Automatic sequences are also non-uniformly morphic

Jean-Paul Allouche  
CNRS, IMJ-PRG  
Sorbonne, 4 Place Jussieu  
F-75252 Paris Cedex 05  
France  
jean-paul.allouche@imj-prg.fr

Jeffrey Shallit  
School of Computer Science  
University of Waterloo  
Waterloo, Ontario N2L 3G1  
Canada  
shallit@uwaterloo.ca

October 22, 2019

Abstract

It is well-known that there exist infinite sequences that are the fixed point of non-uniform morphisms, but not $k$-automatic for any $k$. In this note we show that every $k$-automatic sequence is the image of a fixed point of a non-uniform morphism.

1 Introduction, Definitions, Notation

Combinatorics on words deals with “alphabets”, “words”, “languages”, and “morphisms of monoids”. The first three notions are inspired by the usual meaning of these words in English. Below we recall the precise definitions.

Definition 1. A finite set $A$ is called an alphabet. A word over the alphabet $A$ is a finite (possibly empty) sequence of symbols from $A$. We let $A^*$ denote the set of all words on $A$. A subset of $A^*$ is called a language on $A$. The length of a word $w$, denoted $|w|$, is the number of symbols that it contains (the length of the empty word $\varepsilon$ is 0). The concatenation of two words $w = a_1a_2\cdots a_r$ and $z = b_1b_2\cdots b_s$ of lengths $r$ and $s$, respectively, is the word denoted $wz$ defined by $wz = a_1a_2\cdots a_rb_1b_2\cdots b_s$ of length $r+s$ obtained by gluing $w$ and $z$ in order. The set $A^*$ equipped with concatenation is called the free monoid generated by $A$. The concatenation of a word $w = a_1a_2\cdots a_r$ and a sequence $(x_n)_{n\geq 0}$ is the sequence $a_1a_2\cdots a_rx_0x_1\cdots$, denoted $w(x_n)_{n\geq 0}$. A word $w$ is called a prefix of the word $z$ (or of the infinite sequence $(x_n)_{n\geq 0}$) if there exists a word $y$ with $z = wy$ (respectively a sequence $(y_n)_{n\geq 0}$ with $(x_n)_{n\geq 0} = w(y_n)_{n\geq 0}$).

Let $(u_\ell)_{\ell\geq 0}$ be a sequence of words in of $A^*$, and $(a_n)_{n\geq 0}$ be a sequence over the alphabet $A$. The sequence $(u_\ell)_{\ell\geq 0}$ is said to converge to the sequence $(a_n)_{n\geq 0}$ if the length of the largest prefix of $u_\ell$ that is also a prefix of $(a_n)_{n\geq 0}$ tends to infinity with $\ell$. 

1
Definition 6. An infinite sequence \((\phi_i)_{i \geq 0}\) is said to be pure morphic if there exists a morphism \(\phi\) from \(\mathcal{A}^*\) to \(\mathcal{B}^*\) such that, for all words \(u\) and \(v\), one has \(\phi(uv) = \phi(u)\phi(v)\). A morphism of \(\mathcal{A}^*\) is a morphism from \(\mathcal{A}^*\) to itself.

Remark 5. A morphism \(\phi\) from \(\mathcal{A}^*\) to \(\mathcal{B}^*\) is completely determined by the values of \(\phi(a)\) for \(a \in \mathcal{A}\). Namely, if the word \(u\) is equal to \(a_1a_2\cdots a_n\) with \(a_j \in \mathcal{A}\), then \(\phi(u) = \phi(a_1)\phi(a_2)\cdots\phi(a_n)\).

Definition 7. Let \(\mathcal{A}\) and \(\mathcal{B}\) be two alphabets. A morphism from \(\mathcal{A}^*\) to \(\mathcal{B}^*\) is a map \(\phi\) from \(\mathcal{A}^*\) to \(\mathcal{B}^*\) such that, for all words \(u\) and \(v\), one has \(\phi(uv) = \phi(u)\phi(v)\). A morphism of \(\mathcal{A}^*\) is a morphism from \(\mathcal{A}^*\) to itself.

Remark 2. It is straightforward that \(\mathcal{A}^*\) equipped with concatenation is indeed a monoid: concatenation is associative, and the empty word \(\epsilon\) is the identity element. This monoid is free; intuitively, this means that there are no relations between elements, other than the relations arising from the associative property and the fact that the empty word is the identity element. In particular, this monoid is not commutative if \(\mathcal{A}\) has at least two distinct elements.

