

## ON EXTREMAL PROPERTIES OF THE FIBONACCI WORD

JULIEN CASSAIGNE<sup>1</sup>

**Abstract.** We survey several quantitative problems on infinite words related to repetitions, recurrence, and palindromes, for which the Fibonacci word often exhibits extremal behaviour.

**Mathematics Subject Classification.** 68R15.

### 1. INTRODUCTION

The *Fibonacci infinite word*,

$$\mathbf{f} = abaababaabaababaababaababaababaababaababaababaababa \dots$$

is certainly one of the most often cited examples in the combinatorial theory of infinite words. It is the archetype of a *Sturmian word*, and also the fixed point of a very simple substitution, the *Fibonacci substitution*  $\varphi: a \mapsto ab, b \mapsto a$ .

In many situations, the Fibonacci word happens to have the “best possible” properties, in the sense that some quantity is maximal or minimal for this word. In this paper, we present several such situations, and also a few where the Fibonacci word happens not to be optimal. We consider three different classes of problems: first, problems related to repetition of words; then, problems related to the notion of recurrence; finally, problems involving palindromes.

Throughout the paper,  $A$  is an arbitrary finite alphabet, and  $B = \{a, b\}$  is the binary alphabet. The Fibonacci word is the unique infinite word in  $B^{\mathbb{N}}$  fixed by the substitution  $\varphi$  on  $B$  defined above.

We denote by  $(F_n)$  the classical sequence of *Fibonacci numbers*, with  $F_0 = 0$ ,  $F_1 = 1$  and  $F_{n+1} = F_n + F_{n-1}$ , so that the length of the  $n$ -th iterate of the Fibonacci substitution is  $|\varphi^n(a)| = F_{n+2}$ . The *golden ratio* is denoted  $\Phi = \frac{1+\sqrt{5}}{2}$ .

---

*Keywords and phrases.* Fibonacci word, repetitions, recurrence function, palindromes.

<sup>1</sup> Institut de mathématiques de Luminy, case 907, 163 avenue de Luminy, 13288 Marseille Cedex 9, France; [cassaigne@iml.univ-mrs.fr](mailto:cassaigne@iml.univ-mrs.fr)

Given an infinite word  $\mathbf{u} \in A^{\mathbb{N}}$ ,  $L_n(\mathbf{u})$  denotes the language of factors of length  $n$  of  $\mathbf{u}$  (i.e., finite words of length  $n$  that occur as a block of consecutive letters of  $\mathbf{u}$ ) and  $L(\mathbf{u})$  the language of all factors of  $\mathbf{u}$ .

Let  $\alpha \in [0, 1] \setminus \mathbb{Q}$ . A word  $\mathbf{u} \in B^{\mathbb{N}}$  is called a *Sturmian word of slope  $\alpha$*  when there exists  $\beta \in \mathbb{R}$  such that one of the following holds:

- (i) for all  $n \in \mathbb{N}$ ,  $u_n = a$  if and only if  $\lfloor \alpha n + \beta \rfloor = \lfloor \alpha(n+1) + \beta \rfloor$ ;
- (ii) for all  $n \in \mathbb{N}$ ,  $u_n = a$  if and only if  $\lceil \alpha n + \beta \rceil = \lceil \alpha(n+1) + \beta \rceil$ .

The letter  $b$  occurs then in  $\mathbf{u}$  with frequency  $\alpha$ . The Fibonacci word is a Sturmian word of slope  $2 - \Phi$ . (There are many alternative definitions for Sturmian words, see [18], Chap. 2.)

## 2. REPETITIONS

### 2.1. INDEX

Let  $\mathbf{u} = u_0u_1u_2\dots \in A^{\mathbb{N}}$  be an infinite word.

The *exponent*  $e(w)$  of a word  $w \in A^*$  is the maximum of  $|w|/|z|$  over all words  $z \in A^+$  such that  $w$  is a prefix of  $z^\omega$ . Equivalently,  $e(w) = |w|/(|w| - |x|)$ , where  $x$  is the *maximal border* of  $w$ , i.e., the longest word that is both a proper prefix and a proper suffix of  $w$ . If  $e(w) > 1$ , then  $w$  is called a *repetition* of exponent  $e(w)$  and of period  $z$ .

