Journal of Integer Sequences, Vol. 19 (2016),
Article 16.1.1

Morphisms, Symbolic Sequences, and Their Standard Forms

F. Michel Dekking
DIAM
Delft University of Technology
Mekelweg 4
2628 CD Delft
The Netherlands
f.m.dekking@tudelft.nl

Abstract

Morphisms are homomorphisms under the concatenation operation of the set of words over a finite alphabet. Changing the elements of the finite alphabet does not change the morphism in an essential way. We propose a method to select a unique representative from all these morphisms. This has applications to the classification of the shift dynamical systems generated by morphisms. In a similar way, we propose the selection of a representing sequence out of the class of symbolic sequences over an alphabet of fixed cardinality. Both methods are useful for the storing of symbolic sequences in databases, such as *The On-Line Encyclopedia of Integer Sequences*. We illustrate our proposals with the k-symbol Fibonacci sequences.

1 Introduction

Here is a sequence on an alphabet of six letters:

$$x = 1, 4, 2, 1, 6, 3, 5, 4, 2, 3, 5, 6, 1, 4, 2, 1, \ldots.$$

Why should one be interested in this sequence? Well, if one projects with the letter-to-letter map $1, 3, 5 \to 0,\ 2, 4, 6 \to 1$, one obtains the Thue-Morse sequence

$$0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, \ldots.$$

This is the most famous pure morphic sequence (A010060, A001285, A010059). The sequence x can be obtained from a 3-block morphism (see Section 4) starting from the Thue-Morse morphism

$$0 \to 01, \quad 1 \to 10.$$

This gives that x itself is pure morphic under the square of the morphism

$$\zeta: \qquad 1 \to 23, \quad 2 \to 14, \quad 3 \to 21, \quad 4 \to 56, \quad 5 \to 63, \quad 6 \to 54.$$

It is shown by Coven, Dekking, and Keane [5] that the shift dynamical system generated by ζ is topologically isomorphic to the Thue-Morse system, and that there essentially are only 12 injective morphisms (with associated infinite sequences) that have this property.
How should x be added to the On-Line Encyclopedia of Integer Sequences?

There is a problem here. The choice of the alphabet is quite arbitrary, another popular choice is $\{0, 1, 2, 3, 4, 5\}$. But even with a fixed alphabet of 6 letters, there are 6!=720 morphisms, generating as many sequences, which are essentially the same.

The ternary Thue-Morse sequence generated by $a \to abc$, $b \to ac$, $c \to b$ occurs 12 times in OEIS, exhausting all 3! possibilities of the two alphabets $\{0, 1, 2\}$ and $\{1, 2, 3\}$. It is inconceivable that this could be done for a pure morphic sequence on an alphabet of, say, 20 letters. So what one needs is a canonical way to extract *one* morphism out of these huge collections. This will be done in the next section. In Section 3 we propose a simple way to standardize symbolic sequences. In Section 4 and 5 we illustrate our approach with a discussion of the k-symbol Fibonacci morphisms.

2 Standard forms of morphisms

Let θ be a morphism on r letters. The *standard alphabet* is the set $A = \{1, 2, \dots, r\}$.

Let Π be a permutation on A. The *permuted* version of θ is the morphism θ' given by

$$\theta'(a) = \Pi^{-1}\theta(\Pi(a)), \quad \text{for } a \in A.$$

A morphism θ is called *uniform* if $|\theta(a)| = |\theta(b)|$ for all $a, b \in A$, where $|w|$ denotes the length of a word w.

Definition 1. The *standard form* of a uniform morphism θ is obtained by permuting the letters of the standard alphabet in such a way that the permuted version has the smallest word $\theta'(1)\theta'(2)\cdots\theta'(r)$ in the lexicographical order on words.

As an example, let θ be the Toeplitz or period doubling morphism, see

A096268: Fixed point of the morphism $0 \to 01, 1 \to 00$,

A056832: Fixed point of the morphism $1 \to 12, 2 \to 11$,

A035263: Trajectory of 1 under the morphism $1 \to 10, 0 \to 11$.

The second OEIS entry is on the standard alphabet. The permuted version, with $\Pi = 1 \rightleftarrows 2$ is $1 \to 22, 2 \to 21$. Since $1211 < 2212$, the standard form of the morphism is $1 \to 12, 2 \to 11$.

