On Uniformly Recurrent Morphic Sequences

(Möbius contest version*)

Yuri Pritykin†

September 26, 2007

Abstract

In the paper we mainly deal with two well-known types of infinite words: morphic and uniformly recurrent (=almost periodic). We discuss the problem of finding criterion of uniform recurrence for morphic sequences and give effective polynomial-time such criterion in two particular cases: pure morphic sequences and automatic sequences. We also prove that factor complexity of arbitrary uniformly recurrent morphic sequence is at most linear.

1 Introduction

Many problems of decidability in combinatorics on words are of great interest and difficulty. Here we deal with two well-known types of symbolic infinite sequences — morphic and uniformly recurrent — and try to understand connections between them. Namely, we are trying to find an algorithmic criterion which given a morphic sequence decides whether it is uniformly recurrent.

Though the main problem still remains open, we propose polynomial-time algorithms solving the problem in two important particular cases: for pure morphic sequences (Section 3) and for automatic sequences (Section 4). In Section 6 we discuss the general problem and give a curious result supporting the conjecture of decidability.

Some attempts to solve the problem were already done. In [5] A. Cobham gives a criterion for automatic sequence to be uniformly recurrent. But even if his criterion gives some effective

*This is a joint paper by François Nicolas (University of Turku, nicolas@cs.helsinki.fi) and Yuri Pritykin, being enhanced version of the paper [20]. Generally the work is done by the second author, except the following. The main theorem of Section 3, namely Theorem 3.4, was formulated in [20] only for non-erasing morphisms, and F. N. noticed that it can be modified to hold for all morphisms; he also gave a reference to the paper [8] with similar considerations. F. N. attracted the attention to the subject of factor complexity of uniformly recurrent morphic sequences and noticed that factor complexity of uniformly recurrent **pure** morphic sequences is at most linear. In the current version of the paper this result is covered by Theorem 5.1 for arbitrary uniformly recurrent morphic sequences. F. N. also carefully proofread all the paper, significantly improved explanations in many places, and found many important references.

†Moscow State University, Russia, e-mail: yura@mccme.ru. The work was partially supported by RFBR grants 06-01-00122, 05-01-02803, NSh-5351.2006.1, Kolmogorov grant of Institute of New Technologies, and Möbius Contest Foundation for Young Scientists.

procedure solving the problem (which is not clear from his result), this procedure could not be fast. We construct a polynomial-time algorithm solving the problem. In [13] A. Maes deals with pure morphic sequences and finds a criterion for them to belong to a slightly different class of generalized uniformly recurrent sequences (he calls them almost periodic). And again, his algorithm does not seem to be polynomial-time. The problem of determining ultimate periodicity for pure morphic sequences was solved independently in [10] and [17].

In addition, in Section 5 we deal with very natural and simple combinatorial characteristic of words, namely factor complexity. We prove that factor complexity of a sequence that is both morphic and uniformly recurrent, is at most linear.

2 Preliminaries

Denote the set of natural numbers $\{0, 1, 2, \dots\}$ by $\mathbb{N}$ and the binary alphabet $\{0, 1\}$ by $\mathbb{B}$. Let A be a finite alphabet. We deal with sequences over this alphabet, i. e., mappings $x\colon \mathbb{N} \to A$, and denote the set of these sequences by $A^{\mathbb{N}}$. Sequences are also called infinite words.

Denote by A^* the set of all finite words over A including the empty word Λ. If $i \leqslant j$ are natural, denote by $[i, j]$ the segment of $\mathbb{N}$ with ends in i and j, i. e., the set $\{i, i{+}1, i{+}2, \dots, j\}$. Also denote by $x[i, j]$ a subword $x(i)x(i+1)\dots x(j)$ of a sequence x. A segment $[i, j]$ is an occurrence of a word $u \in A^*$ in a sequence x if $x[i, j] = u$. We say that $u \neq \Lambda$ is *a factor* of x if u occurs in x. A word of the form $x[0, i]$ for some i is called *prefix* of x, and respectively a sequence of the form $x(i)x(i+1)x(i+2)\dots$ for some i is called *suffix* of x and is denoted by $x[i, \infty)$. Denote by $|u|$ the length of a word u. The occurrence $u = x[i, j]$ in x is k-aligned if $k|i$.

A sequence x is periodic if for some T we have $x(i) = x(i + T)$ for each $i \in \mathbb{N}$. This T is called a period of x. We denote by $\mathcal{P}$ the class of all periodic sequences. Let us consider extensions of this class.

A sequence is called *recurrent* if every its factor occurs in this sequence infinitely many times.

A sequence x is called *uniformly recurrent*[1] if for every factor u of x there exists a number l such that every l-length factor of x contains at least one occurrence of u (and therefore u occurs in x infinitely many times). Obviously, to show uniform recurrence of a sequence it is sufficient to check the mentioned condition only for all prefixes but not for all factors (and even for some increasing sequence of prefixes only). Denote by $\mathcal{UR}$ the class of all uniformly recurrent sequences.

Let A, B be finite alphabets. A mapping $\phi\colon A^* \to B^*$ is called *a morphism* if $\phi(uv) = \phi(u)\phi(v)$ for all $u, v \in A^*$. A morphism is obviously determined by its values on single-letter words. A morphism is *non-erasing* if $|\phi(a)| \geqslant 1$ for each $a \in A$. A morphism is *k-uniform* if $|\phi(a)| = k$ for each $a \in A$. A 1-uniform morphism is called *coding*. A morphism is *growing* if $|\phi(a)| \geqslant 2$ for each $a \in A$. A morphism is called *irreducible* if for each $a, b \in A$ there exists n such that $\phi^n(a)$ contains b. A morphism is called *primitive* if there exists n such that for each $a, b \in A$ the word $\phi^n(a)$ contains b. Every primitive morphism is irreducible, but the converse does not hold in general.

[1] It was called *strongly* or *strictly almost periodic* in [15, 18]; sometimes also called *almost periodic* [19, 20].

