



ELSEVIER

Theoretical Computer Science 292 (2003) 9–31

**Theoretical
Computer Science**

www.elsevier.com/locate/tcs

Palindrome complexity

Jean-Paul Allouche^{a, *}, Michael Baake^b, Julien Cassaigne^c, David Damanik^d^a*CNRS, LRI, Université Paris-Sud, (Paris XI), Bât. 490, F-91405 Orsay Cedex, France*^b*Institut für Mathematik und Informatik, Universität Greifswald, Jahnstraße 15 a,
D-17487 Greifswald, Germany*^c*CNRS, IML Case 907, 163 Avenue de Luminy, F-13288 Marseille, Cedex 9, France*^d*Department of Mathematics 253-37, California Institute of Technology, Pasadena,
CA 91125, USA*

Dedicated to Jean Berstel for his 60th birthday with our very best wishes

Abstract

We study the *palindrome complexity* of infinite sequences on finite alphabets, i.e., the number of palindromic factors (blocks) of given length occurring in a given sequence. We survey the known results and obtain new results for some sequences, in particular for Rote sequences and for fixed points of primitive morphisms of constant length belonging to “class P” of Hof–Knill–Simon. We also give an upper bound for the palindrome complexity of a sequence in terms of its (block-)complexity. © 2002 Michael Baake: Published by Elsevier Science B.V. All rights reserved.

1. Introduction

The (block-)complexity function of an infinite sequence on a finite alphabet is the number of factors (blocks) of given length occurring in this sequence. This notion was introduced in 1975 by Ehrenfeucht et al. [27]. The complexity function of a sequence measures in some sense how “complicated” the sequence is: the reader is referred to the surveys [3, 29, 30]. Due inter alia to its applications to physics [32], another interesting “complexity” for an infinite sequence on a finite alphabet is its *palindrome complexity*, i.e., the number of palindromic factors (blocks) of given length occurring in the sequence. Combinatorial results on palindrome complexity for some sequences or classes of sequences were proved in [24, 4, 26, 10, 19, 21]. We survey known results, and give new ones, proving for example that *the palindrome complexity of Rote sequences is constant and equal to 2*. We prove that *the palindrome*

* Corresponding author. Tel.: +33-1-69-15-64-54; fax: +33-1-69-15-65-86.

E-mail addresses: allouche@lri.fr (J.-P. Allouche), mbaake@uni-greifswald.de (M. Baake), cassaigne@iml.univ-mrs.fr (J. Cassaigne), damanik@its.caltech.edu (D. Damanik).

complexity function of a sequence that is a fixed point of a primitive morphism of length d belonging to “class P ” of Hof–Knill–Simon and satisfying some technical conditions is d -automatic. Finally, we give an upper bound for the palindrome complexity in terms of the (usual) complexity.

2. Definitions and notations

2.1. Generalities

We will use the notation $\mathbb{N} = \{0, 1, 2, \dots\}$. We recall the following classical definitions in combinatorics of words. The sequences we consider in this paper are defined on a finite *alphabet* (i.e., on a finite set) \mathcal{A} . Elements of \mathcal{A} are called *letters*. The set \mathcal{A}^* is defined as the set of *words* on \mathcal{A} , i.e., the set of (possibly empty) strings of symbols of \mathcal{A} , equipped with the *concatenation*. In other words, \mathcal{A}^* is the free monoid for the concatenation generated by \mathcal{A} . The *length* of a word w , denoted by $|w|$, is recursively defined by: the empty word has length 0, and for any word w and any letter a , $|wa| = |w| + 1$. A word is called a *factor* of another word or of an infinite sequence if it occurs “without hole” in this word or sequence (factors are also called *subwords*, while the term *substrings* stands for the case where there are holes: 001 is a factor of 11001110 but only a substring of 0101010).

If \mathcal{A} and \mathcal{B} are two alphabets, homomorphisms for the concatenation from \mathcal{A}^* to \mathcal{B}^* are called *morphisms*. A *morphism* on \mathcal{A} (also called *substitution* or *inflation rule*) is a morphism from \mathcal{A}^* into itself. A morphism is defined by its values on the letters. A *constant length morphism* (or *uniform morphism*) is a morphism such that the images of all letters have the same length. If this length is equal to d , the morphism is also called a *morphism of length d* , or a *d -morphism*. A sequence on \mathcal{A} is called *d -automatic* if it is the pointwise image of a fixed point of a morphism of length d on an alphabet \mathcal{B} (pointwise image means, of course, image under a morphism of length 1 from \mathcal{B} to \mathcal{A}). A morphism φ on \mathcal{A} is called *primitive* if there exists an integer $k \geq 1$ such that the image of each letter by φ^k contains at least one occurrence of each letter of \mathcal{A} . A morphism is called *non-erasing* if the image of every letter is different from the empty word.

A (non-erasing) morphism on \mathcal{A} can be extended to infinite sequences with values in \mathcal{A} “by continuity”. (The set of sequences $\mathcal{A}^{\mathbb{N}}$ is equipped with the topology of simple convergence, i.e., the product topology where each copy of \mathcal{A} is equipped with the discrete topology.) An infinite sequence can thus be a *fixed point* of a morphism.

An infinite sequence on a finite alphabet is called *recurrent* if each word that occurs in the sequence occurs infinitely often. The sequence is called *uniformly recurrent* or *minimal* if each word that occurs in the sequence occurs infinitely often and the distance between two consecutive occurrences is bounded (some authors use the term *almost-periodic* for such sequences, while some authors call them *repetitive*). It is easy to prove that a sequence that is a fixed point of a primitive morphism is uniformly recurrent.

2.2. Periodicity

In this section we recall the definition of periodic sequences or words, and we give two theorems and a lemma that will prove useful.

Definition 1.

- An infinite sequence $u = u_0u_1\dots$ is called *periodic* if there exists an integer $T \geq 1$ (called a *period* of the sequence) such that for each $n \geq 0$ we have $u_{n+T} = u_n$. It is called *ultimately periodic* if there exists an integer ℓ such that the sequence v defined by $v_n = u_{\ell+n}$ is periodic.
- Let w be a finite word. Any integer $p \geq 1$ such that w is a prefix of an infinite sequence of period p is called a *period* of w . *The period* of w is the smallest such integer. (For example a period of the word 01101 is 5. Its period is 3.)

Theorems 1 and 2 below will prove useful. They are due respectively to Fine and Wilf [31] and to Lyndon and Schützenberger [37].

Theorem 1 (Fine–Wilf). *Let $u = (u_n)_{n \geq 0}$ and $v = (v_n)_{n \geq 0}$ be two periodic sequences with respective periods T and T' . If $u_n = v_n$ for more than $T + T' - \gcd(T, T')$ consecutive values of n , then the sequences u and v are equal. The value $T + T' - \gcd(T, T')$ is sharp.*

Remark 1. In the literature an easy corollary of this result is often called the theorem of Fine and Wilf, namely that *if a finite word w has two periods T and T' such that $|w| \geq T + T' - \gcd(T, T')$, then $\gcd(T, T')$ is a period of w . The value $T + T' - \gcd(T, T')$ is sharp.*

Theorem 2 (Lyndon–Schützenberger). *Let \mathcal{A} be an alphabet. Let $x, y, z \in \mathcal{A}^*$, with x and z non-empty. Then $xy = yz$ if and only if there exist $u, v \in \mathcal{A}^*$, and an integer $e \geq 0$ such that $x = uv$, $z = vu$, and $y = (uv)^e u = u(vu)^e$.*

Lemma 1.

- If the period T of a word w satisfies $T \leq |w|/2$, then all the periods of w that are $\leq |w|/2$ are divisible by T .
- Let z be a word and let w be a factor of z . If z has period T , if w has period T' , and if $T + T' \leq |w|$, then T' is a period of z .
- Let z be a word and let w and w' be two factors of z . Suppose that T , the period of w , T' , the period of w' , and Θ , the period of z , satisfy $\Theta + T \leq |w|$ and $\Theta + T' \leq |w'|$. Then $T = T'$.

Proof.

- If T is the period of w , if U is another period, then w is a prefix of a T -periodic sequence and of a (possibly distinct) U -periodic sequence. If both periods are $\leq |w|/2$, then $T + U \leq |w|$. Hence the two infinite periodic sequences, coinciding on a prefix

of length $\geq T + U$, must be equal (from Theorem 1). The periodic sequence thus obtained admits T as smallest period (namely T is the smallest period of the word w). Hence T divides U . (Actually it can also be noted that if $T > |w|/2$, then our statement is empty, hence holds.)

- The word z is a prefix of a T -periodic sequence u , and the word w is a prefix of a T' -periodic sequence u' . But the word w is a factor of z , hence a prefix of a sequence of period T , say \hat{u} , obtained from the sequence u by erasing some prefix. Since $T + T' \leq |w|$ we have, from Theorem 1, that $u' = \hat{u}$. Hence T' is a period of u , since u can be obtained by erasing a prefix of \hat{u} . Hence T' is a period of z .
- From the second item above, T and T' are periods of z . Hence clearly T' is a period of w and T is a period of w' . Since T is the minimal period of w and T' the minimal period of w' , we have $T \leq T'$ and $T' \leq T$. Hence $T' = T$. \square

2.3. Palindromes and complexity

Definition 2. If $w = w_1w_2 \dots w_j$ is a word on the alphabet \mathcal{A} , we denote by \tilde{w} the word obtained by reading w backwards, i.e., $\tilde{w} = w_jw_{j-1} \dots w_2w_1$. A *palindrome* is a word w such that $w = \tilde{w}$. (For example the words “level” and “deed” are palindromes in English.)