The *index* (or *critical exponent*) of  $\mathbf{u}$  is the supremum of exponents of repetitions that occur in  $\mathbf{u}$ :

$$\text{ind}(\mathbf{u}) = \sup\{e(w) : w \in L(\mathbf{u})\} \in (1, +\infty].$$

Periodic words, among others, have infinite index, and it is not difficult to construct words with arbitrarily big but finite index. On the other hand, on a given alphabet there is a lowest possible index, and finding that index is a problem known as Dejean's conjecture [14]. Currently, it is solved for alphabets of size  $k \leq 14$  [12,24,25] as well as for alphabets of size  $k \geq 33$  [6]; for the remaining cases, it is conjectured to be equal to  $k/(k-1)$ .

On a binary alphabet, the lowest possible index is 2, as is well known since the work of Thue [28], and the standard example of a word with index 2 is the Prouhet-Thue-Morse word (see Sect. 6 for more on this word).

Here, the Fibonacci word is far from optimal since  $\text{ind}(\mathbf{f}) = \Phi + 2 \simeq 3.618$  [19]. For instance, the cube  $(aba)^3$  occurs in  $\mathbf{f}$  at position 5.

However, it is optimal among Sturmian words. A general formula for the index of a Sturmian word was given independently by Carpi and de Luca [7] and by Damanik and Lenz [13] (see also [3]):

**Theorem 2.1.** *If  $\mathbf{u}$  is a Sturmian word of slope  $\alpha = [0; a_1, a_2, a_3, \dots]$ , then*

$$\text{ind}(\mathbf{u}) = \sup_{n \geq 0} \left( 2 + a_{n+1} + \frac{q_{n-1} - 2}{q_n} \right),$$

where  $q_n$  is the denominator of  $[0; a_1, a_2, a_3, \dots, a_n]$  and satisfies  $q_{-1} = 0, q_0 = 1, q_{n+1} = a_{n+1}q_n + q_{n-1}$ .

From this theorem, we first recover the fact that  $\text{ind}(\mathbf{f}) = \Phi + 2$ , since  $\mathbf{f}$  has slope  $2 - \Phi = [0; 2, 1, 1, \dots]$ . We also deduce that this is the smallest possible index for Sturmian words. Indeed, if the partial quotients are eventually 1, then  $\text{ind}(\mathbf{u}) \geq \lim_{n \rightarrow \infty} (3 + q_{n-1}/q_n) = \Phi + 2$ ; otherwise, choosing  $n$  such that  $q_{n-1} \geq 2$  and  $a_{n+1} \geq 2$ , we find that  $\text{ind}(\mathbf{u}) \geq 4$ .

However, the Fibonacci word and its subshift (*i.e.*, Sturmian words that share the same slope) are not the only Sturmian words that achieve the lowest possible index. Those were classified by Carpi and de Luca [7]:

**Theorem 2.2.** *Let  $\mathbf{u}$  be Sturmian word. Then  $\text{ind}(\mathbf{u}) = \Phi + 2$  if and only if the slope of  $\mathbf{u}$  is one of the six numbers*

$$\begin{array}{l|l} \frac{3-\Phi}{5} = [0; 3, 1, 1, 1, \dots] \simeq .276 & \frac{\Phi+2}{5} = [0; 1, 2, 1, 1, 1, \dots] \simeq .724 \\ 2 - \Phi = [0; 2, 1, 1, \dots] \simeq .382 & \Phi - 1 = [0; 1, 1, 1, \dots] \simeq .618 \\ \frac{\Phi+3}{11} = [0; 2, 2, 1, 1, 1, \dots] \simeq .420 & \frac{8-\Phi}{11} = [0; 1, 1, 2, 1, 1, 1, \dots] \simeq .580. \end{array}$$

Then  $\mathbf{u}$  is in the subshift generated respectively by  $h_1(\mathbf{f}), \mathbf{f}, h_2(\mathbf{f}), E(h_1(\mathbf{f})), E(\mathbf{f}),$  or  $E(h_2(\mathbf{f}))$ , where  $h_1(a) = aab, h_1(b) = a, h_2(a) = ababa, h_2(b) = ab, E(a) = b, E(b) = a$ .