For $x \in A^{\mathbb{N}}$ denote
$$\phi(x) = \phi(x(0))\phi(x(1))\phi(x(2))\dots$$
Further we mainly consider morphisms of the form $A^* \to A^*$ (but codings are of the form $A \to B$, that in fact does not matter, they can be also of the form $A \to A$ without loss of generality). Let $\phi(s) = su$ for some $s \in A$, $u \in A^*$. Then for all natural $m < n$ the word $\phi^n(s)$ begins with the word $\phi^m(s)$, so $\phi^\infty(s) = \lim_{n\to\infty} \phi^n(s) = su\phi(u)\phi^2(u)\phi^3(u)\dots$ is well-defined. If $\forall n\ \phi^n(u) \neq \Lambda$, then $\phi^\infty(s)$ is infinite. In this case we say that ϕ is *prolongable* on s. Sequences of the form $h(\phi^\infty(s))$ for a coding $h\colon A \to B$ are called *morphic*, of the form $\phi^\infty(a)$ are called *pure morphic*.

Note that there exist uniformly recurrent sequences that are not morphic (in fact, $\mathcal{UR}$ has cardinality continuum (e. g., see [15]), while the set of morphic sequences is obviously countable), as well as there exist morphic sequences that are not uniformly recurrent (you will find examples later). Our main goal is to determine whether a morphic sequence is uniformly recurrent or not given its constructive description.

First of all, observe the following

Proposition 2.1. *A sequence $\phi^\infty(s)$ is uniformly recurrent iff s occurs in this sequence infinitely many times with bounded distances.*

Proof. In one direction the statement is obviously true by definition.

Suppose now that s occurs in $\phi^\infty(s)$ infinitely many times with bounded distances. Then for every m the word $\phi^m(s)$ also occurs in $\phi^\infty(s)$ infinitely many times with bounded distances. But every word u occurring in $\phi^\infty(s)$ occurs in some prefix $\phi^m(s)$ and thus occurs infinitely many times with bounded distances. $\square$

For a morphism $\phi\colon \{1,\dots,n\} \to \{1,\dots,n\}$ we can define *an incidence matrix* M_ϕ, such that $(M_\phi)_{ij}$ is a number of occurrences of symbol i into $\phi(j)$. One can easily check that for each l we have $M_\phi^l = M_{\phi^l}$.

Clearly, a morphism ϕ is primitive iff for some l all the entries of M_ϕ^l are positive. For prolongable morphisms the notions of primitiveness and irreducibility coincide.

Let us construct an oriented *incidence graph* G_ϕ of a morphism ϕ. Let its set of vertices be A. In G_ϕ edges go from $b \in A$ to all the symbols occurring in $\phi(b)$.

For $\phi^\infty(s)$ it can easily be found using G_ϕ which symbols from A really occur in this sequence. Indeed, these symbols form the set of all vertices that can be reached from s. So from now on without loss of generality we assume that all the symbols from A occur in $\phi^\infty(s)$.

It is not difficult to formulate a criterion of recurrence for pure morphic sequences.

Proposition 2.2. *Let A be an alphabet, $s \in A$, and let $\phi\colon A^* \to A^*$ be a morphism prolongable on s. The following four assertions are equivalent:*
1) the pure morphic sequence $\phi^\infty(s)$ is recurrent;
2) the letter s occurs infinitely many times in $\phi^\infty(s)$;
3) the letter s occurs at least twice in $\phi^\infty(s)$;
4) the letter s occurs twice in $\phi(s)$ or there exists a letter $a \neq s$ occurring in $\phi^\infty(s)$ such that s occurs in $\phi(a)$.

Proof. Left to the reader. $\square$

The situation is not that easy in the case of uniform recurrence.

A morphism is irreducible if and only if its graph of incidence is strongly connected, i. e., there exists an oriented path between every two vertices. For prolongable morphisms this is also a criterion of primitiveness. This reformulation of the primitiveness notion seems to be more appropriate for computational needs. By Proposition 2.1 (and the observation that codings preserve uniform recurrence) morphic sequences generated by primitive morphisms are always uniformly recurrent. Moreover, in the case of growing morphisms this sufficient condition is also necessary (and this is a polynomial-time algorithmic criterion in that case). However when we generalize this case even on non-erasing morphisms, it is not enough to consider only the incidence graph or even the incidence matrix (which contains more information than the graph), as it can be seen from the following example.

Let ϕ_1 be as follows: $0 \to 01$, $1 \to 120$, $2 \to 2$, and ϕ_2 be as follows: $0 \to 01$, $1 \to 210$, $2 \to 2$. Then these two morphisms have identical matrices, but $\phi_1^\infty(0)$ is uniformly recurrent, while $\phi_2^\infty(0)$ is not. Indeed, in $\phi_2^\infty(0)$ there are arbitrary long segments like $222\ldots22$, so $\phi_2^\infty(0) \notin \mathcal{UR}$. There is no such problem in $\phi_1^\infty(0)$. Since 0 occurs in both $\phi_1(0)$ and $\phi_1(1)$, and 22 does not occur in $\phi_1^\infty(0)$, it follows that 0 occurs in $\phi_1^\infty(0)$ with bounded distances. Thus $\phi_1^m(0)$ for every $m \geqslant 0$ occurs in $\phi_1^\infty(0)$ with bounded distances, so $\phi_1^\infty(0) \in \mathcal{UR}$. See Theorem 3.1 for a general effective criterion of uniform recurrence in the case of pure morphic sequences.

To introduce a bit the notion of uniform recurrence, let us formulate an interesting result on this topic. It seems to be first proved in [5], but also follows from the results of [19]. For $x \in A^{\mathbb{N}}$, $y \in B^{\mathbb{N}}$ define $x \times y \in (A \times B)^{\mathbb{N}}$ such that $(x \times y)(i) = \langle x(i), y(i) \rangle$.

Proposition 2.3. *If x is uniformly recurrent and y is periodic, then $x \times y$ is uniformly recurrent.*

Proof. Prove the proposition for y of the form $y = 012\ldots(m-1)012\ldots(m-1)01\ldots$ over $\Sigma_m = \{0, 1, 2, \ldots, m-1\}$.

We say that u occurs in x modulo i, where $i \in \Sigma_m$, if $u \times [i, i+1, \ldots, m-1, 0, 1, \ldots, m-1, 0, \ldots, i + |u| - 1 \pmod{m}]$ occurs in $x \times y$. Our aim is to prove that if u occurs in x modulo i, then it happens infinitely many times with bounded distances.

Let $A \subseteq \Sigma_m$ be the set of all i such that u occurs in x modulo i at least once, and let $w = x[0, k]$ be a prefix of x such that for each $i \in A$ there exists an occurrence of u modulo i in w. Let $x[p, q]$ be an occurrence of w in x. Then for each $i \in A$ the word u occurs in $x[p, q]$ modulo $i + p \pmod{m}$. Thus $B = \{i + p \pmod{m} : i \in A\} \subseteq A$ by definition of A, but $|B| = |A|$, hence $B = A$.

