There are no iterated morphisms that define the Arshon sequence and the σ-sequence

Sergey Kitaev

12th December 2002

Abstract

In [3], Berstel proved that the Arshon sequence cannot be obtained by iteration of a morphism. An alternative proof of this fact is given here. The σ-sequence was constructed by Evdokimov in order to construct chains of maximal length in the n-dimensional unit cube. It turns out that the σ-sequence has a close connection to the Dragon curve [8]. We prove that the σ-sequence can not be defined by iteration of a morphism.

1 Introduction and Background

In 1937, Arshon gave a construction of a sequence of symbols w over the alphabet $\{1, 2, 3\}$, constructed as follows: Let $w_1 = 1$. For $k \geq 1$, w_{k+1} is obtained by replacing the letters of w_k in odd positions thus:

$$1 \to 123,\ \ 2 \to 231,\ \ 3 \to 312$$

and in even positions thus:

$$1 \to 321,\ \ 2 \to 132,\ \ 3 \to 213.$$

Then

$$w_2 = 123, \quad w_3 = 123132312,$$

and each w_i is the initial subword of w_{i+1}, so the infinite symbolic sequence $w = \lim_{n\to\infty} w_n$ is well defined. It is called the *Arshon sequence*.

This method of constructing w is called the *Arshon Method* (**AM**), and ψ will denote the indicated map of the letters 1, 2, 3 according to position as described above.

We will denote the natural decomposition of w in 3-blocks by lower braces:

$$w = \underbrace{123}\underbrace{132}\underbrace{312}\ldots.$$

The paper by Arshon [2] was published in connection with the problem of constructing a nonrepetitive sequence on a 3-letter alphabet, that is, a sequence that does not contain any subwords of the type $XX = X^2$, where X is any non-empty word of a 3-letter alphabet. The sequence w has that property. The question of the existence of such a sequence, as well as the questions of the existence of sequences avoiding repetitions in other meanings, were studied in algebra [1, 9, 10], discrete analysis [5, 7] and in dynamical systems [14].

Any natural number n can be presented unambiguously as $n = 2^t(4s+\sigma)$, where $\sigma < 4$, and t is the greatest natural number such that 2^t divides n. If n runs through the natural numbers then σ runs through the sequence that we will call the *σ-sequence*. We let w_σ denote that sequence. Obviously, w_σ consists of 1s and 3s. The initial letters of w_σ are 11311331113313 . . . .

In [6, 17], Evdokimov constructed chains of maximal length in the n-dimensional unit cube using the σ-sequence. Originally, the σ-sequence was defined by the following inductive scheme:

$$C_1 = 1, \qquad D_1 = 3$$
$$C_{k+1} = C_k 1 D_k, \qquad D_{k+1} = C_k 3 D_k$$
$$k = 1, 2, \ldots$$

and $w_\sigma = \lim\limits_{k\to\infty} C_k$.

Our definition above of the σ-sequence is equivalent to this one.

The motivation for studying the σ-sequence is that this sequence has a close connection to the well-known *Dragon curve*, discovered by physicist John E. Heighway and defined as follows: we fold a sheet of paper in half, then fold in half again, and again, etc. and then unfold in such way that each crease created by the folding process is opened out into a 90-degree angle. The "curve" refers to the shape of the partially unfolded paper as seen edge on. If one travels along the curve, some of the creases will represent turns to the left and others turns to the right. Now if 1 indicates a turn to the right, and 3 to the left, and we start travelling along the curve indicating the turns, we get the σ-sequence [8].

Let Σ be an alphabet and $\Sigma^\star$ be the set of all words of Σ. A map $\varphi : \Sigma^\star \to \Sigma^\star$ is called a *morphism*, if we have $\varphi(uv) = \varphi(u)\varphi(v)$ for any $u, v \in \Sigma^\star$. It easy to see that a morphism φ can be defined by defining $\varphi(i)$ for each $i \in \Sigma$. The set of all rules $i \to \varphi(i)$ is called a *substitution system*. We create words by starting with a letter from the alphabet Σ and iterating the substitution system. Such a substitution system is called a *D0L (Deterministic, with no context Lindenmayer) system* [12]. D0L systems are a classical object of Formal Language Theory. They are interesting from mathematical point of view, but also have applications in theoretical biology [11].

Suppose a word $\varphi(a)$ begins with a for some $a \in \Sigma$, and that the length of $\varphi^k(a)$ increases without bound. The symbolic sequence $\lim\limits_{k\to\infty} \varphi^k(a)$ is called a *fixed point* of the morphism φ.

