non-automatic for reasons other than Proposition 5.3, that is, such that the dominant eigenvalue of the corresponding incidence matrix is an integer. An example suggested by one of the referees is the fixed point of the morphism $1 \to 2$, $2 \to 211$, (see [1, Section 3.2]). Other examples can be found in the recent

paper [46], which gives, for example, the morphism $a \to ad, b \to adc, c \to b, d \to bc$. Here the dominant eigenvalue is 2, but its fixed point is proved to be not automatic.

Remark 5.5. In the case of a morphic sequence which is not purely morphic, it can happen that the frequencies of letters are rational, while the dominant eigenvalue is irrational, two examples are given by the sequence of moves for the cyclic Hanoi tower [2] and by the sum of digits of the expansions of integers [20] in base $(1 + \sqrt{5})/2$.

We end this section with a theorem which will apply to two morphisms related to other Grigorchuk-like groups (see our Corollary 5.8 below).

Theorem 5.6. *Let* $\mathbf{x} = (x_n)_{n \geqslant 0}$ *be a sequence on some alphabet* $\mathcal{A}$. *Let* $\mathcal{A}'$ *be a proper subset of* $\mathcal{A}$. *Suppose that there exists a sequence* $\mathbf{y} = (y_n)_{n \geqslant 0}$ *on* $\mathcal{A}'$ *with the property that each of its prefixes is a factor of* $\mathbf{x}$. *Let* $d \geqslant 2$. *If no sequence in the closed orbit of* $\mathbf{y}$ *under the shift is* d-*automatic, then* $\mathbf{x}$ *is not* d-*automatic.*

Proof. Define an order on $\mathcal{A}$ such that each element of $\mathcal{A} \setminus \mathcal{A}'$ is larger than each element of $\mathcal{A}'$. The set of sequences on $\mathcal{A}$ is equipped with the lexicographical order induced by the order on $\mathcal{A}$. Let $\mathbf{z} = (z_n)_{n \geqslant 0}$ be the lexicographically least sequence in the orbit closure of $\mathbf{x}$. Since the sequence $\mathbf{y}$ has its values in $\mathcal{A}'$ and since each prefix of $\mathbf{y}$ is a factor of $\mathbf{x}$, the orbit closure of $\mathbf{y}$ under the shift is included in the orbit closure of $\mathbf{x}$. Now, since the elements of $\mathcal{A} \setminus \mathcal{A}'$ are larger than the elements of $\mathcal{A}'$, the least element of the orbit closure of $\mathbf{y}$ is equal to the least element of the orbit closure of $\mathbf{x}$, i.e., is equal to $\mathbf{z}$. Now, if $\mathbf{x}$ were d-automatic for some $d \geqslant 2$, then $\mathbf{z}$ would be d-automatic [6, Theorem 6], which contradicts the hypothesis on the orbit of $\mathbf{y}$. $\qquad\square$

Corollary 5.7. *Let* $\mathbf{x} = (x_n)_{n \geqslant 0}$ *be a sequence on some alphabet* $\mathcal{A}$. *Let* $\mathcal{A}'$ *be a proper subset of* $\mathcal{A}$. *Suppose that there exists a sequence* $\mathbf{y} = (y_n)_{n \geqslant 0}$ *on* $\mathcal{A}'$ *with the property that each of its prefixes is a factor of* $\mathbf{x}$. *Suppose that* $\mathbf{y}$ *is Sturmian, or that* $\mathbf{y}$ *is uniformly recurrent*[3] *and that its complexity is not* $\mathcal{O}(n)$, *then* $\mathbf{x}$ *is not* d-*automatic for any* $d \geqslant 2$.

Proof. If $\mathbf{y}$ is Sturmian, all sequences in its orbit closure are Sturmian (they have complexity $n + 1$), hence cannot be d-automatic. If $\mathbf{y}$ is uniformly recurrent, all the sequences in its orbit closure have the same complexity —which is not $\mathcal{O}(n)$— and thus none of them can be d-automatic. $\qquad\square$

We are ready for our last corollary about the two fixed points of morphisms respectively given in [12, Theorem 2.9] and [8, Theorem 4.5].

Corollary 5.8. *The fixed point beginning with* a *of the morphism* $a \to aca$, $b \to bc$, $c \to b$ *is not automatic. The fixed point beginning with* a *of the morphism* $a \to aca$, $c \to cd$, $d \to c$ *is not automatic.*

Proof. Note that the fixed point of the morphism $b \to bc$, $c \to b$ (respectively $c \to cd$, $d \to c$) is a Sturmian sequence and apply Corollary 5.7 above. $\qquad\square$

Remark 5.9. As suggested by two of the referees, Theorem 5.6 can also be deduced from [16, Theorem A]. Furthermore this approach does not need that $\mathcal{A}'$ be a proper subset of $\mathcal{A}$.