Thus u occurs modulo i in each occurrence of w in x. But $x \in \mathcal{UR}$, and w occurs in x infinitely many times with bounded distances. $\square$

3 Pure Morphic Sequences

Here we consider morphic sequence of the form $\phi^\infty(s)$. We present an algorithm that determines whether a morphic sequence $\phi^\infty(s)$ is uniformly recurrent given an alphabet A, a morphism ϕ and a symbol $s \in A$.

The following definitions are due to Pansiot [16]. A word $w \in A^*$ is called *ϕ-bounded* if the sequence $(w, \phi(w), \phi^2(w), \phi^3(w), \dots)$ is eventually periodic. A word $w \in A^*$ is called *ϕ-growing* if $|\phi^n(w)| \to \infty$ as $n \to \infty$. Obviously, every word from A^* is either ϕ-bounded or ϕ-growing. A word $w \in A^*$ is *ϕ-eventually-erased* if $\phi^n(w) = \Lambda$ for some n.

The following theorem gives criterion for pure morphic sequences to be uniformly recurrent.

Theorem 3.1. *Let A be an alphabet, $s \in A$, and let $\phi\colon A^* \to A^*$ be a morphism prolongable on s. The pure morphic sequence $\phi^\infty(s)$ is uniformly recurrent iff it satisfies the following two properties:*
1) for every ϕ-growing letter a occurring in $\phi^\infty(s)$, there exists an integer $n \in \mathbb{N}$ such that s occurs in $\phi^n(a)$, and
2) only finitely many ϕ-bounded words are factors of $\phi^\infty(s)$.

Proof. $\Rightarrow$. Assume that $\phi^\infty(s)$ is uniformly recurrent. Then there exists a positive integer l such that s occurs in every l-length factor of $\phi^\infty(s)$.

1) Let a be a ϕ-growing letter occurring in $\phi^\infty(s)$. For every $n \in \mathbb{N}$, $\phi^n(a)$ is a factor of $\phi^\infty(s)$, and if n is large enough, then $\phi^n(a)$ has length $\geqslant l$. Hence, s occurs in $\phi^n(a)$ for all n large enough.

2) Since letter s is ϕ-growing, s cannot occur in any ϕ-bounded factor of $\phi^\infty(s)$. Hence, all ϕ-bounded factors of $\phi^\infty(s)$ have lengths smaller than l.

$\Leftarrow$. Assume that both properties 1) and 2) hold. Property 1) implies that there exists a positive integer n such that for every ϕ-growing letter a occurring in $\phi^\infty(s)$, s occurs in $\phi^n(a)$. According to Property 2), there exists a positive integer M such that every ϕ-bounded factor of $\phi^\infty(s)$ has length smaller than M. Let K denote the maximum length of $\phi^n(a)$ over all $a \in A$.

Let w be a factor of $\phi^\infty(s)$ with length $(K+1)M$. There exists an M-length factor v of $\phi^\infty(s)$ such that $\phi^n(v)$ is a factor of w. Since v is longer than every ϕ-bounded factor of $\phi^\infty(s)$, some ϕ-growing letter a occurs in v. Hence, s occurs in $\phi^n(a)$, $\phi^n(a)$ is a factor of $\phi^n(v)$, and $\phi^n(v)$ is a factor of w. It follows that s occurs w.

We have thus shown that s occurs in every factor of $\phi^\infty(s)$ with length $(K+1)M$, and thus $\phi^\infty(s)$ is uniformly recurrent according to Proposition 2.1. □

Now we explain how to get a polynomial-time criterion. First in Proposition 3.2 we give different reformulations of Property 2) from Theorem 3.1. Then we reformulate the uniform recurrence criterion such that it can easily be checked in polynomial time.

Proposition 3.2 (Ehrenfeucht, Rozenberg [8]). *Let A be an alphabet, $s \in A$, and let $\phi\colon A^* \to A^*$ be a morphism prolongable on s. The following three properties are equivalent:*
1) infinitely many ϕ-bounded words are factors of $\phi^\infty(s)$;
2) there exist a natural n, a letter a occurring in $\phi^\infty(s)$ and two words u, $v \in A^*$ satisfying conditions:

(i) u is not ϕ-eventually-erased,

(ii) u is ϕ-bounded, and

(iii) either $\phi^n(a) = uav$ or $\phi^n(a) = vau$.

3) there exists a non-empty ϕ-bounded word w such that w^n is a factor of $\phi^\infty(s)$ for every $n \in \mathbb{N}$.

Proof. 3) $\Rightarrow$ 1) is straightforward. 1) $\Rightarrow$ 2) was proved in [8]. 2) $\Rightarrow$ 3) is easy. $\square$

Suppose we have A, ϕ, and $s \in A$, such that $|A| = n$, $\max_{b\in A} |\phi(b)| = k$, s begins $\phi(s)$. Remember that we suppose that all the symbols from A appear in $\phi^\infty(s)$.

Divide A into two parts. Let I_ϕ be the set of all ϕ-growing (or ϕ-**i**ncreasing) symbols, and let B_ϕ be the set of all ϕ-**b**ounded symbols. Define also $E_\phi \subseteq B_\phi$ to be the set of all ϕ-eventually-**e**rased symbols.

Lemma 3.3. *One can find I_ϕ, B_ϕ, and E_ϕ in* $\mathrm{poly}(n,k)$*-time.*

Proof. First, consider an equivalence relation "$\equiv$" on vertices of G_ϕ: $a \equiv b$ iff a can be reached from b, and vice versa. Obviously, if $a \equiv b$, then $|\phi^m(a)| = \Theta(|\phi^m(b)|)$ as $m \to \infty$. Construct a new graph H_ϕ with vertices being equivalence classes of "$\equiv$". An edge goes from A to B in H_ϕ, if $\exists a \in A\ \exists b \in B$ such that $\phi(a)$ contains b. Define for each vertex A in H_ϕ the number $\kappa_A = \max\{\text{the number of occurrences of symbols from } A \text{ in } \phi(a) : a \in A\}$. Define $S_i = \{B \text{ in } H_\phi : \max\{\kappa_A : A \text{ can be reached from } B\} = i\}$. Obviously, H_ϕ, all κ_A and S_i can be computed in polynomial time.

