International Journal of Foundations of Computer Science
Vol. 20, No. 5 (2009) 919–940

ON UNIFORMLY RECURRENT MORPHIC SEQUENCES

FRANCOIS NICOLAS

Department of Computer Science
P. O. Box 68
00014 University of Helsinki
nicolas@cs.helsinki.fi

YURI PRITYKIN

Moscow State University,
Department of Mechanics and Mathematics,
Mathematical Logic and Theory of Algorithms Division
Leninskie gory, Moscow, 119991 Russia
yura@mccme.ru

Received 13 January 2008
Accepted 22 November 2008
Communicated by Juhani Karhumaki

A pure morphic sequence is a right-infinite, symbolic sequence obtained by iterating a letter-to-word substitution. For instance, the Fibonacci sequence and the Thue–Morse sequence, which play an important role in theoretical computer science, are pure morphic. Define a coding as a letter-to-letter substitution. The image of a pure morphic sequence under a coding is called a morphic sequence.

A sequence x is called uniformly recurrent if for each finite subword u of x there exists an integer l such that u occurs in every l-length subword of x.

The paper mainly focuses on the problem of deciding whether a given morphic sequence is uniformly recurrent. Although the status of the problem remains open, we show some evidence for its decidability: in particular, we prove that it can be solved in polynomial time on pure morphic sequences and on automatic sequences.

In addition, we prove that the complexity of every uniformly recurrent, morphic sequence has at most linear growth: here, complexity is understood as the function that maps each positive integer n to the number of distinct n-length subwords occurring in the sequence.

Keywords: uniformly recurrent sequence; morphic sequence; automatic sequence; subword complexity; polynomial-time algorithm.

1991 Mathematics Subject Classification: 68R15, 03D40, 08A50, 68Q25

1. Introduction

Many problems of decidability in combinatorics on words are of great interest and difficulty. Many difficult problems are connected with free monoid morphisms, also known as substitutions. A good survey on such problems is [12]. A famous example is Post Correspondence Problem (e.g., see [33], Section 5.2, and [12]) which is known to be undecidable.

In this paper we deal with two classes of sequences well-known in combinatorics on words: morphic and uniformly recurrent. Let us extend the domain of substitutions to infinite sequences in the obvious way: the image of an infinite sequence under a substitution is obtained by simultaneously applying the substitution to all symbols of the sequence. A non-trivial, infinite, fixed point of a substitution is called a pure morphic sequence. The image of a pure morphic sequence under a substitution is called a (general) morphic sequence. Such sequences can be effectively described, i.e., by a finite amount of information. Morphic sequences generalize the well-known automatic sequences, i.e., those generated by morphisms in which images of all letters have equal lengths [2]. They appear in symbolic dynamics, number theory, geometry (e.g., see [28, 29, 2, 16, 17]). Another class of interest is the class of uniformly recurrent sequences[a], i.e., sequences in which every factor occurs infinitely often with bounded gaps (the gap size being an arbitrary function of the factor) between its consecutive occurrences. These sequences first appeared in symbolic dynamics, but then turned out to be interesting in connection with combinatorics on words, mathematical logic (e.g., see [2, 20, 30, 31, 26]).

Here we mainly study connections between these two classes of sequences. We discuss the existence of an algorithm that given a morphic sequence (which, recall, can be finitely described) determines whether it is uniformly recurrent. Though the problem in general still remains open, we are successful in two particular cases in which we find polynomial-time algorithms. The first one is limited to pure morphic sequences. (This problem was reported as open in [2], Section 10.12, Problem 5.) The second one considers automatic sequences. Furthermore, besides these criteria, we prove that uniformly recurrent morphic sequences have linearly increasing subword complexity: recall that the subword (=factor) complexity of a sequence is the function mapping each integer $n \geq 0$ to the number of n-length words that occur in the sequence.

The paper is organized as follows. In Section 2 we give all formal definitions, as well as some preliminary results. In Section 3 we deal with pure morphic sequences and formulate two versions of the uniform recurrence criterion in Theorems 4 and 7. In Section 4 we study the case of automatic sequences. Uniform recurrence criterion is discussed in Subsection 4.1. The criterion itself is given in Theorem 18, while in Theorem 19 we explain how to check it in polynomial time. The (non-uniform) recurrence criterion as a related problem is discussed in Subsection 4.2. Factor complexity is studied in Section 5. Theorem 24 says that factor complexity of a uniformly recurrent morphic sequence is at most linear. In Section 6 we discuss the problem of determining uniform recurrence for morphic sequences in general. In Proposition 33 we give the result supporting the conjecture of decidability. In

[a]Also known as *almost periodic* or *minimal*. They were called *strongly* or *strictly almost periodic* in [20, 27].

Proposition 35 we prove $\mathbf{0}'$-decidability of the main problem.

Some attempts to touch these topics were already done. In [5] A. Cobham gives a criterion for an automatic sequence to be uniformly recurrent. However even if his criterion gives some effective procedure solving the problem, this procedure does not look fast. In [18] A. Maes deals with pure morphic sequences and describes a decision procedure determining their membership in the class of "almost-periodic" sequences. Maes calls a sequence x almost-periodic[b] if for every factor u of x, either u occurs in x at most finitely many times, or u occurs in x infinitely often with bounded gaps. Hence, if x is uniformly recurrent or if some suffix of x is uniformly recurrent then x is Maes-almost-periodic. The converse is false in general (see [27, 26]). His algorithm is not polynomial-time, but there are some parallels between his considerations and our Section 3 (though these parallels are sometimes hard to formulate). The problem of determining ultimate periodicity for pure morphic sequences was solved independently in [13] and [23]. Some further remarks can be found in Section 7.