Remark 2. A caveat: the standard form of a morphism has some undesirable properties. For instance, the square of the standard form need not be equal to the standard form of the square. □

Note that the standard form of a morphisms is uniquely defined, since

$$\theta'(1)\theta'(2)\cdots\theta'(r) = \theta(1)\theta(2)\cdots\theta(r) \implies \theta' = \theta.$$

This need not hold in the non-uniform case. For example, if θ is given by

$$\theta(1) = 12,\ \theta(2) = 312,\ \theta(3) = 3123,$$

then $\theta(123) = 123123123$, and the same word appears under the version of θ permuted by $\Pi = 1 \to 2 \to 3$.

Definition 3. The *standard form* of a morphism θ is obtained in the following way. First, permute the letters of the standard alphabet in such a way that the permuted version θ' has the smallest word $w_\chi := \theta'(1)\theta'(2)\cdots\theta'(r)$ in the lexicographical order on words. Among all permutations that have the same w_χ, choose the one with the lexicographically smallest vector $(|\theta'(1)|, |\theta'(2)|, \ldots, |\theta'(r)|)$.

Note that once more, the standard form is uniquely defined.

The standard form of ternary Thue-Morse is the morphism $1 \to 123,\ 2 \to 13,\ 3 \to 2$.

3 Standard forms of symbolic sequences

We now focus on infinite sequences generated by morphisms, and in particular fixed points of morphisms. These are typical examples of what we call symbolic sequences, where a change of symbols does not change the essential properties of the sequence. An example of a non-symbolic sequence on four symbols is $(s_n) =$ A002828, which gives the least number of squares that add up to n.

Without loss of generality we assume that all symbols of $A = \{1, \ldots, r\}$ occur in a sequence over A. We then have the following very simple way to pick a representing sequence.

Definition 4. The *standard form* of a sequence x over A is $\Pi(x)$, where Π is the permutation of A which yields the smallest sequence in lexicographical order.

For example, the standard form of the the Toeplitz or period doubling sequence is $1, 2, 1, 1, 1, 2, 1, 2, 1, 2, 1, 1, 1, 2, 1, 1, \ldots$, and the standard form of the ternary Thue-Morse sequence is $1, 2, 3, 1, 3, 2, 1, 2, 3, 2, 1, 3, 1, \ldots$. It is easily seen that both sequences are the unique fixed point of the corresponding morphism, given in its standard form. This will not happen in general. An exception, for example, is the 6-symbol Thue-Morse morphism of Section 1. For archiving purposes we recommend to store both the fixed points of the standard form of the morphism and the standard form of these fixed points.

4 k-symbol Fibonacci morphisms and sequences

For any N the N-block morphism $\hat{\theta}_N$ of a morphism θ yields pure morphic sequences on an alphabet of $p(N)$ symbols, where $p(\cdot)$ is the subword complexity function of a sequence generated by θ (cf. [7, p. 95]).

In this section we take a look at another famous morphism, the Fibonacci morphism

$$\varphi: \qquad 0 \to 01, \quad 1 \to 0.$$

It is well known (see, e.g.,[7, p. 105]) that $p(N) = N+1$, so $\hat{\varphi}_N$ is a morphism on an alphabet of $N+1$ symbols.

We describe how to obtain for $N = 2$ the 3-symbol Fibonacci morphism $\hat{\varphi}_2$. The blocks of length 2 occurring in the language of φ are 00, 01, and 10. We have

$$\varphi(00) = 0101 =: u_1u_2u_3u_4, \quad \varphi(01) = 010 =: v_1v_2v_3, \quad \varphi(10) = 001 =: w_1w_2w_3.$$

The rule is now that the length of $\hat{\varphi}_2(j_1j_2)$ is equal to the length of $\varphi(j_1)$. So since u_1u_2, $u_2u_3 = 01$, 10, $v_1v_2, v_2v_3 = 01, 10$ and $w_1w_2 = 00$, one obtains on the 2-blocks

$$00 \to 01,\ 10, \quad 01 \to 01,\ 10, \quad 10 \to 00.$$

Coding these to the standard alphabet $\{1, 2, 3\}$ by $00 \leftrightarrow 3, 01 \leftrightarrow 1$ and $10 \leftrightarrow 2$, we obtain the standard form

$$\hat{\varphi}_2(1) = 12 \quad \hat{\varphi}_2(2) = 3, \quad \hat{\varphi}_2(3) = 12.$$

The unique fixed point of $\hat{\varphi}_2$ is

$$x^{(2)} = 1, 2, 3, 1, 2, 1, 2, 3, 1, 2, 3, 1, 2, 1, 2, 3, 1, 2, 1, 2, 3, \ldots.$$