It it not difficult to see that $\forall i \geqslant 2\ \forall A \in S_i\ \forall a \in A\ |\phi^m(a)| \to \infty$ as $m \to \infty$, and thus $a \in I_\phi$. Also, $\forall A \in S_0\ \forall a \in A\ |\phi^m(a)| \to 0$ as $m \to \infty$, and thus $a \in E_\phi$. In fact, every $A \in S_0$ is a singleton.

Now consider the subgraph induced by S_1. We have $S_1 = U \sqcup V$, where $U = \{A : \kappa_A = 0\}$, $V = \{A : \kappa_A = 1\}$. Further, we have $V = X \sqcup Y$, where $X = \{A \in V : \text{some other } B \in V \text{ can be reached from } A\}$, $Y = V \setminus X$. It is not difficult to see that $\bigcup_{A\in X} A \subseteq I_\phi$, $\bigcup_{A\in Y} A \subseteq B_\phi$. Further, if some $B \in X$ can be reached from $A \in U$, then $A \subseteq I_\phi$, otherwise $A \subseteq B_\phi$. Clearly, $\forall A \in S_1\ \forall a \in A\ a \notin E_\phi$.

Obviously, everything here can be checked in polynomial time. $\square$

A word is ϕ-eventually-erased iff it consists of ϕ-eventually-erased symbols. Thus one can easily check whether a given word is ϕ-eventually-erased.

Construct a labeled *prefix graph* L_ϕ. Its set of vertices is I_ϕ. From each vertex b exactly one edge goes off. To construct this edge, find a representation $\phi(b) = ucv$, where $c \in I_\phi$, u is the maximal prefix of $\phi(b)$ containing only symbols from B_ϕ. It follows from the definitions of I_ϕ and B_ϕ that u does not coincide with $\phi(b)$, that is why this representation is correct. Then construct in L_ϕ an edge from b to c and write u on it.

Analogously we construct *a suffix graph* R_ϕ. (In this case we find representation $\phi(b) = vcu$ where $u \in B_\phi^*$, $c \in I_\phi$, and write u on the edge.)

Now we formulate a constructive version of the criterion given in Theorem 3.1.

Theorem 3.4. *A sequence $\phi^\infty(s)$ is uniformly recurrent iff it satisfies the following two properties:*
1) G_ϕ restricted to I_ϕ is strongly connected, and
2) in both graphs L_ϕ and R_ϕ, on each edge of each cycle, ϕ-eventually-erased word is written.

Proof. Property 1) of this theorem is obviously equivalent to Property 1) of Theorem 3.1. Proposition 3.2 explains why the same is true with Properties 2). $\square$

Let us consider examples with ϕ_1 and ϕ_2 from the end of Section 2. In both cases $I_\phi = \{0,1\}$, $B_\phi = \{2\}$. On every edge of R_ϕ in both cases Λ is written. Almost the same is true for L_ϕ: the only difference is about the edge going from 1 to 1. In the case of ϕ_1 an empty word is written on this edge, while in the case of ϕ_2 a word 2 is written. The word 2 is not eventually erased since its image is 2. That is why $\phi_1^\infty(0)$ is uniformly recurrent, while $\phi_2^\infty(0)$ is not.

Corollary 3.5. *For a growing morphism ϕ a sequence $\phi^\infty(s)$ is uniformly recurrent iff ϕ is primitive.*

Corollary 3.6. *There exists a* poly(n,k)*-algorithm that says whether $\phi^\infty(s)$ is uniformly recurrent.*

Proof. Conditions from Theorem 3.4 can easily be checked in polynomial time. □

It also seems useful to formulate an explicit version of the criterion for the binary case.

Corollary 3.7. *For $\phi\colon \mathbb{B} \to \mathbb{B}$ that is prolongable on 0, a sequence $\phi^\infty(0)$ is uniformly recurrent iff one of the following conditions holds:*
1) $\phi(0)$ contains only 0s;
2) $\phi(1)$ contains 0;
3) $\phi(1) = \Lambda$;
4) $\phi(1) = 1$ and $\phi(0) = 0u0$ for some word u.

4 Uniform Morphisms

Now we deal with morphic sequences obtained by uniform morphisms. In Subsection 4.1 we present a polynomial-time criterion of uniform recurrence in this case. Final version of the criterion is in Theorem 4.7. Corollary 4.8 is to explain why the criterion from Theorem 4.7 is polynomial-time. In Subsection 4.2 we are discussing recurrence criterion as a related problem.

Suppose we have an alphabet A, a morphism $\phi\colon A^* \to A^*$, a coding $h\colon A \to B$, and $s \in A$, such that $|A| = n$, $|B| \leqslant n$, $\forall b \in A\ |\phi(b)| = k$, s begins $\phi(s)$. We are interested in whether $h(\phi^\infty(s))$ is uniformly recurrent. Note that the class of sequences of the form $h(\phi^\infty(s))$ with ϕ being k-uniform coincides with the class of so-called k-automatic sequences (see [1]).

4.1 Uniform Recurrence Criterion

For each $l \in \mathbb{N}$ define an equivalence relation on A: $b \sim_l c$ iff $h(\phi^l(b)) = h(\phi^l(b))$. We can easily continue this relation on A^*: $u \sim_l v$ iff $h(\phi^l(u)) = h(\phi^l(v))$. In fact, this means $|u| = |v|$ and $u(i) \sim_l v(i)$ for all i, $1 \leqslant i \leqslant |u|$.

Let B_m be the Bell number, i. e., the number of all possible equivalence relations on a finite set with exactly m elements, see [23]. As it follows from this article, we can estimate B_m in the following way.

Lemma 4.1. $2^m \leqslant B_m \leqslant 2^{Cm\log m}$ *for some constant C.*

Thus the number of all possible relations $\sim_l$ is not greater than $B_n = 2^{O(n \log n)}$. Moreover, the following lemma gives a simple description for the behavior of these relations as l tends to infinity.