---

[3]Uniformly recurrent sequences are also called *minimal*.

6. Final remarks

For more on the Grigorchuk group or similar groups, the reader can also consult, e.g., [33, 26, 25, 36, 30]. Note that *automata groups* appear to be close to morphic or automatic sequences, while *automatic groups* (see, e.g., [22]) seem to be rather away from these sequences. Note that substitutions can also be used, in a different context, to characterize *families* of groups: for example it is proved in [15] that *a finite group is an extension of a nilpotent group by a 2-group if and only if it satisfies a Thue–Morse identity for all elements x, y*, where the nth Thue–Morse identity between x and y is defined by $\varphi_{x,y}^n(x) = \varphi_{x,y}^n(y)$ for every $n \geqslant 0$, and the Thue–Morse substitution $\varphi_{x,y}$ is defined by $\varphi_{x,y}(x) := xy$ and $\varphi_{x,y}(y) := yx$.

7. Appendix: CUP by F. M. Dekking

The second author of the present paper is in an excellent position to take a fresh look at the CUP algorithm, as he gave 7 years ago a simple version of this construction for the special case of higher order Thue–Morse morphisms (see [19]).

Theorem 7.1. (Theorem 11 from [5]) *Let $(a_n)_{n \geqslant 0}$ be an automatic sequence taking values in the alphabet A. Then $(a_n)_{n \geqslant 0}$ is also non-uniformly morphic. Furthermore, if $(a_n)_{n \geqslant 0}$ is the iterative fixed point of a uniform morphism, then there exist an alphabet A' of cardinality $\#A + 2$ and a sequence $(a'_n)_{n \geqslant 0}$ with values in A', such that $(a'_n)_{n \geqslant 0}$ is the iterative fixed point of some non-uniform morphism with domain A'^* and $(a_n)_{n \geqslant 0}$ is the image of $(a'_n)_{n \geqslant 0}$ under a coding.*

Proof. Since the sequence $(a_n)_{n \geqslant 0}$ is the pointwise image of the iterative fixed point $(x_n)_{n \geqslant 0}$ of some uniform morphism, we may suppose, by replacing $(a_n)_{n \geqslant 0}$ with $(x_n)_{n \geqslant 0}$, that $(a_n)_{n \geqslant 0}$ itself is the iterative fixed point of a uniform morphism γ of length q.

Let $a := a_0, b := a_1, c := a_2$. Since (a_n) is the iterative fixed point of γ starting with $a_0 = a$, we have $\gamma(a) = abcw$, where $w = a_3 \ldots a_{q-1}$ (by replacing γ with γ^2, we may assume that $q \geqslant 3$).

Now introduce two new letters b' and c', and define a morphism γ' on $A' = A \cup \{b', c'\}$ by

$$\gamma'(y) := \gamma(y) \text{ for } y \in A' \setminus \{a, b', c'\}, \quad \gamma'(a) := ab'c'w, \quad \gamma'(b') := z, \quad \gamma'(c') := t,$$

where z and t are (any) two non-empty words of unequal length such that $zt = \gamma(bc)$.

By construction, γ' is not uniform. Its iterative fixed point beginning with a clearly exists, and we denote it by $(a'_n)_{n \geqslant 0}$. This sequence has the property—which we call *the UP-property*—that each b' in it is followed by a c' and each c' is preceded by a b'. We let D denote the coding that sends each letter of A to itself, and sends b' to b and c' to c. For every letter y belonging to $A' \setminus \{a, b', c'\}$ we have $\gamma(y) = \gamma'(y)$. Hence $D \circ \gamma'(y) = D \circ \gamma(y) = \gamma(y) = \gamma \circ D(y)$. For the letter a, and the word $b'c'$ we have

$$D \circ \gamma'(a) = D(ab'c'w) = abcw = \gamma(a) = \gamma \circ D(a),$$

$$D \circ \gamma'(b'c') = D(zt) = zt = \gamma(bc) = \gamma \circ D(b'c').$$

We see from this, that if u is a word from A'^* which has the UP-property, then $D \circ \gamma'(u) = \gamma \circ D(u)$.

Claim: $D([\gamma']^n(a) = \gamma^n(a)$ for all $n \geqslant 1$.