Example 4. The Thue-Morse morphism \(\mu\) sending \(0 \rightarrow 01\) and \(1 \rightarrow 10\) is 2-uniform. In contrast, the Fibonacci morphism \(\tau\) sending \(a\) to \(ab\) and \(b\) to \(a\) is non-uniform.

Remark 7. It is immediate that
\[
\varphi(a_0) = a_0x
\]
\[
\varphi^2(a_0) = \varphi(\varphi(a_0)) = \varphi(a_0)x = \varphi(a_0)\varphi(x) = a_0x\varphi(x)
\]
\[
\varphi^3(a_0) = \varphi(\varphi^2(a_0)) = \varphi(a_0x\varphi(x)) = \varphi(a_0)\varphi(x)\varphi^2(x) = a_0x\varphi(x)\varphi^2(x)
\]
and more generally
\[
\varphi^\ell(a_0) = a_0x\varphi(x)\varphi^2(x)\cdots\varphi^{\ell-1}(x)
\]
for all \(\ell \geq 0\).

Definition 8. An infinite sequence \((a_n)_{n \geq 0}\) taking values in \(\mathcal{A}\) is said to be morphic if there exist an alphabet \(\mathcal{B}\) and an infinite sequence \((b_n)_{n \geq 0}\) over the alphabet \(\mathcal{B}\) such that

- the sequence \((b_n)_{n \geq 0}\) is pure morphic;
• there exists a coding from $B^*$ to $A^*$ sending the sequence $(b_n)_{n \geq 0}$ to the sequence $(a_n)_{n \geq 0}$; i.e., the sequence $(a_n)_{n \geq 0}$ is the pointwise image of $(b_n)_{n \geq 0}$.

If the morphism making $(b_n)_{n \geq 0}$ morphic is $k$-uniform, then the sequence $(a_n)_{n \geq 0}$ is said to be $k$-automatic. The word “automatic” comes from the fact that the sequence $(a_n)_{n \geq 0}$ can be generated by a finite automaton (see [2] for more details on this topic).

**Remark 9.** A morphism $\varphi$ of $A^*$ can be extended to infinite sequences with values in $A$ by defining

$$\varphi((a_n)_{n \geq 0}) = \varphi(a_0a_1a_2 \cdots) := \varphi(a_0)\varphi(a_1)\varphi(a_2) \cdots.$$ 

It is easy to see that a pure morphic sequence is a fixed point of (the extension to infinite sequences of) some morphism: actually, with the notation above, it is the fixed point of $\varphi$ beginning with $a_0$. A pure morphic sequence is also called an iterative fixed point of some morphism (because of the construction of that fixed point), while a morphic sequence is the pointwise image of an iterative fixed point of some morphism, and a $k$-automatic sequence is the pointwise image of the iterative fixed point of a $k$-uniform morphism.

## 2 The main result

Looking at the definitions above, we see that every automatic sequence is also a morphic sequence. We will prove that every automatic sequence can be obtained as a morphic sequence where the involved morphism is not uniform.

**Definition 10.** We say a sequence is **non-uniformly pure morphic** if it is the iterative fixed point of a non-uniform morphism. We say that a sequence is **non-uniformly morphic** if it is the image (under a coding) of a non-uniformly pure morphic sequence.

For example, the sequence $abaababa \cdots$ generated by iterating the morphism $\tau$ defined above is non-uniformly pure morphic. This sequence is known as the **(binary) Fibonacci sequence**, since it is also equal to the limit of the sequence of words $(u_n)_{n \geq 0}$ defined by

$$u_0 := a, \quad u_1 := ab, \quad u_{n+2} := u_{n+1}u_n$$

for each $n \geq 0$.

In order to avoid triviality, we certainly assume (as M. Mendes France once pointed out to us) that the alphabet of the non-uniform morphism involved in the above definition is the same as the minimal alphabet of its fixed point. For example, the fact that the morphism $0 \rightarrow 01, \ 1 \rightarrow 10, \ 2 \rightarrow 1101$, whose iterative fixed point beginning with 0 is also the iterative fixed point, beginning with 0, of the morphism $\mu$ — namely the Thue-Morse sequence — does not make that sequence non-uniformly morphic.)