Lemma 4.2. *If $\sim_r$ equals $\sim_s$, then $\sim_{r+p}$ equals $\sim_{s+p}$ for all p.*

Proof. Indeed, suppose $\sim_r$ equals $\sim_s$. Then $b \sim_{r+1} c$ iff $\phi(b) \sim_r \phi(c)$ iff $\phi(b) \sim_s \phi(c)$ iff $b \sim_{s+1} c$. So if $\sim_r$ equals $\sim_s$, then $\sim_{r+1}$ equals $\sim_{s+1}$, which implies the lemma statement. $\square$

This lemma means that the sequence $(\sim_l)_{l \in \mathbb{N}}$ turns out to be ultimately periodic with a period and a preperiod both not greater than B_n. Thus we obtain the following

Lemma 4.3. *For some $p, q \leqslant B_n$ we have for all i and all $t > p$ that $\sim_t$ equals $\sim_{t+iq}$.*

Now let us try to get a criterion which we could check in polynomial time. Notice that the situation is much more difficult than in the pure case because of a coding allowed. In particular, the analogue of Proposition 2.1 for non-pure case does not hold in general.

We move step by step to the appropriate version of the criterion reformulating it several times.

This proposition is quite obvious and follows directly from the definition of uniform recurrence since all $h(\phi^m(a))$ are the prefixes of $h(\phi^\infty(a))$.

Proposition 4.4. *A sequence $h(\phi^\infty(s))$ is uniformly recurrent iff for all m the word $h(\phi^m(s))$ occurs in $h(\phi^\infty(s))$ infinitely often with bounded distances.*

And now a bit more complicated version.

Proposition 4.5. *A sequence $h(\phi^\infty(s))$ is uniformly recurrent iff for all m the symbols that are $\sim_m$-equivalent to s occur in $\phi^\infty(s)$ infinitely often with bounded distances.*

Proof. $\Leftarrow$. If the distance between two consecutive occurrences in $\phi^\infty(s)$ of symbols that are $\sim_m$-equivalent to s is not greater than t, then the distance between two consecutive occurrences of $h(\phi^m(s))$ in $h(\phi^\infty(s))$ is not greater than tk^m.

$\Rightarrow$. Suppose $h(\phi^\infty(s))$ is uniformly recurrent. Let $y_m = 012\ldots(k^m-2)(k^m-1)01\ldots(k^m-1)0\ldots$ be a periodic sequence with a period k^m. Then by Proposition 2.3 a sequence $h(\phi^\infty(s)) \times y_m$ is uniformly recurrent, which means that the distances between consecutive k^m-aligned occurrences of $h(\phi^m(s))$ in $h(\phi^\infty(s))$ are bounded. It only remains to notice that if $h(\phi^\infty(s))[ik^m, (i+1)k^m - 1] = h(\phi^m(s))$, then $\phi^\infty(s)(i) \sim_m s$. $\square$

Let Y_m be the following statement: symbols that are $\sim_m$-equivalent to s occur in $\phi^\infty(s)$ infinitely often with bounded distances.

Suppose for some T that Y_T is true. This implies that $h(\phi^T(s))$ occurs in $h(\phi^\infty(s))$ with bounded distances. Therefore for all $m \leqslant T$ a word $h(\phi^m(s))$ occurs in $h(\phi^\infty(s))$ with bounded distances since $h(\phi^m(s))$ is a prefix of $h(\phi^T(s))$. Thus we do not need to check the statements Y_m for all m, but only for all $m \geqslant T$ for some T.

Furthermore, it follows from Lemma 4.3 that we are sufficient to check the only one such statement as in the following

Proposition 4.6. *For all $r \geqslant B_n$: a sequence $h(\phi^\infty(s))$ is uniformly recurrent iff the symbols that are $\sim_r$-equivalent to s occur in $\phi^\infty(s)$ infinitely often with bounded distances.*

And now the final version of our criterion.

Theorem 4.7. *For all $r \geqslant B_n$: a sequence $h(\phi^\infty(s))$ is uniformly recurrent iff there exists m such that for all $b \in A$ some symbol that is $\sim_r$-equivalent to s occurs in $\phi^m(b)$.*

Indeed, if the symbols of some set occur with bounded distances, then they occur on each k^m-aligned segment for some sufficiently large m.

Now we explain how to check a condition from Theorem 4.7 in polynomial time.

Corollary 4.8. *There exists polynomial-time algorithm checking whether given automatic sequence is uniformly recurrent.*

Proof. We need to show two things: first, how to choose some $r \geqslant B_n$ and to find in polynomial time the set of all symbols that are $\sim_r$-equivalent to s (and this is a complicated thing keeping in mind that B_n is exponential), and second, how to check whether for some m the symbols from this set for all $b \in A$ occur in $\phi^m(b)$.

Let us start from the second. Suppose we have found the set H of all the symbols that are $\sim_r$-equivalent to s. For $m \in \mathbb{N}$ let us denote by $P_m^{(b)}$ the set of all the symbols that occur in $\phi^m(b)$. Our aim is to check whether exists m such that for all b we have $P_m^{(b)} \cap H \neq \varnothing$. First of all, notice that if $\forall b\ P_m^{(b)} \cap H \neq \varnothing$, then $\forall b\ P_l^{(b)} \cap H \neq \varnothing$ for all $l \geqslant m$. Second, notice that the sequence of tuples of sets $((P_m^{(b)})_{b\in A})_{m=0}^\infty$ is ultimately periodic. Indeed, the sequence $(P_m^{(b)})_{m=0}^\infty$ is obviously ultimately periodic with both period and preperiod not greater than 2^n (recall that n is the size of the alphabet A). Thus the period of $((P_m^{(b)})_{b\in A})_{m=0}^\infty$ is not greater than the least common divisor of that for $(P_m^{(b)})_{m=0}^\infty$, $b \in A$, and the preperiod is not greater than the maximal that of $(P_m^{(b)})_{m=0}^\infty$. So the period is not greater than $(2^n)^n = 2^{n^2}$ and the preperiod is not greater than 2^n. Third, notice that there is a polynomial-time-procedure that given a graph corresponding to some morphism ψ (see Section 2 to recall what is the graph corresponding to a morphism) outputs a graph corresponding to morphism ψ^2. Thus after repeating this procedure n^2+1 times we obtain a graph by which we can easily find $(P^{(b)}_{2^{n^2}+2^n})_{b\in A}$, since $2^{n^2+1} > 2^{n^2} + 2^n$.

Similar arguments, even described with more details, are used in deciding our next problem. Here we present a polynomial-time algorithm that finds the set of all symbols that are $\sim_r$-equivalent to s for some $r \geqslant B_n$.