This sequence will not be found[1] in the OEIS. However, the sequence obtained by mapping $1 \to 2$, $2 \to 0$, $3 \to 1$ is sequence A159917, and is described as the "Fixed point of the morphism $0 \to 01$, $1 \to 2$, $2 \to 01$". Except for a relationship with the so-called Fibbinary numbers (A003714), no other properties are given. We will make a connection with the sequence $(a_n) = 0, 1, 0, 2, 1, 0, 2, 0, 2, 1, 0, 2, 1, 0, \ldots$ given in A120614. Its definition is

$$a_n = g_n - g_{n-1} \text{ where } g_k = \lfloor \Phi \lfloor k/\Phi \rfloor \rfloor \text{ and } \Phi = (1 + \sqrt{5})/2. \tag{1}$$

From its comments one can deduce that the sequence $1, 0, 2, 1, 0, 2, 0, 2, 1, 0, 2, 1, 0, \ldots$ obtained after deleting the initial entry 0 is a fixed point of the morphism

$$0 \to 102, \qquad 1 \to 102, \qquad 2 \to 02.$$

Coding this morphism to the standard alphabet via $0 \to 1, 1 \to 3, 2 \to 2$ we obtain the morphism θ given by

$$\theta(1) = 312, \quad \theta(2) = 12, \quad \theta(3) = 312.$$

[1] As of November 2015.

We will compare θ with $\hat{\varphi}_2^2$. Note that

$$\hat{\varphi}_2^2(1) = 123, \quad \hat{\varphi}_2^2(2) = 12, \quad \hat{\varphi}_2^2(3) = 123.$$

We say a morphism α on A can be rotated if $(\alpha(a))_1 = b$ for all $a \in A$, and some fixed $b \in A$. The *rotated morphism* is the morphism denoted $\rho\alpha$, defined by

$$\rho\alpha(a) = b^{-1}\,\alpha(a)\,b \quad \text{for all } a \in A.$$

In the literature such a rotated morphism is also known as a conjugated morphism. Note that $\rho^2\hat{\varphi}_2^2 = \theta$. This implies that $\hat{\varphi}_2^2$ and θ generate the same language (this is not hard to prove, but see also [1, Lemma 3]). Thus arbitrarily long subwords of $x^{(2)}$ will occur in the sequence (a_n) from Equation (1), standardized, and conversely. Actually, a little bit more is true. In the following we write shortly $\eta := \hat{\varphi}_2^2$.

Proposition 5. *Let η be defined by $\eta(1) = 123,\ \eta(2) = 12,\ \eta(3) = 123$, and let θ be defined by $\theta(1) = 312,\ \theta(2) = 12,\ \theta(3) = 312$. Then the fixed point $12312123123\cdots$ of η equals the left shift of the fixed point $31231212312\cdots$ of θ.*

Proof. Note first that for all n

$$\eta^n(2) = \eta^n(1)[\eta^{n-1}(3)]^{-1}, \quad \theta^n(2) = [\theta^{n-1}(3)]^{-1}\theta^n(1),$$

simply because $\eta(2) = \eta(1)3^{-1}$ and $\theta(2) = 3^{-1}\theta(1)$. The proposition will be proved if we show that for all n

$$3\eta^n(1) = \theta^n(1)3.$$

This will be done by induction over two levels. Obviously $3\eta(1) = 3123 = \theta(1)3$, and $3\eta^2(1) = 312312123 = \theta^2(1)3$. We have for $n \geq 2$

$$\begin{aligned}
3\eta^{n+1}(1) &= 3\eta^n(1)\eta^n(2)\eta^n(3) = \theta^n(1)3\,\eta^n(1)[\eta^{n-1}(3)]^{-1}\,3^{-1}\theta^n(1)3 \\
&= \theta^n(1)\,\theta^n(1)3\,[3^{-1}\theta^{n-1}(1)3]^{-1}\,3^{-1}\theta^n(1)3 \\
&= \theta^n(3)\,\theta^n(1)\,[\theta^{n-1}(3)]^{-1}\,\theta^n(1)3 \\
&= \theta^n(3)\,\theta^n(1)\,\theta^n(2)3 = \theta^{n+1}(1)3.
\end{aligned}$$

$\square$

We end this section with 4-symbol Fibonacci. The words of length 3 in the Fibonacci language are 001, 010, 100, and 101. Coding these to the standard alphabet $\{1, 2, 3, 4\}$ by $001 \leftrightarrow 3, 010 \leftrightarrow 1, 100 \leftrightarrow 2$ and $101 \leftrightarrow 4$, we obtain the standard form morphism

$$\hat{\varphi}_3(1) = 12, \quad \hat{\varphi}_3(2) = 3, \quad \hat{\varphi}_3(3) = 14, \quad \hat{\varphi}_3(4) = 3.$$