Proof of the claim: this is true for $n = 1$. Suppose true for n. Then

$$D([\gamma']^{n+1}(a)) = D(\gamma'([\gamma']^n(a)) = \gamma(D([\gamma']^n(a)) = \gamma(\gamma^n(a)) = \gamma^{n+1}(a),$$

since $u = [\gamma']^n(a)$ has the UP-property. Finally the Claim of course implies that $D((a'_n)_{n\geqslant0}) = (a_n)_{n\geqslant0}$. □

Acknowledgements

References

[1] G. Allouche, J.-P. Allouche, and J. Shallit, *Kolam indiens, dessins sur le sable aux îles Vanuatu, courbe de Sierpiński et morphismes de monoïde*, Ann. Inst. Fourier **56** (2006), 2115–2130.

[2] J.-P. Allouche, *Note on the cyclic towers of Hanoi*, Theoret. Comput. Sci. **123** (1994), 3–7.

[3] J.-P. Allouche, M. Queffélec, *A note on substitutions and the first Grigorchuk group*, Unpublished draft (2011).

[4] J.-P. Allouche, J. Shallit, *Automatic Sequences. Theory, Applications, Generalizations*, Cambridge University Press, 2003.

[5] J.-P. Allouche, J. Shallit, *Automatic sequences are also non-uniformly morphic*, in Discrete Mathematics and Applications, A. M. Raigorodskii, M. Th. Rassias eds., Springer Optimization and Its Applications, Springer Nature, vol. 165 (2020), pp. 1–6, available at `https://arxiv.org/abs/1910.08546`.

[6] J.-P. Allouche, N. Rampersad, J. Shallit, *Periodicity, repetitions, and orbits of an automatic sequence*, Theoret. Comput. Sci. **410** (2009), 2795–2803.

[7] M. Baake, J. Roberts, R. Yassawi, *Reversing and extended symmetries of shift spaces*, Discrete Contin. Dyn. Syst. **38** (2018), 835–866.

[8] L. Bartholdi, *Endomorphic presentations of branch groups*, J. Algebra **268** (2003), 419–443.

[9] L. Bartholdi, A. Eschler, *Growth of permutational extensions*, Invent. Math. **189** (2012), 431–455.

[10] L. Bartholdi, R. I. Grigorchuk, *On parabolic subgroups and Hecke algebras of some fractal groups*, Serdica Math. J. **28** (2002), 47–90.

[11] L. Bartholdi, O. Siegenthaler, *The twisted twin of the Grigorchuk group*, Internat. J. Algebra Comput. **20** (2010), 465–488.

[12] M. G. Benli, *Profinite completion of Grigorchuk's group is not finitely presented*, Internat. J. Algebra Comput. **22** (2012), 1250045.

[13] J. Berstel, *Sur la construction de mots sans carré*, Sém. Théorie Nombres Bordeaux **8** (1978-1979), 1–16.

[14] V. Berthé (ed.), M. Rigo (ed.), *Combinatorics, Words and Symbolic Dynamics*, Encyclopedia of Mathematics and its Applications, vol. 159, Cambridge University Press, 2016.

[15] M. Boffa, F. Point, *Identités de Thue–Morse dans les groupes*, C. R. Acad. Sci. Paris, Sér. 1 **312** (1991), 667–670.

[16] J. Byszewski, J. Konieczny, E. Krawczyk, *Substitutive systems and a finitary version of Cobham's theorem*, Preprint 2021, available at `https://arxiv.org/abs/1908.11244`, to appear in Combinatorica.

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

[18] F. M. Dekking, *The spectrum of dynamical systems arising from substitutions of constant length*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **41** (1977/78), 221–239.

[19] F. M. Dekking, *On the structure of Thue–Morse subwords, with an application to dynamical systems*, Theoret. Comput. Sci. **550** (2014), 107–112.

[20] M. Dekking, *The sum of digits function of the base phi expansion of the natural numbers*, Integers **20** (2020), Paper No. A45.

[21] F. Durand, *Cobham's theorem for substitutions*, J. Eur. Math. Soc. **13** (2011), 1799–1814.

[22] D. B. A. Epstein, J. W. Cannon, D. F. Holt, Derek, S. V. F. Levy, M. S. Paterson, W. Thurston, *Word Processing in Groups*, Jones and Bartlett Publishers, Boston, MA, 1992.

[23] N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, Lecture Notes in Mathematics, vol. 1794, Springer-Verlag, Berlin, 2002.

[24] D. Gilat, *Gauss's lemma and the irrationality of roots, revisited*, Math. Mag. **85** (2012), 114–116.