We recursively construct a series of graphs T_i. Let its common set of vertices be the set of all unordered pairs (b, c) such that $b, c \in A$ and $b \neq c$. Thus the number of vertices is $\frac{n(n-1)}{2}$. The set of all vertices connected with (b, c) in the graph T_i we denote by $V_i(b, c)$.

Define a graph T_0. Let $V_0(b, c)$ be the set $\{(\phi(b)(j), \phi(c)(j)) \mid j = 1, \dots, k, \phi(b)(j) \neq \phi(c)(j)\}$. In other words, $b \sim_{l+1} c$ if and only if $x \sim_l y$ for all $(x, y) \in V_0(b, c)$.

Thus $b \sim_2 c$ if and only if for all $(x, y) \in V_0(b, c)$ for all $(z, t) \in V_0(x, y)$ we have $z \sim_0 t$. For the graph T_1 let $V_1(b, c)$ be the set of all (x, y) such that there is a path of length 2 from (b, c) to (x, y) in T_0. The graph T_1 has the following property: $b \sim_2 c$ if and only if $x \sim_0 y$ for all $(x, y) \in V_1(b, c)$. And even more generally: $b \sim_{l+2} c$ if and only if $x \sim_l y$ for all $(x, y) \in V_1(b, c)$.

Now we can repeat operation made with T_0 to obtain T_1. Namely, in T_2 let $V_2(b, c)$ be the set of all (x, y) such that there is a path of length 2 from (b, c) to (x, y) in T_1. Then we obtain: $b \sim_{l+4} c$ if and only if $x \sim_l y$ for all $(x, y) \in V_2(b, c)$.

It follows from Lemma 4.1 that $\log_2 B_n \leqslant Cn \log n$. Thus after we repeat our procedure $r = [Cn \log n]$ times, we will obtain the graph T_l such that $b \sim_{2^r} c$ if and only if $x \sim_0 y$ for all $(x, y) \in V_2(b, c)$. Recall that $x \sim_0 y$ means $h(x) = h(y)$, so now we can easily compute the set of symbols that are $\sim_{2^r}$-equivalent to s. $\square$

4.2 Recurrence Criterion

Here we are discussing recurrence criterion for automatic sequences.

It is not difficult to see that all the arguments of Subsection 4.1 can be applied to the recurrence case with appropriate changes. The only note is that while proving analogue of Proposition 4.5 we should use the following statement instead of Proposition 2.3:

Proposition 4.9. *If x is recurrent and y is periodic, then $x \times y$ is recurrent.*

The proof of Proposition 4.9 is absolutely analogous to the proof of Proposition 2.3 and is left to the reader.

Now we can formulate the recurrence criterion for morphic sequences, analogously to Proposition 4.6:

Proposition 4.10. *For all $r \geqslant B_n$: a sequence $h(\phi^\infty(s))$ is recurrent iff the symbols that are $\sim_r$-equivalent to s occur in $\phi^\infty(s)$ infinitely many times.*

The symbols that are $\sim_r$-equivalent to s occur in $\phi^\infty(s)$ infinitely often if and only if some symbol that is $\sim_r$-equivalent to s occurs infinitely often. It is not difficult to see that for each $a \in A$ we can check in polynomial time analyzing graph G_ϕ, whether a occurs in $\phi^\infty(s)$ infinitely many times. Thus we obtain the following

Corollary 4.11. *There exists polynomial-time algorithm checking whether given automatic sequence is recurrent.*

5 Factor Complexity

Factor complexity is a natural combinatorial characteristic of finite or infinite words. *Factor complexity* of $x \in A^\infty$ is a function $p_x \colon \mathbb{N} \to \mathbb{N}$ where $p_x(n)$ is the number of all n-length factors occurring in x. For a survey on factor complexity see, e. g., [9]. Denote by $F(x)$ the set of all factors of a sequence x, by $F_n(x)$ the set of all n-length factors of a sequence x.

A result from Pansiot [16] states that the factor complexity of arbitrary pure morphic sequence adopts one of the five following asymptotic behaviors: $O(1)$, $\Theta(n)$, $\Theta(n \log \log n)$, $\Theta(n \log n)$ or $\Theta(n^2)$. In fact, factor complexity of ultimately periodic sequences is $O(1)$, while for non-periodic sequences it is always $\Omega(n)$ according to [14].

However, for uniformly recurrent morphic sequences the situation is much easier.

Theorem 5.1. *If x is an uniformly recurrent morphic sequence, then $p_x(n) = O(n)$.*

The proof of the theorem is in following several lemmas. Probably, the keynote lemma containing a funny trick is Lemma 5.7.

Lemma 5.2 (essentially, from Pansiot [16]). *If x is a pure morphic sequence generated by a primitive morphism, then $p_x(n) = O(n)$.*

Lemma 5.3 (Cassaigne, Nicolas [3]). *Let A, B be two alphabets, let $f\colon A^* \to B^*$ be a non-erasing morphism, and let M be the maximal length of $f(a)$ over all $a \in A$. Then $p_{f(x)}(n) \leqslant Mp_x(n)$ for every infinite word $x \in A^{\mathbb{N}}$ and $n \in \mathbb{N}$.*

Proof. For each $v \in F_n(f(x))$ there exists a representation $f(t) = uvw$, where $t \in F_n(x)$, $u, w \in B^*$, and u is chosen with the minimal possible length; clearly, $|u| \leqslant M$. Thus the cardinality of $F_n(f(x))$ is not greater than the number of all pairs $(|u|, t)$ from such representations, i. e., not greater than $M \cdot |F_n(x)| = Mp_x(n)$. □

Lemma 5.4 (Pansiot [16]). *Let A be an alphabet, $s \in A$, and let $\phi\colon A^* \to A^*$ be a morphism prolongable on s. Assume that the set of all ϕ-bounded factors of $\phi^\infty(s)$ is finite. Then $\phi^\infty(s)$ can be written as the image under a non-erasing morphism of a pure morphic sequence generated by a growing morphism.*

Lemma 5.5. *For every two infinite words x and y, if x is uniformly recurrent and $F(y) \subseteq F(x)$, then $F(y) = F(x)$.*

Lemma 5.5 is a well-known minimality property of uniformly recurrent sequences, e. g., see [12].