It has just one fixed point, which is

$$x^{(3)} = 1, 2, 3, 1, 4, 1, 2, 3, 1, 2, 3, 1, 4, 1, 2, 3, \ldots.$$

This happens to be A138967, which indeed deserves to be called the Infinite Fibonacci word on the alphabet $\{1,2,3,4\}$. To prove this, note that $\hat{\varphi}_3^2$ is the morphism

$$1 \to 241, \quad 2 \to 241, \quad 3 \to 24, \quad 4 \to 24.$$

If we make this morphism injective by merging symbols with equals images (cf. [2]), we obtain the morphism $\bar{1} \to \bar{1}\bar{3}\bar{1}, \bar{3} \to \bar{1}\bar{3}$, which is the square of the Fibonacci morphism. Hence, indeed, the recipe in the Comments of A138967: "replace $\bar{1}$ by 1,2,3 and $\bar{3}$ by 1,4 in the infinite 2-symbol Fibonacci word", generates $x^{(3)}$.

5 A tale of two sequences

In the previous section we showed that the sequence $(a_{n+1}) = 1,0,2,1,0,2,0,2,1,0,2,1,0,\dots$ given in A120614 essentially equals the 3-symbol Fibonacci sequence. Its definition is

$$a_n = g_n - g_{n-1} \text{ where } g_k = \lfloor \Phi \lfloor k/\Phi \rfloor \rfloor \text{ and } \Phi = (1+\sqrt{5})/2.$$

Here (g_n) is A120613 given by

$$(g_{n+1}) = 1,1,3,4,4,6,6,8,9,9,11,12,12,14,14,16,17,17,19,19,21,22,22,\dots.$$

This sequence was contributed to OEIS by Cloitre in 2006, but the only reference in its description is to a sequence in the paper [6] by Griffiths, a paper which occurred in 2011. On page 502 in [6] Griffiths states "Incidentally, it is interesting to note that the sequence

$$\left\lfloor \frac{\lfloor (n+1)\Phi \rfloor - 1}{\Phi} \right\rfloor$$

is also related to the golden string. The first few terms are given by $1,1,3,4,4,6,6,8,9,9,11,$ $12,12,14,14,16,17,17,19,19,21,22,22,\dots$."
Griffiths does not give any references to other work.

Problem 6. How to show that $\lfloor \Phi \lfloor \frac{n}{\Phi} \rfloor \rfloor = \left\lfloor \frac{\lfloor n\Phi \rfloor - 1}{\Phi} \right\rfloor$ for all $n \geq 1$?

Using $1/\Phi = \Phi - 1$, the relation is equivalent to

$$\lfloor \Phi \lfloor n\Phi - n \rfloor \rfloor = \lfloor \Phi \lfloor n\Phi \rfloor - \lfloor n\Phi \rfloor - \Phi + 1 \rfloor.$$

Taking $-n$ out of the floor, and bringing $\lfloor n\Phi \rfloor$ to the other side this is equivalent to

$$\lfloor \Phi \lfloor n\Phi \rfloor - n\Phi + \lfloor n\Phi \rfloor \rfloor = \lfloor \Phi \lfloor n\Phi \rfloor - \Phi + 1 \rfloor \quad \text{for all } n \geq 1. \tag{2}$$

This clearly exposes the close relationship between the two sequences: putting $\{n\Phi\} = n\Phi - \lfloor n\Phi \rfloor$, we of course have $-1 \leq -\{n\Phi\} \leq 0$ and $-\Phi + 1 = -0.618\dots$. We thus see that Equation (2) is equivalent to

$$\{\Phi \lfloor n\Phi \rfloor\} < \Phi - 1 \quad \Longleftrightarrow \quad \{\Phi \lfloor n\Phi \rfloor\} < \{n\Phi\} \quad \text{for all } n \geq 1. \tag{3}$$

Let a be the basic sequence A000201 in the papers by Carlitz et al. in [3], and [4]:

$$a(n) = \lfloor n\Phi \rfloor.$$

Then Equation (2) is equivalent to

$$a(a(n) - n) = a(a(n) - 1) - a(n) - 1. \tag{4}$$

But this relation will not be found in the papers by Carlitz et al. The reason is that they consequently avoid the shift operator. For instance, if $e(n) := \lfloor (n+1)/\Phi \rfloor$ is the fifth basic sequence defined by Carlitz et al., then $e(n) = a(n+1) - (n+1)$, but this equation does not occur. We remark that $(e(n))$ is A060143 shifted, and A005206 shifted (Hofstadter G-sequence). It is amusing that Equation (4) can be rewritten as $a(n) = a(a(n) - 1) - a(a(n) - n) - 1$, which is a Hofstadter-Conway type of recursion (cf. A004001).