Definition 3. Let $u := u_0u_1u_2 \dots$ be a sequence on the finite alphabet \mathcal{A} . We denote by $\text{fac}_u(n)$ the number of words of length n that are factors of the sequence u . We denote by $\text{pal}_u(n)$ the number of palindromes of length n that are factors of the sequence u .

Remark 2. The notations used in other papers might differ. In the literature $p_u(n)$ sometimes stands for the block-complexity and sometimes for the palindrome complexity. We hope that our terminology is unambiguous.

2.4. Sturmian sequences

We end this section by recalling the definition of Sturmian sequences. These sequences can be obtained by playing billiard on squares, starting with an irrational slope. They can also be defined by their complexity. The reader is referred to [41, 18, 36, 11, 12].

Definition 4. A *Sturmian sequence* is a sequence $u = (u_n)_{n \geq 0}$ whose (block-)complexity $\text{fac}_u(k)$ satisfies: $\forall k \geq 1, \text{fac}_u(k) = k + 1$.

3. Motivation in physics

Given a uniformly recurrent sequence u , one may consider the associated *LI-class* Ω (also called *hull* or *induced subshift*) which consists of all two-sided infinite sequences that have the same finite factors as u . If u is Sturmian, if u is generated by

a primitive morphism, or if u is derived from a standard cut and project scheme, this gives widely used models of *one-dimensional quasicrystals*. That is, a one-dimensional quasicrystal is modelled by a suitable family of two-sided sequences which are locally indistinguishable since they have the same finite factors.

To such a structure, one may associate a family of discrete one-dimensional Schrödinger operators $(H_\omega)_{\omega \in \Omega}$ as follows: choose an injective function $f: \mathcal{A} \rightarrow \mathbb{R}$ and define, for every $\omega \in \Omega$, the operator H_ω in $\ell^2(\mathbb{Z})$ by

$$(H_\omega \phi)(n) = \phi(n+1) + \phi(n-1) + f(\omega_n)\phi(n).$$

The spectral properties of H_ω determine the “conductivity properties” of the given structure. Roughly speaking, if the spectrum is absolutely continuous, then the structure behaves like a conductor, while in the case of pure point spectrum, it behaves like an insulator. The intermediate spectral type—singular continuous spectrum—is generally expected to give rise to intermediate transport properties, but no one currently understands this correspondence very well.

For classical (periodic or atomic) structures, singular continuous spectra do not occur. However, for one-dimensional quasicrystals, this spectral type appears to be rather typical, and there has been a lot of recent research activity focussing on results of this kind. One important contribution in this direction is the paper of Hof et al. [32], where a sufficient criterion for purely singular continuity of individual such operators is derived in terms of a strong palindromicity property of the underlying sequence.

From a combinatorial point of view, it is interesting to note that the criterion of Hof, Knill, and Simon deduces singular continuous spectrum from explicit combinatorial (and dynamical) properties of u . Suppose that the factors of u occur with well defined, positive frequencies (Ω is strictly ergodic). This is the case for a large class of sequences, including Sturmian sequences and sequences generated by primitive morphisms, but also for sequences derived from the standard or generalized cut and project schemes [47]. Then the following holds [32].

Theorem 3 (Hof–Knill–Simon). *Let u be a sequence on a finite alphabet that is not ultimately periodic (hence $\text{fac}_u(k) \geq k+1$ for every $k \geq 1$), and such that pal_u does not ultimately vanish, i.e., $\limsup_{k \rightarrow \infty} \text{pal}_u(k) > 0$. We suppose that the subshift Ω induced by u is strictly ergodic. Then, for uncountably many $\omega \in \Omega$, the operator H_ω has purely singular continuous spectrum.*

This theorem applies to a large class of sequences generated by primitive morphisms, to all Sturmian sequences, and, more generally, to all sequences defined by circle maps. Furthermore, it also applies to sequences derived from a standard cut and project scheme with inversion-symmetric window (see [10] for details). Note that by uniform recurrence of u , it suffices to assume that u is not periodic. We remark that there is a similar, purely combinatorial, sufficient condition for purely singular continuous spectrum in terms of *powers* occurring in u , see [20] for a survey.

4. Survey of known results

4.1. Rudin–Shapiro and paperfolding sequences

Paperfolding sequences are binary sequences obtained by repeatedly folding a strip of paper (see [23] for example). We recall the definition.

Definition 5. A sequence $(u_n)_{n \geq 1}$ with values in $\{0, 1\}$ is called a *paperfolding sequence* if there exists a sequence i_0, i_1, i_2, \dots , with $i_k \in \{0, 1\}$ (called the sequence of *unfolding instructions*) such that

$$\forall m \geq 0, \forall j \geq 0, \quad u_{2^m(2j+1)} \equiv j + i_m \pmod{2}.$$

(Note that every integer $n \geq 1$ can be uniquely written as $n = 2^m(2j+1)$, with $m, j \geq 0$.)

Generalized Rudin–Shapiro sequences (in the sense of [38]) are obtained by “integrating modulo 2” paperfolding sequences. More precisely

Definition 6. A sequence $(v_n)_{n \geq 0}$ is called a *generalized Rudin–Shapiro sequence* if $v_0 = 0$ and if there exists a paperfolding sequence $(u_n)_{n \geq 1}$ such that, for every $n \geq 0$, there holds $v_n \equiv \sum_{k=1}^n u_k \pmod{2}$. (The “classical” Rudin–Shapiro sequence corresponds to the paperfolding sequence with unfolding instructions $0, 0, 1, 0, 1, 0, 1, 0, 1, \dots$.)

Remark 3. There exist other generalizations of the classical Rudin–Shapiro sequence (see, e.g., [44, 6]) but we restrict here to the ones given in Definition 6 above.

The following theorem was proved in [4]. It was proved again in a more efficient and general way by Baake [10].

Theorem 4 (Allouche). *The palindrome complexity of any paperfolding sequence $u = (u_n)_{n \geq 1}$ satisfies $\text{pal}_u(k) = 0$ for any $k \geq 14$. The palindrome complexity of any generalized Rudin–Shapiro sequence $v = (v_n)_{n \geq 0}$ satisfies $\text{pal}_v(k) = 0$ for any $k \geq 15$.*

4.2. The period-doubling sequence

Definition 7. The *period-doubling sequence* is defined as the infinite fixed point of the morphism $0 \rightarrow 01, 1 \rightarrow 00$.

The following result was proved in [19].

Theorem 5 (Damanik). *The palindrome complexity of the period-doubling sequence u satisfies*

$$\begin{aligned} \forall k \text{ even}, k \geq 4 &\Rightarrow \text{pal}_u(k) = 0, \\ \forall k \text{ odd}, k \geq 5 &\Rightarrow \text{pal}_u(k) = \text{pal}_u(2k - 1) = \text{pal}_u(2k + 1) \end{aligned}$$

and the first few values of $\text{pal}_u(k)$ are given by

k	1	2	3	4	5	6	7
$\text{pal}_u(k)$	2	1	3	0	4	0	3

In particular, the function $k \rightarrow \text{pal}_u(k)$ takes its values in the set $\{0, 1, 2, 3, 4\}$. Furthermore, $\limsup_{k \rightarrow \infty} \text{pal}_u(k) = 4$. (Actually 0, 3, and 4 are the only values that are taken infinitely often.)

4.3. Sturmian sequences

The following nice characterization of Sturmian sequences in terms of palindrome complexity was given in [26] (see also [24] for the Fibonacci sequence, which can be defined as the fixed point of the morphism $0 \rightarrow 01, 1 \rightarrow 0$).

Theorem 6 (Droubay–Pirillo). *A sequence $u = (u_n)_{n \geq 0}$ is Sturmian if and only if its palindrome complexity satisfies: $\forall k$ odd, $\text{pal}_u(k) = 2$ and $\forall k \geq 2$ even, $\text{pal}_u(k) = 1$.*

Remark 4. A generalization of this theorem to the two-dimensional case has been obtained recently by Berthé and Vuillon [13]. For a study of palindromes in *episturmian sequences*, see [25, 33].

4.4. Fixed points of primitive morphisms

Primitive morphisms are often considered because they have dynamical or combinatorial properties that other morphisms may not have. The following result was proved in [21].

Theorem 7 (Damanik–Zare). *The palindrome complexity $\text{pal}_u(k)$ of a fixed point $u = (u_n)_{n \geq 0}$ of a primitive morphism is bounded (hence takes only finitely many values).*

Remark 5. As a consequence of Theorem 12 below, we will have that the conclusion of Theorem 7 also holds for uniform (not necessarily primitive) morphisms.