[25] R. Grigorchuk, *Solved and unsolved problems around one group* in Infinite Groups: Geometric, Combinatorial, and Dynamical Aspects, Prog. Math., vol. 248, Birkhäuser, Basel, 2005, pp. 117–218.

[26] R. Grigorchuk, A. Zuk, *On a torsion-free weakly branch group defined by a three state automaton*, Internat. J. Algebra Comput. **12** (2002), 1–24.

[27] R. Grigorchuk, D. Lenz, T. Nagnibeda, *Schreier graphs of Grigorchuk's group and a subshift associated to a nonprimitive substitution*, in Groups, Graphs and Random Walks, London Math. Soc. Lecture Note Ser., vol. 436, Cambridge Univ. Press, Cambridge, 2017, pp. 250–299.

[28] R. Grigorchuk, D. Lenz, T. Nagnibeda, *Spectra of Schreier graphs of Grigorchuk's group and Schroedinger operators with aperiodic order*, Math. Ann. **370** (2018), 1607–1637.

[29] R. Grigorchuk, D. Savchuk, Z. Šunić, *The spectral problem, substitutions and iterated monodromy*, in Probability and Mathematical Physics, CRM Proc. Lecture Notes, vol. 42, Amer. Math. Soc., Providence, RI, 2007, pp. 225–248.

[30] R. Grigorchuk, Y. Leonov, V. Nekrashevych, V. Sushchansky, *Self-similar groups, automatic sequences, and unitriangular representations*, Bull. Math. Sci. **6** (2016), 231–285.

[31] F. von Haeseler, *Automatic Sequences*, de Gruyter Expositions in Mathematics, vol. 36, Walter de Gruyter & Co., Berlin, 2003.

[32] G. H. Hardy, E. M. Wright, *An Introduction to the Theory of Numbers*, 4th edition, Oxford University Press, 1975.

[33] P. de la Harpe, *Topics in geometric group theory*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 2000.

[34] R. A. Horn, C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 2013.

[35] S. Istrail, *On irreducible* [sic] *languages and nonrational numbers*, Bull. Math. Soc. Sci. Math. R. S. Roumanie **21 (69)** (1977), 301–308.

[36] G. A. Jones, *Maps related to Grigorchuk's group*, Eur. J. Comb. **32**, (2011), 478–494.

[37] M. Lothaire, *Combinatorics on Words*, 2nd ed., Encyclopedia of Mathematics and its Applications, vol. 17, 1997.

[38] M. Lothaire, *Algebraic Combinatorics on Words*, Cambridge University Press, 2002.

[39] M. Lothaire, *Applied Combinatorics on Words*, Encyclopedia of Mathematics and its Applications, vol. 105, Cambridge University Press, 2005.

[40] I. G. Lysënok, *A system of defining relations for a Grigorchuk group*, Mat. Zametki **38** (1985), 503–516, [Math. Notes **38** (1985), 784–792].

[41] Y. Muntian, *Automata groups*, PhD Thesis, 2009, available at `https://oaktrust.library.tamu.edu/handle/1969.1/ETD-TAMU-2009-05-751`.

[42] V. Nekrashevych, *Palindromic subshifts and simple periodic groups of intermediate growth*, Ann. of Math. **187** (2018), 667–719.

[43] *On-Line Encyclopedia of Integer Sequences*, founded by N. J. A. Sloane, available at `http://oeis.org`.

[44] M. Queffélec, *Substitution dynamical systems—Spectral analysis*, Second edition, Lect. Notes in Math., vol. 1294, Springer-Verlag, Berlin, 2010.

[45] M. Rigo, *Formal Languages, Automata and Numeration Systems, vol.1. Introduction to Combinatorics on Words*, ISTE, John Wiley & Sons, 2014.

[46] L. Spiegelhofer, *Gaps in the Thue–Morse word*, Preprint 2021, available at `https://arxiv.org/abs/2102.01018`.

[47] B. Tan, Z.-Y. Wen, *Invertible substitutions and Sturmian sequences*, European J. Combin. **24** (2003), 983–1002.

[48] Y. Vorobets, *On a substitution subshift related to the Grigorchuk group*, in Differential equations and topology. II, Collected papers. In commemoration of the centenary of the birth of Academician Lev Semenovich Pontryagin, (MAIK Nauka/Interperiodica, Moscow, 2010), Tr. Mat. Inst. Steklova **271** (2010), 319–334, [*Proc. Steklov Inst. Math.* **271**, (2010) 306–321].