Sturmian words of slope  $\frac{\Phi+4}{19}$  and  $\frac{15-\Phi}{19}$  have index  $\frac{11}{3} > \Phi + 2$ . All other Sturmian words have index at least 4.

2.2. LONG REPETITIONS

If we consider only arbitrarily long repetitions, we define the *asymptotic index*:

$$\text{ind}^*(\mathbf{u}) = \lim_{n \rightarrow \infty} \sup \{e(w) : w \in L(\mathbf{u}) \text{ and } |w| \geq n\} \in [1, +\infty].$$

Obviously,  $\text{ind}^*(\mathbf{u}) \leq \text{ind}(\mathbf{u})$ .

The asymptotic index for Sturmian words was computed by Vandeth [29] (actually, the theorem is stated there only for Sturmian words that are fixed points of substitutions, but it remains valid in general).

**Theorem 2.3.** *If  $\mathbf{u}$  is a Sturmian word of slope  $\alpha = [0; a_1, a_2, a_3, \dots]$ , then*

$$\text{ind}^*(\mathbf{u}) = 2 + \limsup_{n \rightarrow \infty} [a_n; a_{n-1}, a_{n-2}, \dots, a_1] \in [1, +\infty].$$

With  $\text{ind}^*(\mathbf{f}) = \Phi + 2$ , the Fibonacci word is again optimal among Sturmian words, as well as  $\sigma(\mathbf{f})$  for any Sturmian morphism  $\sigma$ .

We may wonder if the Prouhet-Thue-Morse word  $\mathbf{t}$  is still optimal among all binary infinite words. As any binary fixed point,  $\mathbf{t}$  has arbitrarily long squares, therefore  $\text{ind}^*(\mathbf{t}) = \text{ind}(\mathbf{t}) = 2$ . There exist binary words without long squares [16], but a word without long squares may still have asymptotic index 2. Actually, we found that asymptotic index 1 is achievable, and this is obviously optimal.

**Theorem 2.4.** *There exists a binary infinite word  $\mathbf{u}$  such that  $\text{ind}^*(\mathbf{u}) = 1$ .*

*Proof.* Define  $\mathbf{v} \in \{0, 1, \dots, 7\}^{\mathbb{N}}$  with:

$$\begin{cases} v_{2n} = 2n \pmod 8 \\ v_{2^{k+1}n+2^k-1} = 2\lfloor n/2^k \rfloor + 1 \pmod 8 \end{cases}$$

and let  $\mathbf{u} = \sigma(\mathbf{v})$ , where  $\sigma(i) = a^{8-i}b^{i+1}$ :

$$\mathbf{v} = 012141610321436105214561072147610123416103234361052345610\dots;$$

$$\mathbf{u} = aaaaaaaaaabaaaaaaaaabbbaaaaaaaaaabbbaabbbbaaaaaaaaaabbaab\dots$$

Then  $\text{ind}^*(\mathbf{u}) = \text{ind}^*(\mathbf{v}) = 1$ .

Indeed, if  $w = xyx \in L(\mathbf{v})$  and  $k = \lfloor \log_2 |x| \rfloor$ , then  $|xy| \geq 2^{2k+1} \geq \frac{1}{2}(|x| + 1)^2$  by Lemma 2.5 below. Therefore  $e(w) = 1 + O(1/\sqrt{|w|})$ , and  $\text{ind}^*(\mathbf{v}) = 1$ . By Lemma 2.6 below,  $\text{ind}^*(\mathbf{u}) = 1$  too.  $\square$

**Lemma 2.5.** *Let  $\mathbf{v}$  be defined as in Theorem 2.4,  $k \in \mathbb{N}$ , and  $x \in L(\mathbf{v})$  such that  $2^k \leq |x| < 2^{k+1}$ . Then there exists  $i \in \mathbb{Z}$  such that all occurrences of  $x$  in  $\mathbf{v}$  are at position  $i + 2^{2k+1}n$  for some  $n \in \mathbb{Z}$ .*

*Proof.* It is sufficient to consider the case when  $|x| = 2^k$ . The proof is by induction on  $k$ .