5. Rote sequences

Sequences of (block-)complexity $2k$ were studied in [46]. In that paper, Rote proved in particular [46, Theorem 3] that an infinite 0, 1-sequence $w = (w_n)_{n \geq 0}$ is a complementation-symmetric sequence with block-complexity $2k$ if and only if its first difference (modulo 2) sequence $\beta = (\beta_n)_{n \geq 0}$, is Sturmian, where, for each $n \geq 0$, $\beta_n := w_{n+1} - w_n \pmod 2$. Recall that a complementation-symmetric sequence on a two-letter alphabet, say $\mathcal{A} = \{a, b\}$, is a sequence such that for any block occurring in it,

the block obtained by changing a 's into b 's and b 's into a 's is also a factor. Using Theorem 6 we can compute the palindrome complexity of these sequences.

Theorem 8. *Let $w = (w_n)_{n \geq 0}$ be a complementation-symmetric sequence with complexity $\text{fac}_w(k) = 2k$ for all $k \geq 1$. Then its palindrome complexity satisfies $\text{pal}_w(k) = 2$ for all $k \geq 1$.*

Proof. Since $\text{fac}_w(1) = 2$, we see that the sequence w is a binary sequence. Without loss of generality, we may assume that it is a 0,1-sequence. Define $\beta = (\beta_n)_{n \geq 0}$ by: $\forall n \geq 0 \beta_n := w_{n+1} - w_n \pmod 2$. Then we know that the sequence β is Sturmian.

Define for each $k \geq 1$ the maps Φ_k and Ψ_k^s (where $s = 0, 1$) on words of length k on $\{0, 1\}$ by

$$\begin{aligned} \Phi_k(a_1 a_2 \dots a_k) &:= b_1 b_2 \dots b_{k-1}, \quad \text{where } b_i := a_{i+1} - a_i \pmod 2, \\ \Psi_k^s(b_1 b_2 \dots b_k) &:= s c_1 c_2 \dots c_k, \quad \text{where } s = 0, 1 \text{ and } c_j := s + \sum_{i=1}^j b_i \pmod 2. \end{aligned}$$

It is straightforward if k is even that Φ_k sends palindromes of length k to palindromes of length $k - 1$ whose central letter is 0, and that any palindrome π of length $k - 1$ whose central letter is 0 is the image under Φ_k of exactly two palindromes of length k , namely $\Psi_{k-1}^0(\pi)$ and $\Psi_{k-1}^1(\pi)$.

It is also straightforward if k is odd that Φ_k sends palindromes of length k to palindromes of length $k - 1$, and that any palindrome π of length $k - 1$ is the image under Φ_k of exactly two palindromes of length k , namely $\Psi_{k-1}^0(\pi)$ and $\Psi_{k-1}^1(\pi)$.

Now from this property of the map Φ_k and from the relation between the sequences w and β , we see that the number of palindromes of length k occurring in the sequence w is equal to twice the number of palindromes of length $k - 1$ whose central letter is 0 occurring in β if k is even, and to twice the number of palindromes of length $k - 1$ occurring in β if k is odd.

To conclude the proof it remains to note that any Sturmian sequence has exactly one palindrome of length k if k is even [26], and exactly one palindrome of length k whose central letter is 0 if k is odd: the proof of Proposition 6 of [26] shows there is a bijection between the set of palindromes of length $k + 2$ and the set of palindromes of length k and that this bijection consists of erasing the first and last (identical) letters. Since there is exactly one palindrome of length 1 with central letter 0 and one palindrome of length 1 with central letter 1, we see immediately that, for each odd k , any Sturmian sequence has exactly one palindrome of length k whose central letter is 0 and exactly one palindrome of length k whose central letter is 1. \square

Remark 6.

- Another way of studying palindromes occurring in a sequence is to use the associated dynamical system and its (geometric) symmetry properties. In this direction, the

reader can look at [1] (see for example Theorem 19 in that paper) and [13] (where, in particular, a two-dimensional version of Theorem 8 above is given).

- Taking the first difference sequences modulo 2 of 0,1-sequences for studying factors and complexities of certain sequences was already used in [2, 5, 46, 4] for example.
- What is the palindrome complexity of *any* sequence of complexity $2k$? Rote proves in [46] that the fixed point u of the morphism $0 \rightarrow 001, 1 \rightarrow 111$ has complexity $2k$. With a slight modification of the argument in Theorem 9 below (this morphism is not primitive, but as soon as a word contains a 0, we essentially know from which word it “comes by the morphism”) the reader can prove that $\text{pal}_u(k) = 2$ for each $k \geq 1$. On the other hand, looking at another example of a sequence of complexity $2k$ given in [46], namely the image under the morphism $a \rightarrow 0, b \rightarrow 1, c \rightarrow 10110, d \rightarrow 101$ of the fixed point of the morphism $a \rightarrow ad, b \rightarrow bac, c \rightarrow bacab, d \rightarrow baca$, it is easy to check that this sequence contains only one palindrome of length 6. (More precisely, we have for this sequence $\text{pal}(k) = 2$, for $k = 1, 2, 3, 4, 5$, $\text{pal}(k) = 1$, for $k = 6, 7, 8, 9, 10$, and $\text{pal}(k) = 0$ for $k \geq 11$.)
- Let $v = v_0v_1 \dots$ be the fixed point of the morphism $0 \rightarrow 001, 1 \rightarrow 101$. The complexity of this sequence is given by $\text{fac}_v(k) = 2k$ for $k \geq 1$ (see [28]). The reader can prove that the palindrome complexity of this sequence is 2 for $k \leq 7$ and 0 for $k \geq 8$. (Hint: prove there is no palindrome of length 8 nor of length 9, either by mimicking the method of Theorem 4 above, or by using the recursive definition of the sequence v : $u_{3n} = u_n, u_{3n+1} = 0, u_{3n+2} = 1$.) In the same vein, the reader can prove that the Chacon sequence defined as the infinite fixed point of $0 \rightarrow 0010, 1 \rightarrow 1$ does not contain palindromes of length 13, nor palindromes of length 14, hence it does not contain palindromes of length ≥ 13 . Note that the (usual) complexity of the Chacon sequence is equal to $2k - 1$ for $k \geq 2$ (see [28]).
- We finally note that the (conjectured) palindrome complexity of the Kolakoski sequence is also constant and equal to 2 [35, Section 4.1.3]. Recall that the Kolakoski sequence is the sequence

2 2 1 1 2 1 2 2 1 ...

defined as *the* sequence on the alphabet $\{1, 2\}$ that begins in 2 and such that the sequence of its runlengths is equal to the sequence itself (see [34], see [22] for a recent survey). Recall that a word w on the alphabet $\{1, 2\}$ is said to be differentiable if neither 111 nor 222 occurs, and its derivative w' is the finite sequence of lengths of blocks in w , discarding the first and/or last block if it has length one: $(12211)' = 22, (121)' = 1$. All factors of Kolakoski are C^∞ -words, i.e., words that can be differentiated infinitely many times. The converse is an open conjecture. It is proved in [35, Section 4.1.3] that the number of C^∞ -palindromes of each length ≥ 1 is 2. Hence, the palindrome complexity of the Kolakoski sequence is bounded by 2, and it is conjecturally constant and equal to 2.

6. Fixed points of uniform primitive morphisms

We first recall the definition of *class P morphisms* introduced by Hof et al. [32].

Definition 8 (*Hof–Knill–Simon*). A morphism σ on the (finite) alphabet \mathcal{A} belongs to class P if there exists a palindrome p and for every $a \in \mathcal{A}$ a palindrome q_a such that, for every $a \in \mathcal{A}$, we have $\sigma(a) = pq_a$ (or, for every $a \in \mathcal{A}$, $\sigma(a) = q_a p$). The word p can be empty. If p is not empty, then some (or even all) q_a 's are allowed to be empty.

We show in the following theorem that it is possible to compute the palindrome complexity of fixed points of uniform morphisms that belong to class P.

Theorem 9. *Let $\sigma : \mathcal{A} \rightarrow \mathcal{A}^*$ be primitive and $u \in \mathcal{A}^{\mathbb{N}}$ such that $\sigma(u) = u$. We assume the following:*

- (i) *The morphism σ belongs to class P: there exists a palindrome p and, for every $a \in \mathcal{A}$, a palindrome q_a such that $\sigma(a) = pq_a$ (the case where $\sigma(a) = q_a p$ for every $a \in \mathcal{A}$ is analogous).*
- (ii) *The morphism σ is uniform. Let $l_p := |p|$ and $l_q := |q_a|$, for every $a \in \mathcal{A}$.*
- (iii) *For $a \neq b$, q_a and q_b have distinct first (and hence last) symbols.*

Under these conditions we have the following recursion for the palindrome complexity

$$\exists n_0 \in \mathbb{N} \setminus \{0\}, \forall n \geq n_0, \text{pal}(n) = \sum_{k \in \mathcal{E}} \text{pal}(k),$$

where

$$\mathcal{E} = \{s, n = sl + l_p - 2j, \text{ where } 0 \leq j \leq l - 1\}, \text{ with } l := l_p + l_q.$$

Proof. Let us first explain our idea intuitively. Given any palindrome $x = x_1 \dots x_n$ occurring in u , we can associate with x the palindrome $\sigma(x)p = pq_{x_1}pq_{x_2} \dots pq_{x_n}p$ which is a factor of u since $\sigma(u) = u$. Moreover, we can simultaneously delete symbols at the beginning and end of $\sigma(x)p$, retaining the palindromic form, without deleting q_{x_1} (and hence q_{x_n}) completely. Thus we can associate with x a family of palindromes. By assumption (iii), this map is one-to-one when restricted to the set of palindromes of length n occurring in u . Conversely, we can do an inverse procedure by re-substituting a given palindrome. That is, given a palindrome we can decompose it (essentially uniquely by a result of Mossé [42, 43]) and consider its inverse image under σ which will turn out to be a palindrome. Again by assumption (iii), this process is one-to-one and we can thus establish a bijection between suitable sets of palindromes. As a consequence, we get equality of their cardinalities which yields recursive relations for the palindrome complexity of u .