Preliminary versions of some of the results from this paper originally appeared in [25].

2. Preliminaries

Denote the set of natural numbers $\{0, 1, 2, \dots\}$ by $\mathbb{N}$, and the standard alphabet $\{0, 1, \dots, n-1\}$ by Σ_n. Let A be a finite alphabet. Sequences over A are mappings $x\colon \mathbb{N} \to A$, and the set of sequences over A is denoted 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. Word concatenation is denoted multiplicatively. For $i \le j$, denote by $x[i, j]$ a *subword* (or a *factor*) $x(i)x(i+1)\dots x(j)$ of a sequence x. A segment $[i, j]$ is an occurrence of a word $u \in A^*$ in a sequence x if $x[i, j] = u$. A word of the form $x[0, i]$ for some i is called a *prefix* of x. Denote by $|u|$ the length of a word u.

A sequence x is *periodic* if for some T we have $x(i) = x(i+T)$ for each $i \in \mathbb{N}$. According to a usual common agreement, both T and the word $x(0)\dots x(T-1)$ are called a *period* of x. A sequence which is a concatenation of a finite word and a periodic sequence is called *ultimately periodic*. This finite word, as well as its length, are called a *preperiod* of the sequence. The paper focuses on the following two natural extensions of the class of periodic sequences.

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

A sequence x is called *uniformly recurrent* 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 hence u occurs in x infinitely many times). Clearly, to show uniform recurrence of a sequence, it is sufficient to check the mentioned condition only for all prefixes

[b]Now called *generalized almost periodic*, e.g., see [26].

but not for all factors (and even for some increasing sequence of prefixes only).

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 $u, v \in A^*$. Obviously, a morphism is determined by its values on single-letter words. A morphism is *non-erasing* if $|\phi(a)| \geq 1$ 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 a *coding*.

In what follows, morphisms are mainly codings or endomorphisms (i.e., morphisms from A^* to itself for some alphabet A). For $x \in A^{\mathbb{N}}$ denote

$$\phi(x) = \phi(x(0))\phi(x(1))\phi(x(2))\ldots$$

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. Consider $\phi(0) = 1$ and $\phi(1) = 0$ as a counterexample.

A word $w \in A^*$ is called ϕ*-bounded* if the sequence $(w, \phi(w), \phi^2(w), \phi^3(w), \ldots)$ is ultimately 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. Moreover, a word is ϕ-bounded iff it consists of ϕ-bounded symbols only. A word $w \in A^*$ is ϕ*-eventually-erased* if $\phi^n(w) = \Lambda$ for some n. Analogously, a word is ϕ-eventually-erased iff it consists of ϕ-eventually-erased symbols only.

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.

A morphism $\phi\colon A^* \to A^*$ is called *growing* if every letter in A is ϕ-growing, i.e., $A = I_\phi$.

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 $\phi^\infty(s) = \lim_{n\to\infty} \phi^n(s) = su\phi(u)\phi^2(u)\phi^3(u)\ldots$ is correctly defined. If u is not ϕ-eventually-erased, then $\phi^\infty(s)$ is infinite. In this case we say that ϕ is *prolongable* on s. In other words, ϕ is *prolongable* on s if $\phi(s)$ starts with s, and s is ϕ-growing. Sequences of the form $h(\phi^\infty(s))$ for a coding $h\colon A \to B$ and a morphism ϕ prolongable on s are called *morphic*, of the form $\phi^\infty(s)$ are called *pure morphic*.

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. Sequences that are k-automatic for some k, are called simply *automatic* (this class was introduced in [5] under the name of uniform tag sequences; see also [2]).

Note that there exist uniformly recurrent sequences that are not morphic (in fact, the set of uniformly recurrent sequences has cardinality of the continuum (e.g., see [15, 20]), while the set of morphic sequences is obviously countable), as well as there exist morphic sequences that are not uniformly recurrent (an example can be found below).

Our main goal is to study the decidability of the following problem:

Input: Two finite alphabets A and B, a letter $s \in A$, a morphism $\phi\colon A^* \to A^*$ prolongable on s, and a coding $h\colon A \to B$.
Question: Is the morphic sequence $h(\phi^\infty(s))$ uniformly recurrent?

Remark that recognizing the set of instances of the problem requires checking whether s is ϕ-growing: this can be achieved in polynomial time according to Lemma 6 below.

Before we start discussing our main problem mentioned above, let us show that it is not difficult to formulate a criterion of recurrence for pure morphic sequences.

Proposition 1. *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. □

The situation is not that easy in the case of uniform recurrence. First of all, observe the following

Proposition 2. *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. □

For a morphism $\phi\colon \Sigma_n^* \to \Sigma_n^*$, we define *an incidence matrix* M_ϕ, such that $(M_\phi)_{ij}$ is the number of occurrences of the symbol i into $\phi(j)$. One can easily check that for each $l \in \mathbb{N}$ one has $M_\phi^l = M_{\phi^l}$.

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.

For a morphism $\phi\colon A^* \to A^*$, we define a directed *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 $\phi(b)$. Given a morphism ϕ, many properties of ϕ can be computed from its incidence graph G_ϕ. However, G_ϕ does not contain information about the number of occurrences of i into $\phi(j)$, that is, G_ϕ contains less information about the morphism ϕ than M_ϕ.

For $\phi^\infty(s)$ it can easily be found using G_ϕ which symbols from A actually 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 $\phi^\infty(s)$.

A morphism is irreducible if and only if its graph of incidence is strongly connected, i.e., there exists a directed path between every two vertices. For prolongable morphisms this is also a criterion of primitiveness. Thus the following proposition gives an effective polynomial-time criterion in the case of growing morphisms.