If  $k = 0$ , then  $x$  is a single digit and one can take  $i = x$ . Indeed, by construction, even digits  $x \in \{0, 2, 4, 6\}$  occur only at even positions in  $\mathbf{v}$ , whereas odd digits  $x \in \{1, 3, 5, 7\}$  occur only at odd positions.

Assume that the property holds for a given  $k$ , and consider the word  $x = x_0x_1\dots x_{2^k-1} \in L(\mathbf{v})$  with  $|x| = 2^k$ . By the induction hypothesis, there exists  $i \in \mathbb{Z}$  such that all occurrences of the prefix of length  $2^k$  of  $x$  in  $\mathbf{v}$  are at position  $i + 2^{2k+1}n$  for some  $n \in \mathbb{Z}$ . In particular, this applies to occurrences of  $x$ . Let  $j = 2^k - 1 - i \pmod{2^k}$  and  $m = (i + j - 2^k + 1)/2^{k+1}$ . If  $x$  occurs in  $\mathbf{v}$  at position  $i + 2^{2k+1}n$ , then  $x_j$  occurs at position  $i + 2^{2k+1}n + j = 2^k - 1 + 2^{k+1}(2^kn + m)$ . By the definition of  $\mathbf{v}$  we then have  $x_j = 2(n + \lfloor m/2^k \rfloor) + 1 \pmod 8$ . Therefore  $n$  is determined modulo 4. Let  $i' = i + 2^{2k}(x_j - 1 - 2\lfloor m/2^k \rfloor)$ . Then all occurrences of  $x$  in  $\mathbf{v}$  are at position  $i' + 2^{2k+3}n'$  for some  $n' \in \mathbb{Z}$ .  $\square$

**Lemma 2.6.** *Let  $A = \{c_1, c_2, \dots, c_d\}$  be any finite alphabet, and define the substitution  $\sigma$  from  $A^*$  to  $B^*$  by  $\sigma(c_i) = a^{d+1-i}b^i$  for all  $c_i \in A$ . Then  $\sigma$  preserves  $\text{ind}^*$ , i.e., if  $\mathbf{v}$  is any infinite word in  $A^{\mathbb{N}}$  and  $\mathbf{u} = \sigma(\mathbf{v})$ , then  $\text{ind}^*(\mathbf{u}) = \text{ind}^*(\mathbf{v})$ .*

*Proof.* It is clear that  $e(\sigma(w)) \geq e(w)$  for all  $w \in L(\mathbf{v})$ , therefore  $\text{ind}^*(\mathbf{u}) \geq \text{ind}^*(\mathbf{v})$  (this holds for any uniform substitution).

Conversely, observe that  $ba$  does not occur in  $\sigma(A)$ , but always occurs when two such image words are concatenated. Let  $w' \in L(\mathbf{u})$ ,  $x'$  be its maximal border, and  $z'$  the corresponding period, so that  $w' = z'x'$ . Then either  $|x'| \leq d + 1$  or  $x'$  contains  $ba$ . In the former case  $e(w') = |w'|/(|w'| - |x'|) = 1 + O(1/|w'|)$ . In the latter case one can write  $x' = s\sigma(x)p$  and  $z's = s\sigma(z)$ , with  $|p| \leq d$ ,  $|s| \leq d$ ,

and  $w = zx \in L(\mathbf{v})$ ,  $z$  being a period of  $w$ . We have  $|z'| = (d + 1)|z|$  and  $|w'| \leq (d + 1)|w| + 2d$ , hence  $e(w') = |w'|/|z'| \leq (|w| + 2)/|z| \leq e(w)(1 + O(1/|w|))$ . Therefore  $\text{ind}^*(\mathbf{u}) \leq \text{ind}^*(\mathbf{v})$ .  $\square$

2.3. INITIAL REPETITIONS

Let us now restrict to initial repetitions, *i.e.*, repetitions that occur as prefixes. The *initial critical exponent* of  $\mathbf{u}$  is the supremum of exponents of repetitions that are prefixes of  $\mathbf{u}$ :

$$\text{ice}(\mathbf{u}) = \sup\{e(w) : w \text{ prefix of } \mathbf{u}\} \in [1, +\infty].$$

If only long repetitions are considered, we get the *asymptotic initial critical exponent*:

$$\text{ice}^*(\mathbf{u}) = \lim_{n \rightarrow \infty} \sup\{e(w) : w \text{ prefix of } \mathbf{u} \text{ and } |w| \geq n\} \in [1, +\infty].$$

Among all infinite words,  $\text{ice}(ab^\omega) = \text{ice}^*(ab^\omega) = 1$  is trivially optimal.