Let us be more precise. Given $n \in \mathbb{N} \setminus \{0\}$ (this will be the length of the long palindrome), we are looking for solutions s in $\mathbb{N} \setminus \{0\}$ (this will be the length of the short

palindrome the long palindrome is coming from) of the equation

$$n = sl + l_p - 2j \text{ subject to the condition } 0 \leq j \leq l - 1, \text{ with } l = l_p + l_q. \quad (1)$$

It is clear that there are at most two solutions to this problem, and if there are two solutions s_1, s_2 , then l is even and $|s_1 - s_2| = 1$. We will show the following:

There exists $n_0 \in \mathbb{N} \setminus \{0\}$ such that for every $n \geq n_0$, we have

$$\text{pal}(n) = \sum_{k \text{ solves Equation 1}} \text{pal}(k). \quad (2)$$

Let $\text{Pal}(m)$ denote the set of palindromes of length m that are factors of u . We therefore have $\text{pal}(m) = \#\text{Pal}(m)$. Let $n \in \mathbb{N} \setminus \{0\}$. Assume first that Eq. (1) has two solutions $k, k + 1$. We will define two maps

$$\Phi : \text{Pal}(k) \cup \text{Pal}(k + 1) \rightarrow \text{Pal}(n), \quad \Psi : \text{Pal}(n) \rightarrow \text{Pal}(k) \cup \text{Pal}(k + 1)$$

and show that, for n large enough, they are one-to-one. Since $\text{Pal}(k)$ and $\text{Pal}(k + 1)$ are clearly disjoint, this implies Eq. (2).

The map Φ is defined as explained above. Start with a palindrome x in $\text{Pal}(k) \cup \text{Pal}(k + 1)$ and consider the associated palindrome $\sigma(x)p$. Then, by Eq. (1), $|\sigma(x)p| \geq n$ and we obtain an element of $\text{Pal}(n)$ by “pruning” $\sigma(x)p$ suitably. Call this element $\Phi(x)$. It follows from assumption (iii) and Eq. (1) that Φ restricted to either $\text{Pal}(k)$ or $\text{Pal}(k + 1)$ is one-to-one. To see that Φ is one-to-one also on their union, we invoke the recognizability property of u as proven by Mossé, see [42, 43]. This gives the claim immediately in the case $l_p \neq l_q$ (for n large enough to apply Mossé’s result and to compare two decompositions) and in the case $l_p = l_q$ one uses that u is N th power-free [42] for some N together with this argument to prove the claim.

The definition of Ψ is slightly more complicated. Let $x \in \text{Pal}(n)$ be given. According to Mossé’s result, for n large enough, one can draw one “bar” and hence obtain a unique decomposition of x using assumption (ii). (Recall that a sequence $u = (u_n)_{n \geq 0}$ fixed point of a morphism φ can be written

$$u = \varphi(u) = / \varphi(u_0) / \varphi(u_1) / \dots$$

and any factor of u will “contain” bars—possibly in a non-unique way—according to its position(s) in the sequence u . Note that for a constant-length morphism, knowing the position of one bar gives the positions of the other bars.) We will consider a slightly different decomposition. Namely, we will also draw bars between the p ’s and the q_a ’s, $a \in \mathcal{A}$. It is easy to see that for n large enough, these modified decompositions are unique as well.

We thus have

$$x = \sigma / w_1 / w_2 / \dots / w_r / \pi, \quad (3)$$

where w_i is equal to either p or one of the q_a ’s, $1 \leq i \leq r$, σ is a non-empty suffix and π a non-empty prefix of p or of some q_a . Since x is a palindrome, the “reflected

modified decomposition”

$$x = \tilde{\pi}/w_r/w_{r-1}/\dots/w_1/\tilde{\sigma} \quad (4)$$

corresponds to another modified decomposition of x which must be the same by uniqueness. This shows that $\sigma = \tilde{\pi}$ and $w_i = w_{r-i+1}$, $1 \leq i \leq r$. Moreover, the center of x must also be a center of some w_j in this representation. Finally, we have $w_i = p$ either for all odd i or for all even i . Let us consider the first case. We have

$$x = \sigma/p/q_{x_1}/p/q_{x_2}/p\dots/p/q_{x_2}/p/q_{x_1}/p/\pi \quad (5)$$

and we can associate some q_{x_0} with both σ and π (by assumption (iii)). We therefore get the palindrome $x_0x_1x_2\dots x_2x_1x_0$ which belongs to either $P(k)$ or $P(k+1)$, for otherwise Eq. (1) would have a solution different from $k, k+1$. In the second case we proceed similarly, in this case σ/π are suffix/prefix of p and we obtain a palindrome $x_1x_2\dots x_2x_1$ which, by the same reasoning, belongs to $P(k)$ or $P(k+1)$. Call the obtained palindrome $\Psi(x)$. Using assumption (iii) and the unique decomposition property, we see that Ψ is one-to-one. This establishes the existence of a bijection between $\text{Pal}(k) \cup \text{Pal}(k+1)$ and $\text{Pal}(n)$ and hence proves the assertion of the theorem in the case where Eq. (1) has two solutions.

If there is one solution to Eq. (1), we can prove Eq. (2) along the same lines, parts of the argument being even simpler than in the two-solution case.

On the other hand, if $\text{Pal}(n)$ is non-empty, we can define Ψ as above and our re-substitution argument then shows that Eq. (1) must have a solution. Hence the absence of such a solution implies $\text{Pal}(n) = \emptyset$, that is, $\text{pal}(n) = 0$. \square

Remark 7. This theorem applies to several prominent examples, such as the period doubling morphism ($\mathcal{A} = \{a, b\}$, $p = a$, $q_a = b$, $q_b = a$) and the square of the Thue–Morse morphism ($\mathcal{A} = \{a, b\}$, $p = \varepsilon$, $q_a = abba$, $q_b = baab$). It also applies to the square of some generalizations of the Thue–Morse morphism given in [8] (see Lemma 2 and Theorem 8 of [8] in the case where $m \in \{1, 2\}$).

We know from Theorem 7 that the palindrome complexity of a fixed point u of a primitive morphism takes only finitely many values. It then makes sense to ask whether the sequence $(\text{pal}_u(k))_{k \geq 1}$ itself is generated by a morphism. If the morphism belongs to class P, satisfies some technical conditions, and has constant length, we give an answer in our next theorem.

Theorem 10. *Let $u = (u_n)_{n \geq 0}$ be a sequence that is a fixed point of a primitive morphism belonging to class P and satisfying the extra conditions (i)–(iii) of Theorem 9. In particular the morphism is uniform: let d be its length. Then the palindrome complexity $(\text{pal}_u(k))_{k \geq 1}$ is a d -automatic sequence.*

Proof. Let $i \in [0, l-1]$ be a fixed integer and let n be an integer ≥ 0 ($n \geq 1$ if $i = 0$). We want to compute $\text{pal}_u(ln + i)$ using Theorem 9 above. We distinguish two cases:

- If l is odd, the congruence $l_p - 2j \equiv i \pmod{l}$ has a unique solution, say j_0 , belonging to $[0, l - 1]$. Then, from Theorem 9, we have for n large enough

$$\text{pal}_u(ln + i) = \text{pal}_u\left(n + \frac{i - l_p + 2j_0}{l}\right).$$

Let $f(n) := \text{pal}_u(n - 1)$ for $n \geq 2$, then

$$f(ln + i + 1) = f\left(n + \frac{l + i - l_p + 2j_0}{l}\right). \tag{6}$$

We claim this implies that for $a \in [0, l - 1]$, the sequence $(f(ln + a))_{n \geq \max(n_0, 2)}$ is a linear combination of the sequences $(f(n + p))_{n \geq \max(n_0, 2)}$, with $p \in [0, 3]$.

This is clear for $a \in [1, l - 1]$ from Eq. (6) above. Using Eq. (6) with $i = l - 1$, we obtain for n large enough,

$$f(ln + l) = f\left(n + 1 + \frac{l - 1 - l_p + 2j_0}{l}\right).$$

Hence replacing n by $n - 1$,

$$f(ln) = f\left(n + \frac{l - 1 - l_p + 2j_0}{l}\right),$$

and the claim is proved.