Proposition 3. *Let A be an alphabet, let $s \in A$, and let $\phi\colon A^* \to A^*$ be a growing morphism prolongable on s. Then, $\phi^\infty(s)$ is uniformly recurrent if and only if ϕ is primitive.*

Proof. Assume that ϕ is primitive. Let n be a positive integer such that s occurs in $\phi^n(a)$ for every $a \in A$. Let l denote the maximum length of $\phi^n(a)$ over all $a \in A$. Every factor of $\phi^\infty(s)$ with length $2l$ contains at least one occurrence of s. It now follows from Proposition 2 that $\phi^\infty(s)$ is uniformly recurrent.

Conversely, assume that ϕ is not primitive. There exists $b \in A$ such that for every $n \geq 0$ the word $\phi^n(b)$ does not contain s. For every $n \geq 0$, $\phi^n(b)$ is a factor of $\phi^\infty(s)$. Since $|\phi^n(b)| \to \infty$ as $n \to \infty$, $\phi^\infty(s)$ is not uniformly recurrent. □

Thus the famous Fibonacci sequence

$$f = 0100101001001010010100\ldots$$

generated by the morphism $0 \to 01$, $1 \to 0$, as well as the Thue–Morse sequence

$$t = 0110100110010110100101100\ldots$$

generated by the morphism $0 \to 01$, $1 \to 10$, are uniformly recurrent, as generated by primitive morphisms.

However when we generalize this case even to non-erasing morphisms, it is not sufficient to consider only the incidence graph or even the incidence matrix, 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: $0 \to 01$, $1 \to 210$, $2 \to 2$. They generate sequences

$$\phi_1^\infty(0) = 011201202011202012011\ldots \text{ and } \phi_2^\infty(0) = 0121022100122210010121\ldots$$

These two morphisms have identical incidence 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 of the form $222\ldots22$, so $\phi_2^\infty(0)$ is not uniformly recurrent (but recurrent). 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 $\phi_1^\infty(0)$ is uniformly recurrent by Proposition 2. See Theorem 4 for a general criterion of uniform recurrence in the case of pure morphic sequences.

3. Pure Morphic Sequences

Here we consider morphic sequences of the form $\phi^\infty(s)$. In Theorem 4 we present a compact criterion of uniform recurrence. Theorem 7 gives more effective version that can be checked algorithmically in polynomial time.

Theorem 4. *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 $\geq l$. 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 $KM + K$. 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 in w.

We have thus shown that s occurs in every factor of $\phi^\infty(s)$ with length $KM+K$, and thus $\phi^\infty(s)$ is uniformly recurrent according to Proposition 2. $\square$

Now we explain how to get a polynomial-time criterion. First in Proposition 5 we give different reformulations of Property 2) from Theorem 4. Then we reformulate the uniform recurrence criterion in such a way that it can easily be checked in polynomial time.

Proposition 5 (Ehrenfeucht, Rozenberg [9]) *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) $\phi^n(a) = uav$ *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 plan. 3) ⇒ 1) is straightforward. 1) ⇒ 2) was proved in [9]. 2) ⇒ 3) is easy. □

For a morphism $\phi\colon A^* \to A^*$, recall from Section 2 what the sets I_ϕ, B_ϕ, E_ϕ are. In fact, we can effectively determine them given ϕ.

Lemma 6. *Given an alphabet A and a morphism $\phi\colon A^* \to A^*$, the sets I_ϕ, B_ϕ, and E_ϕ are computable in* $\mathrm{poly}(n, k)$*-time where $n = |A|$ and* $k = \max_{b \in A} |\phi(b)|$.

Proof. First, consider an equivalence relation "≡" 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 "≡". An edge goes from C to D in H_ϕ if $\exists a \in C\ \exists b \in D$ such that $\phi(a)$ contains b. Define for each vertex C in H_ϕ the number $\kappa_C = \max\{$the number of occurrences of symbols from C in $\phi(a) : a \in C\}$. Define $S_i = \{D \text{ in } H_\phi : \max\{\kappa_C : C \text{ can be reached from } D\} = i\}$. Obviously, H_ϕ, all κ_C, and S_i can be computed in polynomial time.

It is not difficult to see that $\forall i \geq 2\ \forall C \in S_i\ \forall a \in C\ |\phi^m(a)|$ grows exponentially as $m \to \infty$, and thus $a \in I_\phi$. Also, every element of S_0 is a singleton included in E_ϕ.

Note that for every strongly connected component C of G_ϕ, $\kappa_C = 1$ if and only if the subgraph of G_ϕ induced by C is a directed cycle.

Now consider the subgraph of H_ϕ induced by S_1. Clearly, $\forall C \in S_1\ \forall a \in C\ a \notin E_\phi$. Let $S_1 = U \cup V$ where $U = \{C \in S_1 : \kappa_C = 0\}$, $V = \{C \in S_1 : \kappa_C = 1\}$. Further, let $V = X \cup Y$ where $X = \{C \in V : \text{some other } D \in V \text{ can be reached from } C\}$, $Y = V \setminus X$. It is not difficult to see that $\bigcup_{C \in X} C \subseteq I_\phi$, $\bigcup_{C \in Y} C \subseteq B_\phi$. Further, notice that every $C \in U$ is a singleton. If some $D \in X$ can be reached from $\{a\} \in U$, then $a \in I_\phi$, otherwise $a \in B_\phi$.

Obviously, everything here can be checked in polynomial time. □

The presented proof of Lemma 6 is essentially taken from [8] by simplifying a general argument from there.

Let $\phi\colon A^* \to A^*$ be a morphism. Recall that 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 out. 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 a representation $\phi(b) = vcu$ where $u \in B_\phi^*$, $c \in I_\phi$, and write u on the edge from b to c.)