One has  $\text{ice}(\mathbf{f}) = \text{ice}^*(\mathbf{f}) = \Phi + 1$ , so we may expect  $\mathbf{f}$  to be optimal among Sturmian words.

Every Sturmian word  $\mathbf{u}$  has infinitely many square prefixes [2] hence  $\text{ice}(\mathbf{u}) \geq \text{ice}^*(\mathbf{u}) \geq 2$ . Berthé *et al.* [4] constructed a class of Sturmian words such that  $\text{ice}^*(\mathbf{u}) = 2$  (see Th. 1.1 of [4]). One such word is:

**Proposition 2.7.** *Let  $\alpha = \sum_{n=1}^{\infty} f_n 2^{-n}$  be the real number whose binary expansion is the Fibonacci word, where the letters are assigned the values  $a = 1$  and  $b = 0$ :  $\alpha = .1011010110110\dots_2 \simeq .710$ . Then the continued fraction expansion of  $\alpha$  is  $\alpha = [0; 2^0, 2^1, 2^1, 2^2, 2^3, 2^5, 2^8, 2^{13}, 2^{21}, 2^{34}, \dots]$ , where exponents are the Fibonacci numbers, and there is a Sturmian word  $\mathbf{u}$  of slope  $\alpha$  such that  $\text{ice}^*(\mathbf{u}) = 2$ .*

*Proof.* We only need to compute the continued fraction expansion of  $\alpha$ , as the last statement follows from [4].

Let  $a_n = 2^{F_{n-1}}$  be the desired partial quotients; then the continued fraction is equal to  $\lim_{n \rightarrow \infty} p_n/q_n$  where  $p_{-1} = 1, p_0 = 0, q_{-1} = 0, q_0 = 1, p_{n+1} = a_{n+1}p_n + p_{n-1}, q_{n+1} = a_{n+1}q_n + q_{n-1}$ . We prove by induction that  $q_n = 2^{F_{n+1}} - 1$ : indeed, if  $q_n = 2^{F_{n+1}} - 1$  and  $q_{n-1} = 2^{F_n} - 1$ , then

$$q_{n+1} = 2^{F_n}(2^{F_{n+1}} - 1) + 2^{F_n} - 1 = 2^{F_{n+2}} - 1.$$

Also by induction, we prove that  $p_n = \sum_{i=1}^{F_{n+1}} f_i 2^{F_{n+1}-i}$  is the integer whose binary expansion is the prefix of length  $F_{n+1}$  of  $\mathbf{f}$ , *i.e.*,  $\varphi^{n-1}(a)$ . Indeed, the relation  $p_{n+1} = 2^{F_n}p_n + p_{n-1}$  amounts to concatenating the binary expansions of  $p_n$  and  $p_{n-1}$ , and we know that  $\varphi^n(a) = \varphi^{n-1}(a)\varphi^{n-2}(a)$ . Finally,

$$\lim_{n \rightarrow \infty} \frac{p_n}{q_n} = \sum_{i=1}^{\infty} f_i 2^{-i} = \alpha. \quad \square$$

## 2.4. INITIAL REPETITIONS IN A SUBSHIFT

The index of an infinite word depends only on its language of factors; consequently, all elements of a minimal subshift have the same index since they all have the same language of factors. On the other hand, the initial critical exponent and its asymptotic counterpart are dependent on the particular infinite word that is considered, so it is interesting to study how they vary within a given subshift.

Let  $I(\mathbf{u})$  be the infimum of  $\text{ice}^*(\mathbf{v})$  where  $\mathbf{v}$  is in the subshift generated by  $\mathbf{u}$ .