- If l is even, the congruence $l_p - 2j \equiv i \pmod{l}$ either has no solution (if l_p and i have opposite parities) or has two solutions, say j_1 and j_2 , belonging to $[0, l - 1]$. In the first case $\text{pal}_u(ln + i) = 0$. In the second case we have for n large enough

$$\text{pal}_u(ln + i) = \text{pal}_u\left(n + \frac{i - l_p + 2j_1}{l}\right) + \text{pal}_u\left(n + \frac{i - l_p + 2j_2}{l}\right).$$

We conclude as above that $f(n) := \text{pal}_u(n - 1)$ for $n \geq 2$ has the property that for every $a \in [0, l - 1]$, the sequence $(f(ln + a))_{n \geq \max(n_0, 2)}$ is a linear combination of the sequences $(f(n + p))_{n \geq \max(n_0, 2)}$, with $p \in [0, 3]$.

Now from a result of [9] quoted below as Theorem 11, we see that the sequence $(f(n))_n$ (and hence the sequence $(\text{pal}_u(n))_{n \geq 1}$) is d -regular in the sense of [7]. But this sequence takes only finitely many values, hence as noted in [7] it must be d -automatic. \square

Before giving the next statement for the sake of completeness, we first recall (see [7]) that a sequence $u = (u_n)_{n \geq 0}$ with values in a Noetherian ring R is called d -regular for some integer $d \geq 1$ if the R -module generated by all subsequences $(u_{d^t n + \ell})_{n \geq 0}$ with $t \geq 0$ and $\ell \leq d^t - 1$ has finite dimension. The following result is proved in [9].

Theorem 11 (Allouche–Shallit). *Let $u = (u_n)_{n \geq 0}$ be a sequence with values in a Noetherian ring R . Suppose there exist an integer $d \geq 2$, an integer $t \geq 0$, an integer $r \geq 0$ and an integer $n_0 \geq 0$ such that each sequence $(u_{d^{t+1}n + \ell})_{n \geq n_0}$ for*

$\ell \in [0, d^{t+1} - 1]$ is a linear combination of the sequences $(u_{d^j n+i})_{n \geq n_0}$ with $j \leq t$, $i \leq d^j - 1$ and of the sequences $(u_{n+p})_{n \geq n_0}$ with $p \leq r$. Then the sequence u is d -regular.

7. Bounding the palindrome complexity in terms of the usual complexity

Looking at some of the examples above we can ask several questions:

- We saw that Sturmian sequences, complement-symmetric Rote sequences, the (nonconstant) fixed point of $0 \rightarrow 001$, $1 \rightarrow 111$, and fixed points of primitive morphisms all have bounded palindrome complexity. On the other hand, for all these sequences the complexity satisfies $\text{fac}(k) = O(k)$. Is it true that the property $\text{fac}(k) = O(k)$ implies $\text{pal}(k) = O(1)$? We will answer this question positively in Theorem 12 below. Note that the converse is not true, by far. For example, apply to a binary sequence u with complexity $\text{fac}_u(k) = 2^k$ the morphism $0 \rightarrow 011001$, $1 \rightarrow 001011$. Then the complexity fac_v of the image v of u under this morphism satisfies $2^{k/6} \leq \text{fac}_v(k) \leq 9 \times 2^{k/6}$, whereas the palindrome complexity drops to 0 from $k = 8$.
- More generally, for which sequences is it true that $\text{pal}(k) = O(\text{fac}(k)/k)$? For which sequences is it true that $0 < \limsup_{k \rightarrow \infty} (k \text{pal}(k)/\text{fac}(k)) < +\infty$? Of course, this last assertion is not true for all sequences: take the sequence of binary digits of a normal real number. The complexity of this sequence is $\text{fac}(k) = 2^k$ and its palindrome complexity is $\text{pal}(k) = 2^{\lfloor (k+1)/2 \rfloor}$. This is not true for the Rudin–Shapiro sequence (Theorem 4) nor for the sequence given in the third item of Remark 6. For this question see also Remark 9 below.
- Is there a Pansiot-like theorem [45] for the palindrome complexity of fixed points of morphisms? Combining with the previous question, is it true that, if the palindrome complexity of a fixed point of a morphism is not ultimately 0, then it satisfies

$$0 < \limsup_{k \rightarrow \infty} (\text{pal}(k)/\varphi(k)) < +\infty,$$

where $\varphi(k)$ is either 1, $\log \log k$, $\log k$ or k ?

This question can be first experimentally addressed by looking at all palindromes of reasonable lengths that occur in fixed points of various morphisms, getting an idea of what the order of magnitude of the palindrome complexity seems to be, and ... proving it. We give two examples.

- we know that the complexity of the (infinite) fixed point u of the morphism $0 \rightarrow 001$, $1 \rightarrow 1$ satisfies $C_1 k^2 \leq \text{fac}_u(k) \leq C_2 k^2$ [45]. If we look for *maximal palindromes* in the sequence, i.e., factors of the sequence that are palindromes and that cannot be extended to longer palindromes occurring in the sequence, we find they are given by the recurrence $w_0 = 0$, $w_{m+1} = 1.\sigma(w_m)$. (Note that by “the palindrome w can be extended to a palindrome occurring in the sequence” we mean of course that there exists a letter a such that awa occurs in the sequence.) It can be proved that $\text{pal}(k)$ is the number of integers m such that $m \leq k + 1$,

$|w_m| = 2^{m+1} + m - 1 \geq k$, and $m - k$ is odd, which implies $\text{pal}(k) = k/2 - 1/2 \log_2(k) + O(1)$.

- In the same vein, the complexity of the fixed point beginning in 0 of $0 \rightarrow 010$, $1 \rightarrow 11$ satisfies $\text{fac}(k) \sim k \log_2 \log_2(k)$ (see [15]). The behavior of the palindrome complexity of this sequence is different for odd and even length: $\text{pal}(2n + 1) = 1$ for $n \geq 4$, whereas $\text{pal}(2n) \sim \log_2 \log_2(2n)$.

Before we state the main theorem of this section, we need a definition and a preliminary result.

Definition 9. Let w be a palindrome on the alphabet \mathcal{A} , and let T be its period.

- If $T > |w|/2$, the palindrome w is called *non-periodic*.
- If $T \leq |w|/2$ and T is odd, the palindrome w is called *a palindrome of odd period*.
- If $T \leq |w|/2$ and T is even, the palindrome w is called *a palindrome of even period*.

Remark 8. In this paper we thus call *periodic* a palindrome whose period (even or odd) is at most half of the length of the palindrome. This means, in particular, that a periodic palindrome is a palindrome that can a priori be written as $A^d B$, where B is a prefix of A with $|B| < |A|$, and necessarily $d \geq 2$. It is not hard to see that the word w is a periodic palindrome if and only if there exist two palindromes B and C , and an integer $d \geq 2$, such that $w = (BC)^d B$ (in the decomposition $w = A^d B$, remember that B is a prefix of A , put $A = BC$, and compare the prefixes of length $|B| + |C|$ of w and \tilde{w}).

Lemma 2. Let w be a palindrome of even period T . The word w is hence a prefix of an infinite sequence $(xyxyxy\dots)$ of smallest period T , with $|x| = |y| = T/2 \leq |w|/4$. The prefix of length $|w|$ of the sequence $(xyxyxy\dots)$ is denoted by w^θ and called the twin of w . Then

- the word w^θ is a periodic palindrome and its period is T . Furthermore $w^\theta \neq w$;
- “taking the twin” is an involution. More precisely the map $w \rightarrow w^\theta$ is an involution on the set of palindromes of even period.

Proof. It is clear that w and w^θ have the same period, and that the map $w \rightarrow w^\theta$ is an involution. Note that $w \neq w^\theta$, otherwise $x = y$ and $T/2$ would be a period of w .

It remains to prove that w^θ is a palindrome. Let us write $w = (xy)^d z$, with $|x| = |y| = T/2$, $d \geq 2$, and z prefix of xy with $|z| < |xy|$. Write $xy = zt$ (with t non-empty). As in Remark 8, the words z and t must be palindromes. Hence $tz = \tilde{t}\tilde{z} = \tilde{z}\tilde{t} = \tilde{x}\tilde{y} = \tilde{y}\tilde{x}$. Hence the word $yz\tilde{y}$, which is a prefix of $yz\tilde{y}\tilde{x}$, is a prefix of $yztz = yx yz$, hence a prefix of $yx yx y$. This shows that the word $y(xy)^{d-1} z \tilde{y} = (yx)^{d-1} yz \tilde{y}$ is a prefix of the sequence $(yx yx yx \dots)$. Since the length of $y(xy)^{d-1} z \tilde{y}$ is equal to $|w|$, we thus have an alternative definition of w^θ :

if w is a palindrome of even period T ($T \leq |w|/2$), with $w = (xy)^d z$, where $|x| = |y| = T/2$, $d \geq 2$, z prefix of xy , and $|z| < |xy|$, then

$$w^\theta = y(xy)^{d-1}z\tilde{y}.$$

Now, with the notations above, and remembering that z and t are palindromes,

$$\begin{aligned} \widetilde{w^\theta} &= yz(\widetilde{xy})^{d-1}\tilde{y} = yz(\widetilde{zt})^{d-1}\tilde{y} \\ &= yz(tz)^{d-1}\tilde{y} = y(zt)^{d-1}z\tilde{y} = y(xy)^{d-1}z\tilde{y} = w^\theta. \quad \square \end{aligned}$$

We now give a theorem that bounds the palindrome complexity in terms of the usual complexity, and that answers positively the first question at the beginning of this section.