We now study classes of sequences that are defined by iterative schemes. There are many techniques to study sequences generated by morphisms [13]. It is reasonable to try to determine if a sequence under consideration can be obtained by iteration of a morphism.

Since the construction of the Arshon sequence w is similar to the iterated morphism scheme, and because w is constructed by two morphisms f_1 and f_2, applied depending on whether the letter position is even or odd, we might expect that there exists a morphism f which generates w. But this turns out not to be true. In [3], Berstel proved that the Arshon sequence cannot be obtained by iteration of a morphism. However, Berstel's proof appeals to advanced machinery, namely a sort of Wilf-Fine theorem due to Cobham [4]. In Theorem 1, we show that such machinery is not needed here, and offer a simple and direct proof.

Naturally a question arises as to the possibility of constructing w_σ using the iteration of a morphism, since such a construction could help us in studying w_σ, and thus in studying the Dragon curve. This also turns out not to be true, as shown in Theorem 10.

2 The Arshon Sequence

Theorem 1. *There does not exist a morphism whose fixed point is the Arshon sequence.*

Note. A corollary of this theorem is the non-existence of a morphism whose iteration gives the Arshon sequence. In fact, if such a morphism exists, it must have the property that 1 is mapped to $1X$ by the action of the morphism, where X is some word, and from this it follows that the Arshon sequence is a fixed point of this morphism.

Proof. It is enough to prove the non-existence of a morphism f with the property $w = f(w)$, since from the definition of a fixed point we have that if w is a fixed point of the morphism f then $w = f(w)$. Suppose there exists a morphism f such that $f(1) = X$, $f(2) = Y$, $f(3) = Z$ and $w = f(w)$. Moreover, in order to avoid dealing with the identity morphism, we assume that $|f(123)| > 3$. We do this since w is a fixed point of the identity morphism, but iterations of this morphism obviously do not give us the Arshon sequence. From all such morphisms f we choose a morphism with minimal length of X.

Clearly not all of $|X|$, $|Y|$ and $|Z|$ can be 0. Suppose for example then that $|X| = 0$, $|YZ| \neq 0$. Then w contains the subword $f(123132) = XYZXZY = YZZY$. If $|Z| \neq 0$ then w contains the repetition ZZ. Otherwise w contains repetition YY. This is impossible, hence $|X| \neq 0$. One can check in the same way that $|Y| \neq 0$ and $|Z| \neq 0$. Now $|X| + |Y| + |Z| \neq 3$, since otherwise $|f^\ell(1)| = 1$ for $\ell = 1, 2...$, and w is not a fixed point of the morphism f. Now

$$f(w) = w = XYZXZYZXY\dots,$$

hence X consists of $|X|$ of the first letters of w, Y is $|Y|$ of the following letters, and Z is $|Z|$ of the letters following that.

We will use upper braces to show the decomposition of w into f-blocks (that is, to show the disposition of the words X, Y and Z in w). We have

$$w = \overbrace{\underbrace{123}\,\underbrace{132}\dots a_{|X|}}^{X}\,\overbrace{a_{|X|+1}\dots a_{|X|+|Y|}}^{Y}\,\overbrace{a_{|X|+|Y|+1}\dots a_{|X|+|Y|+|Z|}}^{Z}\,\overbrace{a_{|X|+|Y|+|Z|+1}\dots\dots}^{X}\dots,$$

where all a_i are letters of the alphabet $\{1, 2, 3\}$.

Lemma 2. *We have* $|X| + |Y| + |Z| \equiv 0 \pmod 3$.

Proof. From the structure of w, the frequencies of 1, 2, 3 in w coincide, hence the frequencies of these letters in $f(w) = w$ coincide as well. But this is only possible when $|X| + |Y| + |Z| \equiv 0 \pmod 3$. Indeed, otherwise there are two letters, whose frequencies in $f(123) = XYZ$ do not coincide, which implies that the frequencies of these letters in $f(w) = w$ do not coincide as well, since w can be written as $w = W_1 W_2 W_3 \ldots$, where W_i is a permutation of the letters X, Y and Z. $\square$

Lemma 3. *The situation* $|X| \equiv |Y| \equiv |Z| \equiv 0 \pmod 3$ *is impossible.*