Now we formulate a constructive version of the criterion given in Theorem 4.

Theorem 7. *Let A be an alphabet, $s \in A$, and let $\phi\colon A^* \to A^*$ be a morphism prolongable on s. 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, a ϕ-eventually-erased word is written.

Recall that we assume all the symbols from A appear in $\phi^\infty(s)$.

Proof. Property 1) of this theorem is equivalent to Property 1) of Theorem 4. Proposition 5 explains why the same is true with Properties 2). □

In fact, the equivalence between Property 2) from Theorem 4 and Property 2) from Theorem 7 was first noticed and proved in [18] for the case of non-erasing morphism ϕ.

Let us consider the examples with ϕ_1 and ϕ_2 concluding Section 2. For $\phi \in \{\phi_1, \phi_2\}$, $I_\phi = \{0, 1\}$ and $B_\phi = \{2\}$. 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 8. *There exists a* poly(n, k)*-algorithm that decides whether $\phi^\infty(s)$ is uniformly recurrent.*

Proof. Conditions from Theorem 7 can easily be checked in polynomial time. □

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

Corollary 9. *For a morphism $\phi\colon \{0,1\}^* \to \{0,1\}^*$ 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 0 *and no* 1*;*
2) $\phi(1)$ contains 0*;*
3) $\phi(1) = \Lambda$;
4) $\phi(1) = 1$ and $\phi(0) = 0u0$ for some word u.

The example with ϕ such that $\phi(0) = 0010$, $\phi(1) = 1$ (the Chacon morphism, see [10, 28]), non-trivially illustrates Case 4 of Corollary 9: ϕ is not primitive, $\phi^\infty(0)$ is not ultimately periodic, and $\phi^\infty(0)$ is uniformly recurrent.

4. Uniform Morphisms

Now we deal with automatic sequences. In Subsection 4.1 we present a polynomial-time criterion of uniform recurrence in this case. Final version of the criterion is in Theorem 18. Theorem 19 is to explain why the criterion from Theorem 18 is polynomial-time. In Subsection 4.2 we are discussing a (non-uniform) recurrence criterion as a related problem.

Suppose we have two alphabets A and B, a morphism $\phi\colon A^* \to A^*$, a coding $h\colon A \to B$, and $s \in A$, such that $|A| = n$, $|B| \leq n$, $\forall b \in A$ $|\phi(b)| = k$, and $\phi(s)$ starts with s.

4.1. *Uniform recurrence criterion*

Here we are interested in whether $h(\phi^\infty(s))$ is uniformly recurrent.

For each $l \in \mathbb{N}$ define an equivalence relation on A: $b \sim_l c$ iff $h(\phi^l(b)) = h(\phi^l(c))$. We can easily generalize this relation to 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, $1 \leq i \leq |u|$.

Let $R = R(h, \phi)$ be the number of all possible relations $\sim_l$.

Lemma 10. $R \leq 2^{n^2}$.

Proof. The total number of binary relations over a set with n elements is not greater than 2^{n^2}. □

Some properties of $(\sim_l)_{l\in\mathbb{N}}$ are given below.

Lemma 11. *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. □

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

Lemma 12. *For some $p, q \in \mathbb{N}$ with $p + q = R$ we have for all i and all $t \geq p$ that $\sim_t$ equals $\sim_{t+iq}$.*

The following question has independent interest and is not directly connected with our considerations, while its partial solutions might help in various investigations of automatic sequences.

Question 13. *What are non-trivial lower and upper bounds for $R = R(h, \phi)$, depending on ϕ and h? In particular, may R be exponentially large in terms of n? And more generally, describe the behavior of $(\sim_l)_{l\in\mathbb{N}}$, depending on ϕ and h.*

Let us remark that $R \leq B_n$ where B_n is the n-th Bell number, i.e., the number of all possible equivalence relations on a finite set with n elements. It is known that $B_n = 2^{O(n\log n)}$, but this bound is not necessary for our purpose (Lemma 10 is sufficient).

The following proposition is also useful. It was first proved in [5]. 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 14. *If x is uniformly recurrent and y is periodic, then $x \times y$ is uniformly recurrent.*

Proof. Let us prove the proposition for y of the form $y = 012\dots(m-1)012\dots(m-1)01\dots$ over Σ_m. Then the proposition in full generality would follow by applying a coding.

We say that u occurs in x in position i modulo m, where $0 \leq i \leq m-1$, if there exists $q \in \mathbb{N}$ such that $u = x[mq+i, mq+i+|u|-1]$. Our aim is to prove that if u occurs in x in position k modulo m, then this happens infinitely many times with bounded distances.

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

Thus u occurs in position k modulo m in each occurrence of w in x. But x is uniformly recurrent, therefore w occurs in x infinitely many times with bounded distances. □

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. In particular, the analogue of Proposition 2 for non-pure case does not hold in general. The primitiveness notion also does not play such a role as in pure case. For instance, if $\phi\colon \Sigma_3^* \to \Sigma_3^*$ is as follows: $\phi(0) = 02$, $\phi(1) = 12$, $\phi(2) = 21$, and $h\colon \Sigma_3 \to \{1, 2\}$ is as follows: $h(0) = h(1) = 1$, $h(2) = 2$, then $\phi^\infty(0) = 0221211221121221\dots$, and $h(\phi^\infty(0)) = 1221211221121221\dots$ is non-ultimately-periodic uniformly recurrent (since it is Thue–Morse of 1 and 2), while ϕ is not primitive. Nevertheless it is worth mentioning that a uniformly recurrent automatic sequence can be generated by a primitive uniform morphism (see [5]).