Lemma 5.6. *Let B be an alphabet and let $\phi\colon B^* \to B^*$ be a growing morphism. There exist a natural n and a letter $t \in B$ such that ϕ^n is prolongable on t.*

Proof. Let b be an element of B. Since B is finite, there exist i, j with $i < j$ such that $\phi^i(b)$ and $\phi^j(b)$ start with the same letter, say t. Hence $\phi^{j-i}(t)$ begins with t. Since ϕ is growing, ϕ^{j-i} is growing too. Thus ϕ^{j-i} is prolongable on t. □

Lemma 5.7. *For every pure morphic sequence x generated by a growing morphism, there exists a pure morphic sequence y generated by a primitive morphism such that $F(y) \subseteq F(x)$.*

Proof. Suppose $x = \phi^\infty(s)$ where ϕ is growing. Let B be a strongly connected component in the incidence graph G_ϕ with no outgoing edges. Then ϕ restricted to B induces a growing irreducible morphism from B^* to B^*. According to Lemma 5.6, there exist $t \in B$ and n such that ϕ^n is prolongable on t. If ϕ^n is primitive, then we are done and $(\phi^n)^\infty(t)$ is a suitable choice for y, since t and therefore all its morphic powers occur in x.

Suppose ϕ^n is not primitive. It means that B is a proper subgraph of G_ϕ, because otherwise ϕ^n is both prolongable and irreducible, and thus primitive. Now repeat the procedure: consider G_{ϕ^n} (which is B in fact), find some its strongly connected component with no outgoing edges, consider an appropriate power of ϕ^n which is prolongable on some letter, and so on.

Thus on each step of this argument we either find a suitable y, or decrease the size of the current subgraph. So we are done by induction. □

Now we are ready to prove the main theorem of this section. Recall that it establishes the factor complexity of arbitrary uniformly recurrent morphic sequence to be at most linear.

Proof of Theorem 5.1. Suppose $x = h(\phi^\infty(s))$ is an uniformly recurrent morphic sequence with $\phi\colon A^* \to A^*$, $h\colon A \to B$. There are two possibilities.

1) There exist infinitely many ϕ-bounded factors in $\phi^\infty(s)$. Then by Proposition 3.2 there exists a non-empty $w \in A^*$ such that w^n occurs in x for each n. Therefore $(h(w))^n$ occurs in x for each n, and hence x is periodic, which means that its complexity is $O(1)$.

2) There are only finitely many ϕ-bounded factors in $\phi^\infty(s)$. Then by Lemma 5.4 $\phi^\infty(s)$ can be represented as $\phi^\infty(s) = g(\psi^\infty(t))$ for some $\psi\colon C^* \to C^*$, $g\colon C^* \to D^*$ with ψ growing and g non-erasing. By Lemma 5.7 there exists pure morphic sequence y generated by a primitive morphism such that $F(y) \subseteq F(\psi^\infty(t))$. Hence $F(h(g(y))) \subseteq F(x)$, but x is uniformly recurrent, therefore by Lemma 5.5 we have $F(h(g(y))) = F(x)$. Thus for some constant M

$$p_x(n) = p_{h(g(y))}(n) \leqslant M p_y(n) = O(n),$$

where the second inequality holds by Lemma 5.3 and the last equality holds by Lemma 5.2. □

Interestingly, almost nothing is known about factor complexity of arbitrary morphic sequences. Probably, the only progress is done in [4]. It is shown there that for morphic sequences there exist at least infinitely many complexity classes of the form $\Theta(n^{1+\frac{1}{k}})$ for $k \in \mathbb{N}$. However, recently two conjectures were made [7].

Conjecture 5.8 (Rostislav Devyatov). *The factor complexity of arbitrary morphic sequence is either of the form $\Theta(n^{1+\frac{1}{k}})$ for some $k \in \mathbb{N}$, or of the form $O(n \log n)$.*

Conjecture 5.8 seems to be proved by its author but the proof is extremely difficult and has to be additionally rechecked several times. That is why it is formulated as a conjecture here.

To the contrary, the status of the following conjecture is much weaker, and no proof is known so far.

Conjecture 5.9 (Rostislav Devyatov). *The factor complexity of arbitrary morphic sequence adopts one of the following asymptotic behaviors: $O(1)$, $\Theta(n)$, $\Theta(n \log\log n)$, $\Theta(n \log n)$, $\Theta(n^2)$ or $\Theta(n^{1+\frac{1}{k}})$ for some $k \in \mathbb{N}$.*

6 Arbitrary Morphic Sequences

It is not still known whether the problem of determining uniform recurrence of arbitrary morphic sequence is decidable, though we believe that it is true.

Conjecture 6.1. *It is decidable given arbitrary morphic sequence, whether this sequence is uniformly recurrent or not.*

Proposition 6.2 given below somehow supports Conjecture 6.1 (but in fact even does not follow from it!).

A very natural characteristic of uniformly recurrent sequence is uniform recurrence regulator. An *uniform recurrence regulator* of an uniformly recurrent sequence x is a function $f\colon \mathbb{N} \to \mathbb{N}$ such that every n-length factor u of x occurs in each $f(n)$-length factor of x, and $f(n)$ is chosen to be minimal satisfying this condition. So an uniform recurrence regulator somehow regulates how (uniformly) recurrent a sequence is.

Proposition 6.2. *If x is both morphic and uniformly recurrent, then its uniform recurrence regulator is computable.*

Proof. First, notice that the set of factors of morphic sequence is decidable. In other words, there exists an algorithm that given a morphic sequence and a word, says whether this word occurs in the sequence.

Second, if an uniformly recurrent sequence is computable and its set of factors is decidable, then the uniform recurrence regulator of this sequence is computable. Indeed, suppose we want to check whether $l \geqslant f(n)$. For that we find all n-length factors and all l-length factors, we can do it due to decidability of the set of factors. Then we check whether each of l-length factors contains all n-length factors. If so, then $l \geqslant f(n)$. Thus to find precise value of $f(n)$, we can check all natural numbers starting from n until some of them works. $\square$

Proposition 6.2 can easily be made uniform: there exists an algorithm that given a morphic sequence computes its uniform recurrence regulator whenever this sequence is uniformly recurrent.

However, Proposition 6.2 does not imply the decidability of uniform recurrence for morphic sequences. In fact, this decidability also does not imply Proposition 6.2. By the way, Proposition 6.2 allows us to hope that this algorithm of decidability exists.

Monadic theory of morphic sequence is decidable, e. g., see [2]. In fact, it also follows from the result from [6] that finite transduction of a morphic sequence is morphic, see also [1].