Proof. Suppose $|X| \equiv |Y| \equiv |Z| \equiv 0 \pmod 3$. Then X, Y and Z consist of a whole number of 3-blocks. The properties of ψ mean that the morphism g given by $g(1) = \psi^{-1}(X)$, $g(2) = \psi^{-1}(Y)$, $g(3) = \psi^{-1}(Z)$ satisfies $w = g(w)$. By the minimality of f, $|g(123)| = 3$, so that $g(123) = 123$. This implies that $f(1) = 123$, $f(2) = 132$, $f(3) = 312$. However, this impossible, since then $f(1231) = 123132312123$ is not a prefix w. $\square$

We define the *Nth f-block X* to be the block X obtained by considering the decomposition of w into f-blocks and then taking the Nth block that corresponds to the word X (possibly skipping some blocks that corresponds to the words Y and Z without counting them). Thus, for example, when we say "the 4th f-block X", we mean not the following marked block $w = XYZ\mathbf{X}ZY\ldots$, but the following marked block $w = XYZXZYZXYZY\mathbf{X}\ldots$.

Lemma 4. *With the assumption of the existence of the morphism f,* $|X| \ \leq \ 5$.

Proof. Suppose $|X| \geq 6$, that is, $X = 123132\ldots$. If $|X| \equiv 2 \pmod 3$ ($|X| \equiv 1 \pmod 3$), then $|X| \geq 7$ and using Lemma 2 we consider the 4th f-block $X = 12\underbrace{313}\ldots$ ($X = 1\underbrace{231}\underbrace{323}\ldots$). This contradicts the **AM**. Hence $|X| \equiv 0 \pmod 3$.

It follows from Lemma 3 that the situation $|Y| \equiv 0 \pmod 3$ is impossible. If $|Y| \equiv 1 \pmod 3$ ($|Y| \equiv 2 \pmod 3$), then we consider the 10th (3rd) f-block $X = 12\underbrace{313}2\ldots$ and it brings us to a contradiction with the **AM**. Hence if $|X| \geq 6$ then the morphism f can not exist. $\square$

Lemma 5. *With the assumption of the existence of the morphism f,* $|X| \ \neq \ 1$.

Proof. If $|X| = 1$, then $X = 1$ and the length of the words $f^k(1)$ for $k = 1, 2, \ldots$ does not increase, whence w is not a fixed point of the morphism f. This is a contradiction. $\square$

Lemma 6. *With the assumption of the existence of the morphism f,* $|X| \ \neq \ 2$.

Proof. Suppose $|X| = 2$, that is $X = 12$.

We have $|X| \equiv 2 \pmod 3$, hence, using Lemma 2, we have $|Y|+|Z| \equiv 1 \pmod 3$.

We consider the 2nd f-block X and the f-block Z next after it. It can be seen that Z begins with 3. We consider the 4th f-block X and Y preceding it and find that Y ends with 3. But then, considering YZ, which is a subword of w, we see, that 33 is a subword of w, which is impossible. That is for $|X| = 2$ the morphism f cannot exist. □

The 3-blocks 123, 231, 312 are said to be *odd* 3-blocks. All other 3-blocks are said to be *even*.

Lemma 7. *With the assumption of the existence of the morphism f, $|X| \neq 3$.*

Proof. Suppose $|X| = 3$, that is $X = 123$.

We have $|X| \equiv 0 \pmod 3$, hence, using Lemma 2 we have $|Y|+|Z| \equiv 0 \pmod 3$. Considering the **AM**, the 2nd f-block X must be an odd 3-block, hence $|Y| + |Z| \equiv 1 \pmod 2$.

Let $|Z| \geq 2$. Then the 2nd f-block Z begins with an even 3-block, and the 3rd Z begins with an odd 3-block. This is impossible. (Note that 2 letters define the evenness of the 3-block unambiguously.) Thus $|Z| = 1$.

Let $|Y| \geq 2$. In XYZ (or in an arbitrary permutation of these letters) there is an even number of 3-blocks, so the 9th f-block Y begins with an odd 3-block, but the 1st Y begins with an even 3-block. Hence $|Y| = 1$.

This is a contradiction with $|Y| + |Z| \equiv 0 \pmod 3$ (and also a contradiction with $|Y| + |Z| \equiv 1 \pmod 2$). That is for $|X| = 3$ the morphism f cannot exist. □

Lemma 8. *With the assumption of the existence of the morphism f, $|X| \neq 4$.*

Proof. Suppose $|X| = 4$, that is $X = 1231$.