Although most non-uniformly morphic sequences are not automatic (e.g., the binary Fibonacci sequence is not automatic), some sequences can be simultaneously automatic and non-uniformly morphic. An example is the sequence $Z$ formed by the lengths of the blocks of 1’s between two consecutive zeros in the Thue-Morse sequence.

$$Z = 2 \ 1 \ 0 \ 2 \ 0 \ 1 \ 2 \cdots$$
As is well known \cite{[3]}, this sequence is both the fixed point of the map sending 2 \to 210, 1 \to 20, and 0 \to 1, and also the image, under the coding 0 \to 2, 1 \to 1, 2 \to 0, 3 \to 1 of the fixed point of the map 0 \to 01, 1 \to 20, 2 \to 23, and 3 \to 02.

In view of this example, one can ask which non-uniformly morphic sequences are also \(k\)-automatic for some integer \(k \geq 2\), or which automatic sequences are also non-uniformly morphic. We prove here that all automatic sequences are also non-uniformly morphic.

**Theorem 11.** Let \((a_n)_{n \geq 0}\) be an automatic sequence taking values in the alphabet \(A\). Then \((a_n)_{n \geq 0}\) is also non-uniformly morphic. Furthermore, if \((a_n)_{n \geq 0}\) is the iterative fixed point of a uniform morphism, then there exist an alphabet \(B\) of cardinality \((3 + \#A)\) and a sequence \((a'_n)_{n \geq 0}\) with values in \(B\), such that \((a'_n)_{n \geq 0}\) is the iterative fixed point of some non-uniform morphism with domain \(B^*\) and \((a_n)_{n \geq 0}\) is the image of \((a'_n)_{n \geq 0}\) under a coding.

**Proof.** We start with the first assertion. First, we may suppose that the first letter of \((a_n)_{n \geq 0}\) is different from all \(a_j\) for \(j \geq 1\). If not, take a letter \(a\) not in \(A\) and consider the sequence \(\alpha a j a \cdots\). This sequence is automatic and the morphism \(\alpha \to a_0\) and \(a \to a\) for all letters \(a\) in \(A\) sends it to \((a_n)_{n \geq 0}\).

We may also suppose that the sequence \((a_n)_{n \geq 0}\) is not ultimately periodic (otherwise the result is trivial: if \(u\) and \(v\) are two words over the alphabet \(A\), the sequence \(uvwv \cdots\) is the iterative fixed point of the morphism \(\alpha \to u\) and \(a \to v\) for all \(a \in A\), where \(j\) is chosen so that \(|j|v| \neq |u|\)).

Thus we may now start with an automatic non-ultimately periodic sequence, still called \((a_n)_{n \geq 0}\), with \(a_0 = \alpha \neq a_1\). Since the sequence \((a_n)_{n \geq 0}\) is the pointwise image of the iterative fixed point \((x_n)_{n \geq 0}\) of some uniform morphism, we may suppose, by replacing \((a_n)_{n \geq 0}\) with \((x_n)_{n \geq 0}\), that \((a_n)_{n \geq 0}\) itself is the iterative fixed point beginning with \(a_0 = \alpha \neq a_j\) for all \(j \geq 1\) of a uniform morphism \(\gamma\) with domain \(A^*\), and still non-ultimately periodic.

We claim that there exists a 2-letter word \(bc\) such that \(\gamma(bc)\) contains \(bc\) as a factor. Namely, since \(\gamma\) is uniform, it has exponential growth (that is, iterating \(\gamma\) on each letter gives words of exponentially growing length). Hence there exists a letter \(b\) that is expanding, i.e., such that some power of \(\gamma\) maps \(b\) to a word that contains at least two occurrences of \(b\) (see, e.g., \cite{[4]}). By replacing \(\gamma\) with this power of \(\gamma\), we can write \(\gamma(b) = uvbwv\) for some words \(u, v, w\). By replacing this new \(\gamma\) with \(\gamma^2\), we can also suppose that both \(u\) and \(w\) are nonempty. Let \(c\) be the letter following the prefix \(ub\) of \(uvbwv\). Now there are two cases:

- if \(c \neq b\), then \(v = cy\) for some word \(y\), and \(\gamma(b) = ubcyw\), and \(\gamma(bc) = \gamma(b)\gamma(c)\) contains \(bc\) as a factor;
- if \(c = b\), then \(\gamma(b) = ubbz\) for some word \(z\), and \(\gamma(bb) = ubbzubbz\) contains \(bb\) as a factor.