Theorem 12. *Let $u = u_0u_1u_2\dots$ be an infinite non-ultimately periodic sequence on a finite alphabet. Then, for all $k \geq 1$, we have*

$$\text{pal}_u(k) < \frac{16}{k} \text{fac}_u \left(k + \left\lfloor \frac{k}{4} \right\rfloor \right).$$

Proof. We first suppose that the sequence u is recurrent or that it is indexed by \mathbb{Z} ($u = \dots u_{-2}u_{-1}u_0u_1u_2\dots$). In the latter case we suppose the sequence is not ultimately periodic “on the right”. Let $k \geq 1$ be a fixed integer. We split $\text{Pal}_u(k)$ the set of palindromes of length k occurring in the sequence u into three classes according to their periods T :

$$\begin{aligned} \text{Pal}_u^{(0)}(k) &:= \{w \in \text{Pal}_u(k), T > k/2\} \\ &\text{(this is the set of non-periodic palindromes of length} \\ &\quad k \text{ occurring in } u), \end{aligned}$$

$$\begin{aligned} \text{Pal}_u^{(1)}(k) &:= \{w \in \text{Pal}_u(k), T \leq k/2 \text{ and } T \text{ odd}\} \\ &\text{(this is the set of palindromes of length } k \text{ and of odd} \\ &\quad \text{period occurring in } u), \end{aligned}$$

$$\begin{aligned} \text{Pal}_u^{(2)}(k) &:= \{w \in \text{Pal}_u(k), T \leq k/2 \text{ and } T \text{ even}\} \\ &\text{(this is the set of palindromes of length } k \text{ and of even} \\ &\quad \text{period occurring in } u). \end{aligned}$$

- For each $w \in \text{Pal}_u^{(0)}(k)$, i.e., for each non-periodic palindrome w of length k in the sequence u we choose an index ℓ (assuming furthermore, which is possible, that $\ell > \lfloor k/4 \rfloor$ if the sequence is recurrent) such that an occurrence of w in u begins at index ℓ . Then, we associate with w the language $S(w)$ that consists of the $(\lfloor k/4 \rfloor + 1)$ words of length $(k + \lfloor k/4 \rfloor)$ that occur in u and begin at indexes between $(\ell - \lfloor k/4 \rfloor)$ and ℓ . (Note that these indexes are non-negative in the case where the sequence is

recurrent, and well defined if the sequence is indexed by \mathbb{Z} .) These words are all distinct; namely if the words beginning at indexes $\ell - i$ and $\ell - j$ (where $0 \leq i, j \leq \lfloor k/4 \rfloor$) were equal, the word w would have a period $\leq |i - j| \leq k/4 \leq k/2$ (use Theorem 2), which contradicts the non-periodicity of w . We thus have

$$\forall w \in \text{Pal}_u^{(0)}(k), \quad \#(S(w)) = \lfloor k/4 \rfloor + 1.$$

- For each $w \in \text{Pal}_u^{(1)}(k)$, i.e., for each palindrome w of length k occurring in the sequence u and having an odd period T , we choose an index ℓ (assuming furthermore, which is possible, that $\ell > \lfloor k/4 \rfloor$ if the sequence is recurrent) such that an occurrence of w in u begins at index ℓ , and such that the word of length $|w|$ beginning at index $\ell + T$ is different from w (this is possible since the sequence u is not ultimately periodic). We define $S(w)$ as above. The words in $S(w)$ are pairwise distinct: if the words of $S(w)$ beginning at $\ell - i$ and $\ell - j$, say z_i and z_j , (where $0 \leq i, j \leq \lfloor k/4 \rfloor$) were equal, then there would exist two words A and B of length $|j - i|$ such that $Az = zB$, where $z = z_i = z_j$. From Theorem 2 this implies the existence of two words γ and δ and of an integer $e \geq 0$ such that $A = \gamma\delta$, $B = \delta\gamma$, and $z = (\gamma\delta)^e\gamma$. Hence $Az = zB$ is a prefix of a periodic sequence of period $|\gamma\delta| = |j - i|$. Since w is a factor of $Az = zB$ (this is the factor of length k beginning at index ℓ in the sequence u , and $Az = zB$ is the factor of length $k + \lfloor k/4 \rfloor$ beginning at index $\ell - \max(i, j)$ in the sequence u), this implies from Lemma 1 that Az has period T (note that $|j - i| + T \leq j + i + k/2 \leq 2\lfloor k/4 \rfloor + k/2 \leq k = |w|$). Now, the factor of the sequence u of length k beginning at index $\ell + T$ is a factor of Az since $T \leq |j - i|$ (namely $|j - i|$ is a period of $Az = zB$, which contains w , hence $|j - i|$ is a period of w , and by Lemma 1 it is a multiple of T , so $T \leq |j - i|$) hence is equal to the factor of length k beginning at ℓ , i.e., to w , which gives a contradiction.
- We consider now the palindromes of even period. For each $w \in \text{Pal}_u^{(2)}(k)$, i.e., for each palindrome w of length k and of even period T (remember that means in particular that $T \leq k/2$), we consider its twin w^θ (see Lemma 2). Since the map $w \rightarrow w^\theta$ is involutive on the set of palindromes of even period, and $w^\theta \neq w$ (Lemma 2), palindromes of even period can be grouped in pairs of twins. For each pair of twins (w, w^θ) of length k and (even) period T , such that at least one of w, w^θ occurs in the sequence u , we choose an index ℓ (and we can suppose $\ell \geq \lfloor k/4 \rfloor$ in the case where u is recurrent) such that one of the words w, w^θ occurs in u at index ℓ , and such that its twin does not occur at index $\ell + T/2$, this is possible since the sequence u is not ultimately periodic: if it were true that for every index ℓ such that the word w (resp. w^θ) occurs in the sequence u at index ℓ , then the word w^θ (resp. w) would occur in the sequence u at index $\ell + T/2$, then both words w and w^θ would occur in u , and, for every index ℓ such that w occurs at ℓ , w would also occur at index $\ell + T$, which would imply that u is ultimately periodic. We construct the set $S(w)$ (or $S(w^\theta)$) as above, and the words in $S(w)$ (or $S(w^\theta)$) are again pairwise distinct.

We thus constructed at least $\#\text{Pal}_u^{(0)}(k) + \#\text{Pal}_u^{(1)}(k) + \#\text{Pal}_u^{(2)}(k)/2 \geq \text{pal}_u(k)/2$ languages, each of which contains $\lfloor k/4 \rfloor + 1$ words of length $k + \lfloor k/4 \rfloor$ that are distinct. We

prove now that these sets are pairwise disjoint. If there exists a word z belonging to $S(w_1) \cap S(w_2)$ (hence $|z| = k + \lfloor k/4 \rfloor$), then there exist indexes ℓ_1 and ℓ_2 , non-negative integers i_1 and i_2 both $\leq \lfloor k/4 \rfloor$, such that w_1 occurs at index ℓ_1 in the sequence u , w_2 occurs at index ℓ_2 in u , and z occurs in u both at indexes $\ell_1 - i_1$ and $\ell_2 - i_2$. Then w_1 occurs at position i_1 in z and w_2 at position i_2 in z . Let $h := i_2 - i_1$ (assuming $i_2 \geq i_1$). If $h = 0$, then $w_1 = w_2$. Otherwise we have $h > 0$. Then the prefix of length $k - h$ of w_2 is equal to the suffix of length $k - h$ of w_1 . We write $w_1 = At$ and $w_2 = tB$, where $|t| = k - h$ and $|A| = |B| = h$. Now let $z' := AtB$ be the factor of z of length $k + h$ obtained by “superposing” w_1 and w_2 . Since w_1 and w_2 are both palindromes, we have

$$(A\tilde{B})\tilde{t} = A(\tilde{B}\tilde{t}) = A(tB) = z' = (At)B = (\tilde{t}\tilde{A})B = \tilde{t}(\tilde{A}B).$$

Hence, from Theorem 2 there exist two words C and D and an integer $q \geq 0$ such that $\tilde{t} = (CD)^q C$, $A\tilde{B} = CD$, and $\tilde{A}B = DC$. Hence $z' = (CD)^{q+1} C$, and $|CD| = |A\tilde{B}| = 2h$ is a period of z' . In particular $2h$ is a period of w_1 . Since $2h \leq 2i_2 \leq 2\lfloor k/4 \rfloor \leq k/2$, the word w_1 is a periodic palindrome whose period, say T , divides $2h$: this is a consequence of Lemma 1 which also gives that T is a period of z' and hence of w_2 .