We have $|X| \equiv 1 \pmod 3$, hence, using Lemma 2, we have $|Y|+|Z| \equiv 2 \pmod 3$.

We have $|Y| \geq 2$, since otherwise $Y = 3$ and hence XYX which is a subword of w, contains 3131, which is impossible. Hence $Y = 32\ldots$. We consider ZX and ZY and see that Z ends with 2, since otherwise w contains $Z1$ and $Z3$, and thus the square 11 or 33. Now $|Z| \geq 2$, since otherwise $Z = 2$ and XZX which is a subword of w, contains 1212, which is impossible. Hence $Z = \ldots 32$, or $Z = \ldots 12$. The former is impossible since 3232 is contained in ZY, and hence in w. The latter is impossible too, since considering the 9th f-block Z and the f-block X following it, we obtain $ZX = \ldots \underbrace{121}\underbrace{231}$, which contradicts the **AM**. That is for $|X| = 4$ the morphism f cannot exist. □

Lemma 9. *With the assumption of the existence of the morphism f, $|X| \neq 5$.*

Proof. Suppose $|X| = 5$, that is $X = 12313$.

We have $|X| \equiv 2 \pmod 3$, hence, using Lemma 2, we have $|Y|+|Z| \equiv 1 \pmod 3$. Then the 4th f-block is $X = 12\underbrace{313}$, which is a contradiction with the **AM**. That is if $|X| = 5$ then the morphism f cannot exist. □

From Lemmas 4 – 9 we have a contradiction with the assumption of the existence of the morphism f. This proves Theorem 1. □

Remark. In [2], Arshon gave the construction of a nonrepetitive sequence w_n for an n-letter alphabet, where n is any natural number greater than or equal to 3. It is easy to see that, for even n, there exists a morphism f_n that defines w_n. Namely, for $1 \leq i \leq n$, one has:

$$f_n(i) = \begin{cases} i(i+1)\dots n12\dots i-1, & \text{if } i \text{ is odd,} \\ (i-1)(i-2)\dots 1n(n-1)\dots i, & \text{if } i \text{ is even.} \end{cases}$$

Theorem 1 shows that for $n = 3$ such a morphism does not exist. However, whether there exists a morphism defining w_n for arbitrary odd n is still an open question.

3 The σ-sequence

Theorem 10. *There does not exist a morphism whose iteration defines the sequence w_σ.*

Proof. Suppose there exists a morphism f, such that $f(1) = X$, $f(3) = Y$ and $w_\sigma = \lim_{k\to\infty} f^k(1)$. Obviously, X consists of the first $|X|$ letters of w, where $|X|$ is the length of X.

Lemma 11. *The subsequence of w_σ consisting of the letters in odd positions is the alternating sequence of* 1*s and* 3*s:* $1313131\dots$.

Proof. The odd positions of w_σ correspond to the odd numbers $n = 2^0(4s+\sigma) = 4s+\sigma$, so clearly σ alternates between 1 and 3. □

Lemma 12. *If there exists a morphism f whose iteration gives w_σ then* $|X| \equiv 0 \pmod 4$.

Proof. It is easy to see that $f(1) = 1X^{(1)}$, where $|X^{(1)}| \geq 1$, since otherwise $|f^k(1)| = 1$, for $k = 1, 2, 3\dots$, so w_σ cannot be obtained by iterating f.

Suppose $|X^{(1)}| = 1$, that is $f(1) = 11$. But then w_σ consists of 1s only, which is impossible, hence $f(1) = 11X^{(2)}$, where $|X^{(2)}| \geq 1$.

Suppose $|X^{(2)}| = 1$, that is $f(1) = 113$. Since w_σ has the subword 111, then w_σ has a subword $f(111) = 113113113$. If $f(111)$ begins with a letter in an odd position, then the marked letters **1**1**3**1**1**3**1**1**3**, read from left to right will make up consecutive letters of w_σ in odd positions. This contradicts Lemma 11. If $f(111)$ begins with a letter in an even position, then marking letters in odd positions will lead to the same contradiction with Lemma 11, hence $f(1) = 113X^{(3)}$, where $|X^{(3)}| \geq 1$.

Suppose $|X^{(3)}| = 1$, that is $f(1) = 1131$. Then $f^2(1) = 1131\mathbf{1}131Y1131$ and the marked letter does not coincide with the letter of w_σ standing in the same place, hence $f(1) = 1131X^{(4)}$, where $|X^{(4)}| \geq 1$.