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

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

Proposition 15. *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. An occurrence $u = x[i,j]$ in x is p-aligned if p divides i.

Proposition 16. *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 $y = 012\dots(k^m-2)(k^m-1)01\dots(k^m-1)0\dots$ be a periodic sequence with a period k^m. Then by Proposition 14 the sequence $h(\phi^\infty(s)) \times y$ 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 $m \le T$ the 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 $m \ge T$ for some T.

Furthermore, it follows from Lemma 12 that it is sufficient to check the only one such statement, as in the following

Proposition 17. *For all $r \ge R$: 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 18. *For all $r \ge R$: 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 in each k^m-aligned k^m-length segment for some sufficiently large m.

Now we explain how to check a condition from Theorem 18 in polynomial time.

Theorem 19. *There exists a polynomial-time algorithm deciding whether a given automatic sequence $h(\phi^\infty(s))$ is uniformly recurrent.*

Proof. We need to show two things: first, how to choose some $r \geq R$ 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 R might be exponentially large), and second, how to check whether for some m the symbols from this set occur in $\phi^m(b)$ for all $b \in A$.

Let us start with the second. Suppose we have found the set H of all 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 there exists m such that for all b we have $P_m^{(b)} \cap H \neq \varnothing$. First of all, observe that if $\forall b\ P_m^{(b)} \cap H \neq \varnothing$, then $\forall b\ P_l^{(b)} \cap H \neq \varnothing$ for all $l \geq m$. Second, notice that the sequence of n-tuples of sets $(\langle P_m^{(b)} \rangle_{b\in A})_{m=0}^\infty$ is ultimately periodic (recall that n is the size of the alphabet A). Indeed, the sequence $(P_m^{(b)})_{m=0}^\infty$ for each b is obviously ultimately periodic with both period and preperiod not greater than 2^n. Thus the period of $(\langle P_m^{(b)} \rangle_{b\in A})_{m=0}^\infty$ is not greater than the least common multiple of that for $(P_m^{(b)})_{m=0}^\infty$, $b \in A$, and the preperiod is not greater than the maximal that of $(P_m^{(b)})_{m=0}^\infty$, $b \in A$. So the period is not greater than $(2^n)^n = 2^{n^2}$ and the preperiod is not greater than 2^n. This means, by the first observation, that it is sufficient to choose one fixed $m \geq 2^{n^2} + 2^n$ and then to compute $\langle P_m^{(b)} \rangle_{b\in A}$ and to check intersections with H. Third, notice that there is a polynomial-time procedure that given an incidence graph of a morphism ψ (see Section 2 to recall what an incidence graph is) outputs an incidence graph of a morphism ψ^2. Thus after repeating this procedure n^2+1 times we obtain a graph by which we can easily find $\langle P_{2^{n^2+1}}^{(b)} \rangle_{b\in A}$. Note that $2^{n^2+1} > 2^{n^2} + 2^n$.

Similar arguments 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 $r \geq R$.

We recursively construct a series of graphs T_i. Let their 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 $\{\{\phi(b)(j), \phi(c)(j)\} \mid j = 1, \ldots, k, \phi(b)(j) \neq \phi(c)(j)\}$. 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 $\{x, y\} \in V_0(b, c)$ and 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 the 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 $\{x, y\} \in V_2(b, c)$.

In the same way all T_i are constructed. More explicitly, $V_i(b,c)$ is the set of all pairs of the form $\left\{\phi^{2^i}(b)(j), \phi^{2^i}(c)(j)\right\}$ with $1 \leq j \leq k^{2^i}$ and $\phi^{2^i}(b)(j) \neq \phi^{2^i}(c)(j)$. Remark also that $\phi^{2^i}(b) = \phi^{2^i}(c)$ if and only if $\{b,c\}$ has out-degree 0 in T_i.

It follows from Lemma 10 that $\log_2 R \leq n^2$. Thus after we repeat our procedure n^2 times, we will obtain the graph T_{n^2} such that $b \sim_{2^{n^2}} c$ if and only if $x \sim_0 y$ for all $\{x,y\} \in V_{n^2}(b,c)$. Recall that $x \sim_0 y$ means $h(x) = h(y)$, so now we can compute the set of symbols that are $\sim_{2^{n^2}}$-equivalent to s. □

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 after appropriate changes. The only note is that while proving analogue of Proposition 16 we should use the following statement instead of Proposition 14:

Proposition 20. *If x is recurrent and y is periodic, then $x \times y$ is recurrent.*

The proof of Proposition 20 is absolutely analogous to the proof of Proposition 14 and is left to the reader.

Now we can formulate the recurrence criterion for morphic sequences, analogously to Proposition 17:

Proposition 21. *For all $r \geq R$: 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.

Lemma 22. *The following problem can be solved in polynomial time: given an alphabet A, a morphism $\phi\colon A^* \to A^*$, and two symbols $s, a \in A$ such that ϕ is prolongable on s, decide whether a occurs infinitely many times in $\phi^\infty(s)$.*

Proof. Compute $u \in A^*$ such that $\phi(s) = su$. Let C denote the set of all letters occurring in u. The sequence $\phi^\infty(s)$ contains infinitely many occurrences of a if and only if there exists $c \in C$ such that a occurs in $\phi^n(c)$ for infinitely many integers $n \geq 0$. For each $c \in C$, compute the set A_c of all $b \in A$ such that there exist a path from c to b and a path from b to a in G_ϕ. The symbol a occurs in $\phi^n(c)$ for infinitely many integers $n \geq 0$ if and only if the subgraph of G_ϕ induced by A_c contains a cycle. □

In fact, the decidability without time constraints in Lemma 22 easily follows from the monadic logic approach, see Section 6.