- If T divides h , then the equality $z' = A\tilde{B}\tilde{t}$ shows that $A = \tilde{B}$ (remember that $|A| = |B| = h$). Hence $w_1 = At$ and $w_2 = tB = \tilde{B}\tilde{t}$ coincide on their prefixes of length $|A| = h$, hence $w_1 = w_2$.
- If T does not divide h , let $A\tilde{B} = E^r$, with $|E| = T$ and $r \geq 1$. Since $2h = |A\tilde{B}| = |E^r| = rT$ and T does not divide h , T must be even, and r must be odd. Let $E = xy$, with $|x| = |y| = T/2$. We have $A\tilde{B} = (xy)^r$, which implies, since r is odd, that $A = (xy)^{(r-1)/2}x$ and $\tilde{B} = y(xy)^{(r-1)/2}$. The word xy is a prefix of $A\tilde{B}$, hence of $A\tilde{B}\tilde{t} = z'$. Since w_1 is a prefix of z' , and since $|w_1| \geq 2T = 2|x y|$, we see that xy is a prefix of w_1 . Finally w_1 is the prefix of length k of the sequence $(xyxyxy\dots)$ and the smallest period of this sequence is T .

Now $\tilde{B}\tilde{t}D = \tilde{B}(CD)^{q+1} = y(xy)^{r(q+1)+(r-1)/2}$. Hence yx is a prefix of $\tilde{B}\tilde{t}D$. Since $w_2 = \tilde{B}\tilde{t}$ is a prefix of $\tilde{B}\tilde{t}D$, and $|w_2| \geq 2|x y|$, the word yx is a prefix of w_2 . Now $T = |xy|$ is a period of w_2 , hence we see that w_2 is the prefix of length k of the sequence $(yxxyxy\dots)$. Since the smallest period of this sequence must be T (as for the sequence $(xyxyxy\dots)$), w_2 is the twin of w_1 , which is a contradiction, since we chose exactly one representative in each pair of twins.

We thus have

$$\frac{\text{pal}_u(k)}{2} (\lfloor k/4 \rfloor + 1) \leq \text{fac}_u(k + \lfloor k/4 \rfloor).$$

Since $\lfloor k/4 \rfloor + 1 > k/4$, we have that

$$\text{pal}_u(k) < \frac{8}{k} \text{fac}_u \left(k + \left\lfloor \frac{k}{4} \right\rfloor \right).$$

Now, if we take a non-ultimately periodic sequence $u = (u_n)_{n \geq 0}$ on an alphabet \mathcal{A} and if u is not recurrent, let ω be a letter that does not belong to \mathcal{A} . Define the

Note that this is true for sequences u with maximal complexity, or even for sequences with complexity $\Theta(\alpha^k k^{-a})$ for some $\alpha > 1$ and $a \geq 0$ (use the easy bound $\text{pal}_u(k) \leq \text{fac}_u(\lfloor (k+1)/2 \rfloor)$). This is also true from Theorem 12 above if $\text{fac}_u(k) = O(k^{3/2})$, or if $\text{fac}_u(k) = \Theta(k^2)$. Hence, using Pansiot's theorem [45], this is true for fixed points of non-trivial morphisms.

8. On a question of Hof, Knill, and Simon

To end this paper, we recall a question of Hof et al. [32, Remark 3, p. 153]: *are there (minimal) sequences containing arbitrarily long palindromes that arise from substitutions none of which belongs to class P?* This question is still open. But we prove in Theorem 13 below that we can restrict ourselves to particular substitutions in class P, as well as to nonperiodic (minimal) sequences. We first prove the following result.

Lemma 3. *Let u be a fixed point of a primitive morphism σ on the alphabet \mathcal{A} . Suppose that there exists a non-empty word x that is a prefix of $\sigma(a)$ for all $a \in \mathcal{A}$ (resp. a suffix of $\sigma(a)$ for all $a \in \mathcal{A}$). Write $\sigma(a) = xz_a$ (resp. $\sigma(a) = z_a x$). Let $\sigma_{\#}$ be the morphism defined by $\sigma_{\#}(a) = z_a x$ (resp. $\sigma_{\#}(a) = x z_a$). Then $\sigma_{\#}$ is primitive, and any fixed point v of a power of $\sigma_{\#}$ (there always exists at least one such fixed point) has the same factors as u .*

Proof. The morphism $\sigma_{\#}$ is clearly primitive. Suppose now, for example, that $\sigma(a) = xz_a$ for all $a \in \mathcal{A}$. Then, for all $a \in \mathcal{A}$ we have $x\sigma_{\#}(a) = \sigma(a)x$. Hence, for any word w on \mathcal{A} , we have $x\sigma_{\#}(w) = \sigma(w)x$. Taking for w prefixes of increasing length of u , we get a limit when this length goes to infinity: $x\sigma_{\#}(u) = \sigma(u) = u$. Hence, $x\sigma_{\#}(x)\sigma_{\#}^2(x)\dots = u$. From this equality we see that, calling α the first letter of x (remember that x is not empty), any factor of u is a factor of $\sigma_{\#}^j(\alpha)$ for $j \geq j_0$ (remember that u is minimal, since σ is primitive). Trivially any factor of $\sigma_{\#}^j(\alpha)$ for any j is a factor of u . Now there exist integers $\ell > 0$ such that $\sigma_{\#}^{\ell}$ admits a fixed point, say v . For any such ℓ it is clear from what precedes that v and u have the same factors. The case where x is a suffix of $\sigma(a)$ for all a is similar. \square

Theorem 13.

- *Let u be a sequence that is a fixed point of a primitive morphism in class P. Then the set of factors of u is the same as the set of factors of a sequence $u_{\#}$ that is a fixed point of a primitive morphism in class P where the palindrome p occurring in Definition 8 has length 0 or 1. We can even choose one of the two forms $a \rightarrow pq_a$ for all a , or $a \rightarrow q_a p$ for all a .*
- *Let u be a periodic sequence that contains arbitrarily long palindromes, then u is a fixed point of a morphism in class P.*

Proof. Let u be a sequence that is a fixed point of the primitive morphism σ on the alphabet \mathcal{A} . Suppose there exists a palindrome p and, for any $a \in \mathcal{A}$ a palindrome q_a , with $\sigma(a) = pq_a$.

If p is empty, the first assertion of Theorem 13 is satisfied.

If p has even length, let $p = r\tilde{r}$ for some word r . Define the morphism $\sigma_{\#}$ by $\sigma_{\#}(a) := \tilde{r}q_ar$ for all $a \in \mathcal{A}$. Applying Lemma 3 with $x = r$, we know that this morphism is primitive; furthermore there exists an integer $\ell > 0$ such that $\sigma_{\#}^{\ell}$ admits a fixed point, say v , and the sequences v and u have the same factors. Finally, the definition of $\sigma_{\#}$ shows that the image of any letter is a palindrome, hence that the image of any palindrome is also a palindrome. This proves that the image of any letter by $\sigma_{\#}^{\ell}$ is a palindrome. In other words, $\sigma_{\#}^{\ell}$ belongs to class P, and the corresponding palindrome p is empty.

If p has odd length, write $p = rb\tilde{r}$ for some letter b and some word r . Define, for all $a \in \mathcal{A}$, $\sigma_{\#}$ by $\sigma_{\#}(a) := b\tilde{r}q_ar$. Applying Lemma 3 with $x = r$, we mimic the proof just above in the case where the length of p is even, leading to a morphism in class P, whose corresponding palindrome p has length 1.

Finally, if the morphism has, for example, the form $a \rightarrow pq_a$ for all a , we can apply Lemma 3 with $x = p$ to obtain a morphism of the form $a \rightarrow q_ap$ for all a .

We now prove that the answer to the question of Hof, Knill and Simon is negative for a periodic sequence (that is a fixed point of a primitive morphism). Let $u = www\dots$ be a periodic sequence that contains arbitrarily long palindromes. Let s be a palindromic factor of u such that $|s| \geq 2|w|$. We can write $s = xw^ky$, where $k \geq 1$, and $0 \leq |x|, |y| < |w|$. Since s is a palindrome, we have $s = \tilde{y}(\tilde{w})^k\tilde{x}$. Hence \tilde{w} is a factor of s , hence a factor of u . Thus \tilde{w} must be a factor of ww . Let $ww = A\tilde{w}B$. Since $|\tilde{w}| = |w|$, we see that $|w| = |A| + |B|$, then we must have $w = AB$ (A is a prefix of w and B is a suffix of w). The equality $ww = A\tilde{w}B$ then implies that $\tilde{w} = BA$. Hence $w = A\tilde{B}$, which shows that A and B are palindromes. We conclude by noting that the sequence $u = ww\dots$ is a fixed point of the morphism τ defined by, for all $a \in \mathcal{A}$, $\tau(a) := w = AB$ that is in class P. \square

Note added in proof (October 2002). A. Frid has generalized our Theorem 8 above by showing that the palindrome complexity of the (Rote) sequences $(w_n)_{n \geq 0}$ defined by

$$w_n := \begin{cases} 0, & \text{if } (c + n\theta) \bmod 1 \in [0, \varphi) \\ 1, & \text{if } (c + n\theta) \bmod 1 \in [\varphi, 1) \end{cases}$$

(where c, φ, θ are real numbers with $0 < \varphi < 1, 0 < \theta < \min(\varphi, 1 - \varphi)$ and θ is irrational) is constant and equal to 2.