The property of recurrence for $x \in \Sigma^{\mathbb{N}}$ can be written as

$$\text{(for each prefix } u \text{ of } x) \text{ (there are infinitely many occurrences of } u \text{ in } x).$$

The property "there are infinitely many occurrences of u in x" for morphic x can be algorithmically checked, since it can be written in monadic language. Thus the problem of determining recurrence for morphic sequences is in the class Π^0_1 of Kleene hierarchy.

It is not difficult to see that the problem of determining uniform recurrence for morphic sequences is in Π^0_2, since it can be written as

$$\text{(for each prefix } u \text{ of } x) \text{ (there exists } l) \text{ such that } (u \text{ occurs on each } l\text{-length segment of } x),$$

where the last property can be algorithmically checked for morphic sequences again by monadic logic reasons.

However, it turns out that this problem lies in Π^0_1. Indeed, by Lemma 5.7 and all the proof of Theorem 5.1, for a morphic x we can find morphic y such that $x \in \mathcal{UR}$ whenever $F(x) = F(y)$, and thus uniform recurrence of x can be expressed by "$\forall\ \exists$"-formula.

Finally, notice that Theorem 7.5.1 from [1] allows us to represent an arbitrary morphic sequence as $h(\phi^\infty(s))$ with ϕ non-erasing. So it is sufficient to solve our main problem for morphic sequences generated by non-erasing morphisms.

7 Conclusion

We have described two polynomial-time algorithms, but without any precise bound for their working time. Of course, it can be done after deep analyzing of all the previous, but is probably not so interesting.

The problem of finding an effective periodicity criterion in the case of arbitrary morphic sequences is also of great interest, as well as criteria for variations with periodicity and uniform recurrence: ultimate periodicity, generalized uniform recurrence (called *generalized almost periodicity* in [19]), ultimate uniform recurrence, etc. If one notion is a particular case of another, it does not mean that corresponding criterion for the first case is more difficult (or less difficult) than for the second.

Of course, to continue investigations about factor complexity is also the problem of great interest. In particular, one can try to investigate factor complexity for morphic sequences of some special types.

Acknowledgements

J. Cassaigne, R. Devyatov, A. Frid, N. Rampersad, M. Raskin, K. Saari are kindly acknowledged for fruitful discussions, and An. Muchnik and A. Semenov are especially acknowledged for their permanent support.

References

[1] J.-P. Allouche, J. Shallit. *Automatic Sequences.* Cambridge University Press, 2003.

[2] O. Carton, W. Thomas. *The Monadic Theory of Morphic Infinite Words and Generalizations.* Information and Computation 176, N 1, pp. 51–65, 2002.

[3] J. Cassaigne, F. Nicolas. *Quelques propriétés des mots substitutifs.* Bulletin of the Belgian Mathematical Society—Simon Stevin 10, N 5, pp. 661–677, 2003.

[4] J. Cassaigne, F. Nicolas. In preparation.

[5] A. Cobham. *Uniform tag sequences.* Mathematical Systems Theory 6, N 1–2, pp. 164–192, 1972.

[6] F. M. Dekking. *Iteration of maps by an automaton.* Discrete Mathematics 126, N 1–3, pp. 81–86, 1994.

[7] R. Devyatov, personal communication, 2007.

[8] A. Ehrenfeucht, G. Rozenberg. *Repetition of Subwords in DOL Languages.* Information and Control 59, N 1–3, pp. 13–35, 1983.

[9] S. Ferenczi. *Complexity of sequences and dynamical systems.* Discrete Mathematics 206, N 1, pp. 145–154, 1999.

[10] T. Harju, M. Linna. *On the periodicity of morphisms on free monoids.* RAIRO Informatique Théorique et Applications 20, N 1, pp. 47–54, 1986.

[11] Kolmogorov Seminar: `http://lpcs.math.msu.su/kolmogorovseminar/eng/`.

[12] M. Lothaire. *Combinatorics on Words.* Cambridge University Press, 2nd edition, 1997.

[13] A. Maes. *More on morphisms and almost-periodicity.* Theoretical Computer Science 231, N 2, pp. 205–215, 2000.

[14] M. Morse, G. A. Hedlund. *Symbolic dynamics.* American Journal of Mathematics 60, N 4, pp. 815–866, 1938.

[15] An. Muchnik, A. Semenov, M. Ushakov. *Almost periodic sequences.* Theoretical Computer Science 304, N 1–3, pp. 1–33, 2003.

[16] J.-J. Pansiot. *Complexité des Facteurs des Mots Infinis Engendrés par Morphimes Itérés.* Proceedings of ICALP'84, Lecture Notes in Computer Science 172, pp. 380–389, 1984.

[17] J.-J. Pansiot. *Decidability of periodicity for infinite words.* RAIRO Informatique Théorique et Applications 20, N 1, pp. 43–46, 1986.

[18] Yu. L. Pritykin. *Finite-Automaton Transformations of Strictly Almost-Periodic Sequences.* Mathematical Notes 80, N 5, pp. 710–714, 2006 (translated from Russian, Matematicheskie Zametki 80, N 5, pp. 751–756, 2006). Preprint on `http://arXiv.org/abs/cs.DM/0605026`.

[19] Yu. Pritykin. *Almost Periodicity, Finite Automata Mappings and Related Effectiveness Issues.* Proceedings of WoWA, St. Petersburg, Russia (satellite to CSR 2006). To appear in "Izvestia VUZov. Mathematics". Preprint on `http://arXiv.org/abs/cs.DM/0607009`.

[20] Yu. Pritykin. *On Almost Periodicity Criteria for Morphic Sequences in Some Particular Cases.* Proceedings of 11th International Conference on Developments in Language Theory (DLT 2007), Turku, Finland, Lecture Notes in Computer Science 4588 (Springer), pp. 361–370. Preprint on `http://arXiv.org/abs/0704.0218`.

[21] A. L. Semenov. *On certain extensions of the arithmetic of addition of natural numbers.* Mathematics of USSR–Izvestia 15, N 2, pp. 401–418, 1980.

[22] A. L. Semenov. *Logical theories of one-place functions on the set of natural numbers.* Mathematics of USSR–Izvestia 22, N 3, pp. 587–618, 1983.

[23] Eric W. Weisstein. *Bell Number.* From MathWorld—A Wolfram Web Resource. `http://mathworld.wolfram.com/BellNumber.html`