Thus we obtain the following

Corollary 23. *There exists a polynomial-time algorithm deciding whether a given automatic sequence is recurrent.*

5. Factor Complexity

A factor (or subword) complexity is a natural combinatorial characteristic of words. A *factor complexity* of $x \in A^{\mathbb{N}}$ 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., [1], [10], or [2], Chapter 10. 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 [21] 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, the factor complexity of ultimately periodic sequences is $O(1)$, while for non-periodic sequences it is always $\Omega(n)$ according to [19]. It is also known that the factor complexity of automatic sequences is $O(n)$ (see [5]).

The factor complexity of uniformly recurrent sequences might be very large, for instance, exponential: for every α such that $0 < \alpha < 1$ there exists a uniformly recurrent sequence over m letters with factor complexity larger than or equal to $m^{\alpha n}$ (see [11] or [24]). Notice that α can not be put equal to 1 here. Indeed, for every letter a and every uniformly recurrent sequence $x \neq aaaa\ldots$, a^n occurs in x for at most finitely many $n \in \mathbb{N}$.

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

Theorem 24. *If x is a uniformly recurrent morphic sequence, then $p_x(n) = O(n)$.*

The proof of the theorem is in the following several lemmas. Probably, the keynote lemma is Lemma 30. Other important lemmas are Lemma 25, Lemma 27, and Lemma 28. Lemmas 26 and 29 are technical.

Lemma 25. *If x is a pure morphic sequence generated by a primitive morphism, then $p_x(n) = O(n)$.*

Lemma 25 is explicitly presented in [29] or in [2] (Theorem 10.4.12), but also follows from the results of [21].

Lemma 26. *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 $p_{f(x)}(n) \leq Mp_x(n)$ for every infinite word $x \in A^{\mathbb{N}}$ and $n \in \mathbb{N}$.*

Lemma 26 can be found in [4] or in [2] (Theorem 10.2.4).

Lemma 27 (Pansiot [21]) *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.*

We prove this lemma for completeness (especially because in [21] it is published in French).

Proof. Recall that I_ϕ is the set of ϕ-growing letters, and B_ϕ is the set of ϕ-bounded letters (see Section 2).

Let an alphabet C consist of all symbols $[twt']$ where twt' is a factor of $x = \phi^\infty(s)$, $t, t' \in I_\phi$, and $w \in B_\phi^*$. According to the conditions of the theorem, C is finite.

Define $\psi\colon C^* \to C^*$ as follows:

$$\psi([twt']) = [t_1 w_1 t_2][t_2 w_2 t_3] \ldots [t_k w_k t_{k+1}],$$

where $\phi(tw) = w_0 t_1 w_1 t_2 w_2 \ldots t_k w_k'$, $\phi(t')$ starts from $w_k'' t_{k+1}$, and $w_k = w_k' w_k''$, with $t_i \in I_\phi$ and $w_i, w_k', w_k'' \in B_\phi^*$. It can easily be seen that ψ is growing. Define also $g\colon C^* \to A^*$ as follows:

$$g([twt']) = tw.$$

Let x be represented as $t_0 w_0 t_1 w_1 t_2 w_2 \ldots$, where $t_i \in I_\phi$, $w_i \in B_\phi^*$; clearly $s = t_0$ is ϕ-growing. It can easily be seen that $\phi^i(s)$ is a prefix of $g(\psi^i([t_0 w_0 t_1]))$ for each $i \geq 1$ (therefore in particular ψ is prolongable on $[t_0 w_0 t_1]$). Thus $\phi^\infty(s) = g(\psi^\infty([t_0 w_0 t_1]))$. □

Lemma 28. *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 28 is a well-known minimality property of uniformly recurrent sequences, e.g., see [16].

Lemma 29. *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 30. *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 is a growing irreducible morphism from B^* to B^* (here we identify B with its set of vertices). According to Lemma 29, 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 occurs in x and therefore $(\phi^n)^m(t)$ for all m occur in x.

Suppose ϕ^n is not primitive. This means that B is a proper subgraph of G_ϕ, because otherwise ϕ^n is both prolongable and irreducible, and thus primitive. Now let us denote by ψ the morphism ϕ^n restricted to B, and repeat the procedure for ψ and G_ψ (which is actually B): find some strongly connected component of G_ψ with no outgoing edges, consider an appropriate power of ψ 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 Theorem 24.

Proof of Theorem 24. Suppose $x = h(\phi^\infty(s))$ is a 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 5 there exists a non-empty $w \in A^*$ such that w^n occurs in $\phi^\infty(s)$ for each n. Therefore $(h(w))^n$ occurs in x for each n, and hence x is periodic, which means its complexity is $O(1)$.

2) There are only finitely many ϕ-bounded factors in $\phi^\infty(s)$. Then by Lemma 27 $\phi^\infty(s)$ can be represented as $\phi^\infty(s) = g(\psi^\infty(t))$ for some morphisms $\psi\colon C^* \to C^*$ and $g\colon C^* \to A^*$ with ψ growing and g non-erasing. By Lemma 30 there exists a 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 28 we have $F(h(g(y))) = F(x)$. Thus for some constant M

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

where the middle inequality holds by Lemma 26 and the last equality holds by Lemma 25. □

Interestingly, almost nothing was known so far about factor complexity of arbitrary morphic sequences. Probably, the only progress was done in [22], where infinitely many examples with complexity of the form $\Theta(n^{1+\frac{1}{k}})$ for all $k = 1, 2, 3, \dots$ were constructed.

However, recently the following conjecture about the complexity of morphic sequences in general was formulated.

Conjecture 31 (Devyatov [8]) *The factor complexity of an 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* $k \in \mathbb{N}$.