**Theorem 2.8.** *The Fibonacci word is maximal for  $I$  among all non periodic words.*

*Proof.* Obviously,  $I(\mathbf{u})$  is infinite when  $\mathbf{u}$  is periodic, so periodic words should be excluded. Mignosi *et al.* [20] proved that  $I(\mathbf{u}) \leq \Phi + 1$  for any non periodic  $\mathbf{u}$ .

Berthé *et al.* [4] proved that if  $\mathbf{u}$  is in the Fibonacci subshift, and is not in the shift orbit of  $\mathbf{f}$ , then it begins in arbitrarily long cubes:  $\text{ice}^*(\mathbf{u}) \geq 3$  (see Prop. 4.3 of [4]). Therefore the minimum is attained in the shift orbit of  $\mathbf{f}$ , where  $\text{ice}^*(\mathbf{u}) = \text{ice}^*(\mathbf{f}) = \Phi + 1$ . Hence  $I(\mathbf{f}) = \Phi + 1$ .  $\square$

## 3. RECURRENCE

The *recurrence function* of an infinite word  $\mathbf{u}$  was introduced by Morse and Hedlund [22]. It is defined by

$$R(n) = \inf\{N \in \mathbb{N} : \forall w \in L_N(\mathbf{u}), L_n(w) = L_n(\mathbf{u})\} \in \mathbb{N} \cup \{+\infty\}$$

and the *recurrence quotient* of  $\mathbf{u}$  by

$$\rho^*(\mathbf{u}) = \limsup_{n \rightarrow \infty} \frac{R(n)}{n} \in [1, +\infty].$$

The recurrence quotient of Sturmian words can be easily computed from the continued fraction expansion of their slope [10]:

**Theorem 3.1.** *If  $\mathbf{u}$  is a Sturmian word of slope  $\alpha = [0; a_1, a_2, a_3, \dots]$ , then*

$$\rho^*(\mathbf{u}) = 2 + \limsup_{n \rightarrow \infty} [a_n; a_{n-1}, a_{n-2}, \dots, a_1].$$

Consequently, as was already known by Morse and Hedlund [23], the Fibonacci word has  $\rho^*(\mathbf{f}) = \Phi + 2$  and this is the lowest possible value for a Sturmian word, for as soon as  $a_n$  is not eventually 1,  $\rho^*(\mathbf{u}) \geq 3 + \sqrt{2}$ .

Actually,  $\mathbf{f}$  seems to be also optimal among non-periodic words, as conjectured by Rauzy [27].

## 3.1. RECURRENCE QUOTIENT AND ASYMPTOTIC INDEX

We observe that a bound on the asymptotic index can be derived from the recurrence quotient.

**Proposition 3.2.** *For any infinite word  $\mathbf{u}$ ,  $\text{ind}^*(\mathbf{u}) \geq 1 + \frac{1}{\rho^*(\mathbf{u})-1}$ .*

*Proof.* If  $\rho^*(\mathbf{u}) = +\infty$ , the inequality obviously holds as  $\text{ind}^*(\mathbf{u}) \geq 1$ .

Assume now that  $\rho^*(\mathbf{u})$  is finite, so  $R(n)$  is finite too. Let  $x_n$  be the prefix of length  $n$  of  $\mathbf{u}$ , and  $z_n$  be the shortest prefix of  $\mathbf{u}$  such that  $z_n x_n$  is also a prefix of  $\mathbf{u}$  (in other words,  $|z_n|$  is the position of the second occurrence of  $x_n$  in  $\mathbf{u}$ ). Observe that the word obtained by removing the first and last letters in  $z_n x_n$  does not contain  $x_n$ , hence  $|z_n| + n - 2 < R(n)$ . Then obviously

$$\text{ind}^*(\mathbf{u}) \geq \text{ice}^*(\mathbf{u}) = \limsup_{n \rightarrow \infty} \frac{n}{|z_n|} + 1$$

and

$$\rho^*(\mathbf{u}) = \limsup_{n \rightarrow \infty} \frac{R(n)}{n} \geq \liminf_{n \rightarrow \infty} \frac{R(n)}{n} \geq \liminf_{n \rightarrow \infty} \frac{|z_n|}{n} + 1$$

and the result follows from

$$\limsup_{n \rightarrow \infty} \frac{n}{|z_n|} = \left( \liminf_{n \rightarrow \infty} \frac{|z_n|}{n} \right)^{-1}. \quad \square$$