After this digression we will now prove the equivalence in Equation (3).

Proposition 7. *For all positive integers n:*

$$\{\lfloor n\Phi \rfloor \Phi\} < \Phi - 1 \quad \Longleftrightarrow \quad \{\lfloor n\Phi \rfloor \Phi\} < \{n\Phi\}.$$

We prove this proposition with the following lemma.

Lemma 8. *Let $\Phi = \frac{1}{2}(1 + \sqrt{5})$ be the golden mean. Then for all positive integers n*

$$\{\lfloor n\Phi \rfloor \Phi\} = (1 - \Phi)\{n\Phi\} + 1.$$

Proof. All that we use is the relation $\Phi^2 = \Phi + 1$.

$$\begin{aligned}
\{\lfloor n\Phi \rfloor \Phi\} &= \lfloor n\Phi \rfloor \Phi - \lfloor \lfloor n\Phi \rfloor \Phi \rfloor \\
&= n\Phi^2 - \Phi\{n\Phi\} - \lfloor \lfloor n\Phi \rfloor \Phi \rfloor \\
&= n\Phi + n - \Phi\{n\Phi\} - \lfloor \lfloor n\Phi \rfloor \Phi \rfloor \\
&= \lfloor n\Phi \rfloor + \{n\Phi\} + n - \Phi\{n\Phi\} - \lfloor \lfloor n\Phi \rfloor \Phi \rfloor \\
&= (1 - \Phi)\{n\Phi\} + n + \lfloor n\Phi \rfloor - \lfloor \lfloor n\Phi \rfloor \Phi \rfloor \\
&= (1 - \Phi)\{n\Phi\} + 1.
\end{aligned}$$

Here the last step holds because the integer $n + \lfloor n\Phi \rfloor - \lfloor \lfloor n\Phi \rfloor \Phi \rfloor$ has to be equal to 1, since $0 \leq \{\lfloor n\Phi \rfloor \Phi\} \leq 1$, and $-0.73 < 1 - \Phi \leq (1 - \Phi)\{n\Phi\} \leq 0$. ☐

Proposition 7 follows directly from Lemma 8 by noting that the line $y = (1 - \Phi)x + 1$ has a negative slope, and cuts the line $y = x$ at $x = \Phi - 1$.

The following well known relationship (see, e.g., [4, Equation (1.11)]) is a corollary to the proof of Lemma 8.

Corollary 9. *For all positive integers n: $\lfloor \lfloor n\Phi \rfloor \Phi \rfloor = \lfloor n\Phi \rfloor + n - 1$.*

References

[1] Jean-Paul Allouche, Michael Baake, Julien Cassaigne, and David Damanik, Palindrome complexity, *Theoret. Comput. Sci.* **292** (2003), 9–31.

[2] François Blanchard, Fabien Durand, and Alejandro Maass, Constant-length substitutions and countable scrambled sets, *Nonlinearity* **17** (2004), 817–833.

[3] L. Carlitz, V. E. Hoggatt Jr., and Richard Scoville, Fibonacci representations, *Fibonacci Quart.* **10** (1972), 1–28.

[4] L. Carlitz, R. Scoville, and T. Vaughan, Some arithmetic functions related to Fibonacci numbers, *Fibonacci Quart.* **11** (1973), 337–386.

[5] Ethan M. Coven, F. Michel Dekking, and Michael S. Keane, Topological conjugacy of constant length substitution dynamical systems, preprint, 2015, http://arxiv.org/abs/1401.0126.

[6] Martin Griffiths, The golden string, Zeckendorf representations, and the sum of a series, *Amer. Math. Monthly* **118** (2011), 497–507.

[7] Martine Queffélec, *Substitution Dynamical Systems – Spectral Analysis*, Lecture Notes in Mathematics, Vol. 1294, 2nd edition, Springer, 2010.

2010 *Mathematics Subject Classification*: Primary 68R15; Secondary 37B10, 11B85.
Keywords: morphism, substitution, pure morphic, Fibonacci word, standard form.

(Concerned with sequences A000078, A000201, A001285, A002828, A003714, A004001, A005206, A007413, A010059, A010060, A035263, A056832, A060143, A080843, A096268, A120613, A120614, A138967, A159917, A216190, A254990, and A255014.)

Received August 31 2015; revised version received December 7 2015. Published in *Journal of Integer Sequences*, December 16 2015.

Return to Journal of Integer Sequences home page.