Moreover, the following result partially solving Conjecture 31, was presented in [8]: the factor complexity of an arbitrary morphic sequence is either of the form $\Theta(n^{1+\frac{1}{k}})$ for some $k \in \mathbb{N}$, or of the form $O(n \log n)$. So it only remains to investigate the case of $O(n \log n)$ to solve Conjecture 31.

6. Arbitrary Morphic Sequences

Here we give some remarks concerning the general case.

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

Conjecture 32. *It is decidable, given an arbitrary morphic sequence, whether this sequence is uniformly recurrent or not.*

Proposition 33 given below in a sense supports Conjecture 32.

A very natural characteristic of a uniformly recurrent sequence is a uniform recurrence regulator. An *uniform recurrence regulator* of a 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 the uniform recurrence regulator somehow regulates how (uniformly) recurrent a sequence is.

Proposition 33. *Given a morphic sequence, one can compute its uniform recurrence regulator whenever this sequence is uniformly recurrent.*

Proof. First, notice that the set of factors of a morphic sequence is decidable. And even more, there exists an algorithm that given a morphic sequence and a word, decides whether this word occurs in the sequence. (For instance, this follows from the fact that the monadic theory of a morphic sequence is decidable — see below. There should be an explicit proof as well.)

Second, if a 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 $l \geq f(n)$. 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 $l \geq f(n)$. Thus to find precise value of $f(n)$, we can check all natural numbers starting from n until some of them works. □

Note that the monadic theory of a morphic sequence is decidable, e.g., see [3]. In fact, it also follows from [7] where it is shown that a finite transduction of a morphic sequence is morphic (see also [2], Theorem 7.9.1). (See more on monadic theories in context of combinatorics on words, e.g., in [31, 30, 34].)

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

(for each prefix u of x) (there are infinitely many occurrences of u in x).

The property "there are infinitely many occurrences of u in x" for morphic x can be algorithmically checked, since this property can be expressed in monadic language. Probably, the decidability of this property can also be proved directly (generalizing Lemma 22).

Suppose now that we have an oracle for halting problem, so-called $\mathbf{0}'$-oracle, and consider computations with this oracle. There exists the following simple description of $\mathbf{0}'$-computable functions. A (not necessarily total) function $f\colon \mathbb{N} \to \mathbb{N}$ is $\mathbf{0}'$-computable iff there exists a computable total function $F\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that $f(x) = \lim_{n\to\infty} F(x, n)$. The notion of $\mathbf{0}'$-computability (as well as the notion of computability itself) can be generalized to functions on objects that can be encoded by natural numbers. One can also define the notions of $\mathbf{0}'$-decidable and $\mathbf{0}'$-enumerable sets. In particular, every enumerable set is $\mathbf{0}'$-decidable (while the converse does not hold in general). More details on $\mathbf{0}'$-computations can be found, e.g., in [32].

The problem of determining recurrence for morphic sequences is $\mathbf{0}'$-decidable. Indeed, suppose we have a morphic sequence. Let us check for each prefix whether there are infinitely many occurrences of this prefix in the sequence. If we find some prefix for which it does not hold, we output the negative answer. Otherwise, if we never find such a bad prefix (that is what we can check using $\mathbf{0}'$-oracle), then we output a positive answer.

Thus we obtain the following

Proposition 34. *The problem of determining recurrence for morphic sequences is* $\mathbf{0}'$*-decidable.*

It is not difficult to see that the problem of determining uniform recurrence for morphic sequences can be written as

(for each prefix u of x) (there exists l) such that (u occurs in each l-length segment of x),

where the last property can be algorithmically checked for morphic sequences again by monadic logic reasons (and again probably a direct proof of this fact exists).

However, it turns out that this problem can be simplified.

Proposition 35. *The problem of determining uniform recurrence for morphic sequences is* $\mathbf{0}'$*-decidable.*

Proof. It follows from a careful examination of Section 5 that, given a morphic sequence x, one can compute a morphic sequence y satisfying the following three properties:

1) y is either periodic or generated by a primitive morphism,
2) $F(y) \subseteq F(x)$, and
3) x is uniformly recurrent if and only if $F(x) \subseteq F(y)$.

Thus, uniform recurrence of x can be expressed as

(for each factor u of x) (u occurs in y). □

7. Conclusion

Conjecture 32 remains the main open problem here. We solved it in the two particular cases of pure morphic sequences and automatic sequences by presenting the polynomial-time algorithms. Other particular cases also might be of great interest. Probably the most important is the case of $h(\phi^\infty(s))$ with ϕ growing, though this case does not seem to immediately imply Conjecture 32 in its general form. Notice that Theorem 7.5.1 from [2] allows us to represent an arbitrary morphic sequence as $h(\phi^\infty(s))$ with ϕ non-erasing, so it is sufficient to solve Conjecture 32 only for non-erasing morphisms.

Besides determining uniform recurrence for morphic sequences, similar problems can be formulated for variations with periodicity and uniform recurrence: ultimate periodicity, generalized uniform recurrence (called generalized almost periodicity in [26] and almost periodicity in [18, 20]), ultimate uniform recurrence, recurrence, ultimate recurrence, etc. If one notion is a particular case of another, it does not mean that corresponding criterion for the former is more difficult (or less difficult) than for the latter.

In this context the following problem can be stated.

Question 36. *Given two morphic sequences, can one determine whether their sets of subwords are equal?*

Besides equality, the variant of Question 36 for inclusion can be formulated as well. The positive answer would imply the positive solution of Conjecture 32, as we managed to show in Section 6 (while the converse implication does not seem to hold). Question 36 might be connected with the difficult open problem of determining equality of two given morphic sequences (see [2], Section 7.11, Problem 1; it was shown to be decidable in pure morphic case, see [6, 14]).