If $|X|$ is odd, then the marked letters in $f^2(1) = 1131X^{(4)}\mathbf{1131}X^{(4)}\ldots$ are two consecutive letters in odd places. This contradicts Lemma 11. Hence $|X|$ is even.

We have $f^2(1) = XX\ldots = X1131X^{(4)}\ldots$, whence the next-to-last letter of X is in an odd position and is equal to 3, since otherwise two consecutive 1 in w_σ stand at odd places, which contradicts Lemma 11. The natural number which corresponds to the next-to-last letter of X is written as $2^0(4s+3)$, the next number is equal to $|X|$ and to $2^0(4s+3)+1 = 4(s+1) \equiv 0 \pmod 4$. □

The following Lemma is straightforward to prove.

Lemma 13. *If $n_1 = 2^{t_1}(4s_1+1)$, $n_2 = 2^{t_2}(4s_2+1)$, $n_3 = 2^{t_3}(4s_3+3)$ and $n_4 = 2^{t_4}(4s_4+3)$ then n_1n_2, n_3n_4 can be written as $2^t(4s+1)$, and n_1n_3 as $2^t(4s+3)$.*

It follows from Lemma 12 that $|X| = 4t$.

Suppose X ends with 1 (the case when X ends with 3 is similar), that is at the $(4t)$th position in X we have 1. According to the multiplication by 2 does not change σ, so at the $(2t)$th position in X we have 1.

Consider $f^2(1) = X\mathbf{X}\ldots$. The letters of the marked X occupy the positions of $f^2(1)$ from $(4t+1)$th to $(8t)$th. Since $X = \mathbf{X}$, then at the $(6t)$th place we have 1. But $6t = 3(2t)$, whence, by Lemma 13, at the $(2t)$th and the $(6t)$th places there must stand different letters. This is a contradiction and Theorem 10 is proved. □

References

[1] Adian S. I.: *The Burnside problem and identities in groups.* Translated from the Russian by John Lennox and James Wiegold. Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas], 95. Springer-Verlag, Berlin-New York, (1979). xi+311 pp.

[2] Arshon S. E.: The Proof of the Existence of n - letter Non-Repeated Asymmetric Sequences, Math. collected papers, New series, Publ. H. AS USSR, Vol. **2**, Issue 3, Moscow (1937), 769–779. (Russian)

[3] Berstel J.: Mots sans carré et morphismes itérés, Discrete Math. **29** (1979), 235–244.

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

[5] Dekking F. M.: Strongly non-repetitive sequences and progression-free sets, Journal Com. Theory, Vol. **27-A**, No. 2 (1979), 181–185.

[6] Evdokimov A. A.: On the Maximal Chain Length of an Unit n-dimensional Cube, Maths Notes **6**, No. 3 (1969), 309–319. (Russian)

[7] Evdokimov A. A.: Strongly asymmetric sequences generated by a finite number of symbols, Dokl. Akad. Nauk SSSR, 179 (1968), 1268–1271. (Russian) English translation in: Soviet Math. Dokl., 9 (1968), 536–539.

[8] Evdokimov A. A.: Private communication (2001).

[9] Justin J.: Characterization of the repetitive commutative semigroups, Journal of Algebra (1972), no. 21, 87–90.

[10] Kolotov A. T.: Aperiodic sequences and functions of the growth of algebras, Algebra i Logika **20** (1981), no. 2, 138–154. (Russian)

[11] Lindenmayer A.: Mathematical models for cellular interaction in development, Parts I and II, Journal of Theoretical Biology, 18 (1968), 280–315.

[12] Lindenmayer A., Rozenberg G.: *Automata, languages, development*, North-Holland Publishing Co., Amsterdam-New York-Oxford (1976), viii+529 pp.

[13] Lothaire M.: *Combinatorics on Words*, Encyclopedia of Mathematics, Vol. **17**, Addison-Wesley (1986).

[14] Morse M., Hedlung G.: Unending chess, symbolic dinamics and a problem in semigroups, Duke Math. Journal, Vol. **11**, No. 1 (1944), 1–7.

[15] Pleasants P.: Non-repetitive sequences, Proc. Camb. Phil. Soc., Vol. **68** (1970), 267–274.

[16] Salomaa A.: *Jewels of Formal Language*, Theory Computer Science Press, (1981).

[17] *Discrete Mathematics and Mathematical Questions of Cybernetics*, Vol. **1**, Moscow (1974), 112–116. (Russian)