Acknowledgements

JPA, MB, and DD began to discuss together on palindromes during the summer school and workshop “The Mathematics of Aperiodic Order” at the University of

Alberta at Edmonton. They want to thank very warmly R. V. Moody for having organized this conference and having permitted many people from different fields to confront questions and knowledges. The second author thanks the German Research Council (DFG) for its support. The fourth author thanks the German Academic Exchange Service for financial support through Hochschulsonderprogramm III (Postdoktoranden). Finally we would like to thank A. Rodenhausen for her questions on the Kolakoski sequence, S. Brlek for his helpful comments on this sequence, and V. Berthé for several useful comments on a previous version of this paper.

References

- [1] P. Alessandri, V. Berthé, Three distance theorems and combinatorics on words, *Enseign. Math.* 44 (1998) 103–132.
- [2] J.-P. Allouche, The number of factors in a paperfolding sequence, *Bull. Austral. Math. Soc.* 46 (1992) 23–32.
- [3] J.-P. Allouche, Sur la complexité des suites infinies, *Bull. Belg. Math. Soc.* (1) (1994) 133–143.
- [4] J.-P. Allouche, Schrödinger operators with Rudin–Shapiro potentials are not palindromic, *J. Math. Phys.* (Special Issue “Quantum Problems in Condensed Matter Physics”) 38 (1997) 1843–1848.
- [5] J.-P. Allouche, M. Bousquet-Mélou, Facteurs des suites de Rudin–Shapiro généralisées, *Bull. Belg. Math. Soc.* 1 (1994) 145–164.
- [6] J.-P. Allouche, P. Liardet, Generalized Rudin–Shapiro sequences, *Acta Arith.* 60 (1991) 1–27.
- [7] J.-P. Allouche, J. Shallit, The ring of k -regular sequences, *Theoret. Comput. Sci.* 98 (1992) 163–187.
- [8] J.-P. Allouche, J. Shallit, Sums of digits, overlaps, and palindromes, *Discrete Math. Theoret. Comput. Sci.* 4 (2000) 1–10. <http://dmtcs.loria.fr/volumes/abstracts/dm040101.abs.html>.
- [9] J.-P. Allouche, J. Shallit, The ring of k -regular sequences, II, submitted, 2002.
- [10] M. Baake, A note on palindromicity, *Lett. Math. Phys.* 49 (1999) 217–227; math-ph/9907011.
- [11] J. Berstel, Recent results in Sturmian words, in: J. Dassow, G. Rozenberg, A. Salomaa (Eds.), *Developments in Language Theory II*, World Scientific, Singapore, 1996, pp. 13–24.
- [12] J. Berstel, P. Séébold, Sturmian words, in: M. Lothaire (Ed.), *Algebraic Combinatorics on Words*, Cambridge Univ. Press, 2002, Chapter 3, preprint version available at: <http://www-igm.univ-mlv.fr/~berstel/Lothaire/index.html>.
- [13] V. Berthé, L. Vuillon, Palindromes and two-dimensional Sturmian sequences, *J. Automat. Lang. Combin.* to appear.
- [14] J. Cassaigne, Special factors of sequences with linear subword complexity, in: J. Dassow, G. Rozenberg, A. Salomaa (Eds.), *Developments in Language Theory, II*, World Scientific, Singapore, 1996, pp. 25–34.
- [15] J. Cassaigne, Complexité et facteurs spéciaux, *Bull. Belg. Math. Soc.* 4 (1997) 67–88.
- [16] C. Choffrut, private communication, January 2001.
- [17] A. Cobham, Uniform tag sequences, *Math. Systems Theory* 6 (1972) 164–192.
- [18] E.M. Coven, G.A. Hedlund, Sequences with minimal block growth, *Math. Systems Theory* 7 (1973) 138–153.
- [19] D. Damanik, Local symmetries in the period-doubling sequence, *Discrete Appl. Math.* 100 (2000) 115–121.
- [20] D. Damanik, Gordon-type arguments in the spectral theory of one-dimensional quasicrystals, in: M. Baake, R.V. Moody (Eds.), *Directions in Mathematical Quasicrystals*, CRM Monograph Series, vol. 13, AMS, Providence, RI, 2000, pp. 277–304; math-ph/9912005.
- [21] D. Damanik, D. Zare, Palindrome complexity bounds for primitive substitution sequences, *Discrete Math.* 222 (2000) 259–267.
- [22] F.M. Dekking, What is the long range order in the Kolakoski sequence?, in: R.V. Moody (Ed.), *The Mathematics of Long-Range Aperiodic Order*, NATO ASI Series, Series C., Mathematics of Physical Science, vol. 489, Kluwer, Dordrecht, 1997, pp. 115–125.
- [23] M. Dekking, M. Mendès France, A.J. van der Poorten, FOLDS!, *Math. Intell.* 4 (1982) 130–138, 173–181, 190–195.

- [24] X. Droubay, Palindromes in the Fibonacci word, *Inform. Process. Lett.* 55 (1995) 217–221.
- [25] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.* 255 (2001) 539–553.
- [26] X. Droubay, G. Pirillo, Palindromes and Sturmian words, *Theoret. Comput. Sci.* 223 (1999) 73–85.
- [27] A. Ehrenfeucht, K.P. Lee, G. Rozenberg, Subword complexities of various classes of deterministic developmental languages without interaction, *Theoret. Comput. Sci.* 1 (1975) 59–75.
- [28] S. Ferenczi, Les transformations de Chacon: combinatoire, structure géométrique, lien avec les systèmes de complexité $2n + 1$, *Bull. Soc. Math. France* 123 (1995) 271–292.
- [29] S. Ferenczi, Complexity of sequences and dynamical systems, *Discrete Math.* 206 (1999) 145–154.
- [30] S. Ferenczi, Z. Kása, Complexity for finite factors of infinite sequences, *Theoret. Comput. Sci.* 218 (1999) 177–195.
- [31] N.J. Fine, H.S. Wilf, Uniqueness theorems for periodic functions, *Proc. Amer. Math. Soc.* 16 (1965) 109–114.
- [32] A. Hof, O. Knill, B. Simon, Singular continuous spectrum for palindromic Schrödinger operators, *Commun. Math. Phys.* 174 (1995) 149–159.
- [33] J. Justin, G. Pirillo, Episturmian words and episturmian morphisms, *Theoret. Comput. Sci.* 276 (2002) 281–313.
- [34] W. Kolakoski, Elementary problem 5304, *Amer. Math. Monthly* 72 (1965) 674; Solution 73 (1966) 681–682.
- [35] A. Ladouceur, Outil logiciel pour la combinatoire des mots, Mémoire de Maîtrise, UQAM, Montréal, 1999.
- [36] W.F. Lunnon, P.A.B. Pleasants, Characterization of two-distance sequences, *J. Austral. Math. Soc. Ser. A* 53 (1992) 198–218.
- [37] R.C. Lyndon, M.P. Schützenberger, The equation $a^M = b^N c^P$ in a free group, *Michigan Math. J.* 9 (1962) 289–298.
- [38] M. Mendès France, G. Tenenbaum, Dimension des courbes planes, papiers pliés et suites de Rudin–Shapiro, *Bull. Soc. Math. France* 109 (1981) 207–215.
- [39] P. Michel, Sur les ensembles minimaux engendrés par les substitutions de longueur non constante, Thèse, Université de Rennes, 1975.
- [40] P. Michel, Stricte ergodicité d’ensembles minimaux de substitution, in: J.-P. Conze, M.S. Keane (Eds.), *Théorie Ergodique, Actes des Journées Ergodiques, Rennes 1973/1974, Lecture Notes in Mathematics*, vol. 532, Springer, Berlin, 1976, pp. 189–201.
- [41] M. Morse, G.A. Hedlund, Symbolic dynamics, II, Sturmian trajectories, *Amer. J. Math.* 62 (1940) 1–42.
- [42] B. Mossé, Puissances de mots et reconnaissabilité des points fixes d’une substitution, *Theoret. Comput. Sci.* 99 (1992) 327–334.
- [43] B. Mossé, Reconnaisabilité des substitutions et complexité des suites automatiques, *Bull. Soc. Math. France* 124 (1996) 329–346.
- [44] M. Queffélec, Une nouvelle propriété des suites de Rudin–Shapiro, *Ann. Inst. Fourier* 37 (1987) 115–138.
- [45] J.-J. Pansiot, Complexité des facteurs des mots infinis engendrés par morphismes itérés, in: *Automata, Languages and Programming, Antwerp, 1984, Lecture Notes in Computer Science*, vol. 172, Springer, Berlin, 1984, pp. 380–389.
- [46] G. Rote, Sequences with subword complexity $2n$, *J. Number Theory* 46 (1994) 196–213.
- [47] M. Schlottmann, Generalized model sets and dynamical systems, in: M. Baake, R.V. Moody (Eds.), *Directions in Mathematical Quasicrystals, CRM Monograph Series*, vol. 13, AMS, Providence, RI, 2000 pp. 143–159.