Of course, to continue investigations about factor complexity is also the problem of great interest. In particular, one can try to continue investigations about 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. J. Karhumäki is acknowledged for general support and hosting in University of Turku while working on some parts of the paper. The second author also especially thanks An. Muchnik and A. Semenov for their permanent support and discussions.

The work was supported by the Academy of Finland under grants 204785 (Automata Theory and Combinatorics on Words), 7523004 (Algorithmic Data Analysis) for the first author, and 208278 (the International Visitors Program) for the second author, as well as by grants 05-01-02803 (Russian Foundation for Basic Research) and 5351.2006.1 (Scientific Schools Supporting Council) for the second author.

References

[1] J.-P. Allouche. Sur la complexité des suites infinies. *Bulletin of the Belgian Mathematical Society—Simon Stevin*, 1(2):133–143, 1994.

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

[3] O. Carton and W. Thomas. The monadic theory of morphic infinite words and generalizations. *Information and Computation*, 176:51–65, 2002.

[4] J. Cassaigne and F. Nicolas. Quelques propriétés des mots substitutifs. *Bulletin of the Belgian Mathematical Society—Simon Stevin*, 10(5):661–677, 2003.

[5] A. Cobham. Uniform tag sequences. *Mathematical Systems Theory*, 6:164–192, 1972.

[6] K. Culik II and T. Harju. The ω-sequence equivalence problem for D0L systems is decidable. *Journal of the ACM*, 31(2):282–298, 1984.

[7] F. M. Dekking. Iteration of maps by an automaton. *Discrete Mathematics*, 126(1–3):81–86, 1994.

[8] R. Devyatov. On subword complexity of morphic sequences. In *Proceedings of CSR 2008*, volume 5010 of *Lecture Notes in Computer Science*, pages 146–157. Springer-Verlag, 2008.

[9] A. Ehrenfeucht and G. Rozenberg. Repetition of subwords in DOL languages. *Information and Control*, 59(1–3):13–35, 1983.

[10] S. Ferenczi. Complexity of sequences and dynamical systems. *Discrete Mathematics*, 206(1):145–154, 1999.

[11] C. Grillenberger. Constructions of strictly ergodic systems I. Given entropy. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 25:323–334, 1973.

[12] T. Harju and J. Karhumäki. Morphisms. In G. Rozenberg and A. Salomaa, editors, *Handbook of formal languages*, volume 1, pages 439–510. Springer-Verlag, 1997.

[13] T. Harju and M. Linna. On the periodicity of morphisms on free monoids. *RAIRO Informatique Théorique et Applications*, 20(1):47–54, 1986.

[14] J. Honkala. The D0L ω-equivalence problem. *International Journal of Foundations of Computer Science*, 18(1):181–194, 2007.

[15] K. Jacobs. *Maschinenerzeugte 0-1-Folgen.* Selecta Mathematica II. Springer-Verlag, 1970.

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

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

[18] A. Maes. More on morphisms and almost-periodicity. *Theoretical Computer Science*, 231(2):205–215, 2000.

[19] M. Morse and G. A. Hedlund. Symbolic dynamics. *American Journal of Mathematics*, 60(4):815–866, 1938.

[20] An. Muchnik, A. Semenov, and M. Ushakov. Almost periodic sequences. *Theoretical Computer Science*, 304(1–3):1–33, 2003.

[21] J.-J. Pansiot. Complexité des facteurs des mots infinis engendrés par morphismes itérés. In *Proceedings of ICALP'84*, volume 172 of *Lecture Notes in Computer Science*, pages 380–389. Springer-Verlag, 1984.

[22] J.-J. Pansiot. Subword complexities and iteration. *Bulletin of the European Association of Theoretical Computer Science*, 26:55–62, 1985.

[23] J.-J. Pansiot. Decidability of periodicity for infinite words. *RAIRO Informatique Théorique et Applications*, 20(1):43–46, 1986.

[24] Yu. Pritykin. Information in infinite words. In *Proceedings of WORDS 2007*, pages 254–261, 2007.

[25] Yu. Pritykin. On almost periodicity criteria for morphic sequences in some particular

cases. In *Proceedings of DLT 2007*, volume 4588 of *Lecture Notes in Computer Science*, pages 361–370. Springer-Verlag, 2007.

[26] Yu. Pritykin. Almost periodicity, finite automata mappings and related effectiveness issues. *Russian Mathematics (Izvestia VUZ. Matematika)*, to appear. Preprint arXiv:cs/0607009.

[27] Yu. L. Pritykin. Finite-automaton transformations of strictly almost-periodic sequences. *Mathematical Notes*, 80(5):710–714, 2006. Translated from Russian, *Matematicheskie Zametki*, 80(5):751–756, 2006.

[28] N. Pytheas Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, 2002.

[29] M. Queffélec. *Substitution Dynamical Systems—Spectral Analysis*, volume 1284 of *Lecture Notes in Mathematics*. Springer-Verlag, 1987.

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

[31] A. L. Semenov. Decidability of monadic theories. In *Proceedings of MFCS'84*, volume 176 of *Lecture Notes in Computer Science*, pages 162–175. Springer-Verlag, 1984.

[32] A. Shen and N. K. Vereshchagin. *Computable functions*, volume 19 of *Student Mathematical Library*. AMS, 2003. Translated from Russian: N. Vereshchagin, A. Shen. *Computable functions*. MCCME, 3rd edition, 2008.

[33] M. Sipser. *Introduction to the Theory of Computation*. PWS Publishing Company: Boston, 2nd edition, 2005.

[34] W. Thomas. Automata on infinite objects. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 133–191. Elsevier, 1990.