In both cases, there exist two letters \(b\) and \(c\), not necessarily distinct, such that \(\gamma(b) = w_1bcw_2\) and \(\gamma(bc) = w_1bcw_3\), where \(w_1, w_2\) are non-empty words. Note, in particular, that \(b\) can be chosen distinct from \(a_0\) (\(w_1\) is non-empty and \(a_0 = \alpha\) is different from all \(a_j\) for \(j \geq 1\)).

Now define a new alphabet \(A' := A \cup \{b', c'\}\), where \(b', c'\) are two new letters not in \(A\). Define the morphism \(\gamma'\) with domain \(A'\) as follows: if the letter \(y\) belongs to \(A \setminus \{b\}\), then
\[ \gamma'(y) := \gamma(y). \] If \( y = b \), define \( \gamma'(b) := w_1b'c'w_2 \). Finally, define \( \gamma'(b') \) and \( \gamma'(c') \) as follows: first recall that \( \gamma(bc) = w_1bcw_3 \); cut the word \( w_1bcw_3 \) into (any) two non-empty words of unequal length, say \( w_1b'c'w_3 \); define \( \gamma'(b') := z \), \( \gamma'(c') := t \).

By construction, \( \gamma' \) is not uniform. Its iterative fixed point beginning with \( a_0 \) clearly exists, and we denote it by \( (a'_n)_{n \geq 0} \). This sequence has the 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 \( \mathcal{A} \) to itself, and sends \( b' \) to \( b \) and \( c' \) to \( c \). For every letter \( x \) belonging to \( \mathcal{A}' \setminus \{b, b', c'\} \) we have \( \gamma(x) = \gamma'(x) \). Hence \( D \circ \gamma'(x) = D \circ \gamma(x) = \gamma(x) \). For \( x = b \), we have \( D \circ \gamma'(b) = D(w_1b'c'w_2) = w_1bcw_3 = \gamma(b) \). Furthermore, we have \( D \circ \gamma'(b') = D(zt) = zt = w_1bcw_3 = \gamma(bc) \).

Now let \( P_k \) be the prefix of the sequence \( (a'_n)_{n \geq 0} \) that ends with \( c' \) and contains exactly \( k \) occurrences of the letter \( c' \). Each occurrence of \( c' \) must be preceded by a \( b' \), so that \( P_k \) can be written \( P_k = p_1b'c'p_2b'c' \cdots p_kb'c' \) where the \( p_i \)'s are words over the alphabet \( \mathcal{A} \). We have

\[
D \circ \gamma'(P_k) = D \circ \gamma'(p_1b'c'p_2b'c' \cdots p_kb'c') \\
= (D \circ \gamma'(p_1))(D \circ \gamma'(b')c')(D \circ \gamma'(p_2))(D \circ \gamma'(b'c')) \cdots (D \circ \gamma'(p_k))(D \circ \gamma'(b'c')) \\
= \gamma(p_1)\gamma(bc)\gamma(p_2)\gamma(bc) \cdots \gamma(p_k)\gamma(bc) \\
= \gamma(p_1bcp_2bc \cdots p_kbc) \\
= \gamma \circ D(p_1b'c'p_2b'c' \cdots p_kb'c') \\
= \gamma \circ D(P_k).
\]

Letting \( k \) go to infinity, we obtain that \( D \circ \gamma'((a'_n)_{n \geq 0}) = \gamma \circ D((a'_n)_{n \geq 0}) \), but \( \gamma'((a'_n)_{n \geq 0}) = (a'_n)_{n \geq 0} \), so that \( D((a'_n)_{n \geq 0}) = \gamma \circ D((a'_n)_{n \geq 0}) \). Hence \( D((a'_n)_{n \geq 0}) \) is the iterative fixed point of \( \gamma \) beginning with \( a_0 \). Hence it is equal to the sequence \( (a_n)_{n \geq 0} \).

The second assertion is a consequence of the fact that we introduced at most only three new letters \( \alpha, b', c' \) in the proof above.

\[ \square \]

**References**

[1] J.-P. Allouche and J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence, in *Sequences and Their Applications*, Proceedings of SETA’98, C. Ding, T. Helleseth and H. Niederreiter (Eds.), (Springer, 1999), pp. 1–16.

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

[3] J. Berstel, Sur la construction de mots sans carré, Sém. Théor. Nombres Bordeaux, 1978–1979, Exposé 18, 18-01–18-15.

[4] A. Salomaa, On exponential growth in Lindenmayer systems, Nederl. Akad. Wetensch. Proc. Ser. A 76 (= Indag. Math. 35 (1973)), 23–30.