In particular, when  $\text{ind}^*(\mathbf{u}) = 1$ , then  $\rho^*(\mathbf{u})$  has to be infinite. Apart from this case, and the periodic case where  $\text{ind}^*(\mathbf{u}) = +\infty$  and  $\rho^*(\mathbf{u}) = 1$ , equality cannot hold, as a consequence of the result of [11] that  $R(n)/n$  cannot converge to a finite limit when  $\mathbf{u}$  is not periodic.

**Open problem 1.** What is the infimum of  $(\text{ind}^*(\mathbf{u}) - 1)(\rho^*(\mathbf{u}) - 1)$  over all words  $\mathbf{u}$  for which both  $\text{ind}^*(\mathbf{u})$  and  $\rho^*(\mathbf{u})$  are finite?

The above inequality may suggest that  $\text{ind}^*$  and  $\rho^*$  vary somehow in opposite directions. However, this is not at all the case for Sturmian words:

**Theorem 3.3.** *If  $\mathbf{u}$  is a Sturmian word, then  $\rho^*(\mathbf{u}) = \text{ind}^*(\mathbf{u})$ .*

*Proof.* Just observe that Theorems 2.1 and 3.1 contain exactly the same formula. If  $\mathbf{u}$  is a Sturmian word of slope  $\alpha = [0; a_1, a_2, a_3, \dots]$ , then

$$\text{ind}^*(\mathbf{u}) = 2 + \limsup_{n \rightarrow \infty} [a_n; a_{n-1}, a_{n-2}, \dots, a_1]$$

and

$$\rho^*(\mathbf{u}) = 2 + \limsup_{n \rightarrow \infty} [a_n; a_{n-1}, a_{n-2}, \dots, a_1]. \quad \square$$

**Open problem 2.** How can this equality be explained? Does it characterize Sturmian words? (Compare, for instance, with the Prouhet-Thue-Morse word which has  $\text{ind}^*(\mathbf{t}) = 2$  and  $\rho^*(\mathbf{t}) = 10$ , see Prop. 6.1.)

3.2. FIRST OCCURRENCE

We now consider prefixes. Analogously to the recurrence function, we define

$$R'(n) = \inf\{N \in \mathbb{N} : L_n(u_0 \dots u_{N-1}) = L_n(\mathbf{u})\}$$

and

$$\rho'^*(\mathbf{u}) = \limsup_{n \rightarrow \infty} \frac{R'(n)}{n} \in [1, +\infty].$$

Note that  $R'(n) - n + 1$  is the maximal position where a factor of length  $n$  occurs for the first time.

When  $\mathbf{u}$  is eventually periodic, obviously  $\rho'^*(\mathbf{u}) = 1$ . Surprisingly, the lowest possible value for  $\rho'^*$  among non eventually periodic words is not attained by the Fibonacci word, for which  $\rho'^*(\mathbf{f}) = \Phi + 1$ , but by another Sturmian word [9]:

**Theorem 3.4.** *Let  $\mathbf{u}$  be the fixed point of  $a \mapsto abaababa, b \mapsto aba$ :*

$$\mathbf{u} = abaababaabaabaababaabaabaabaabaabaabaabaabaabaabaabaabaabaabaaba \dots$$

$\mathbf{u}$  is a non-standard Sturmian word, with slope  $\frac{5-\sqrt{10}}{5}$ . Then

$$\rho'^*(\mathbf{u}) = \frac{29 - 2\sqrt{10}}{9} \simeq 2.519 < 2.618 \simeq \Phi + 1,$$

and this is optimal.

4. PALINDROMES

4.1. PALINDROME DENSITIES

A palindrome is a finite word  $w$  which is equal to its mirror image  $\tilde{w}$ . The only infinite word all factors of which are palindromes is the constant word  $a^\omega$ ; other eventually periodic words may have a positive proportion of palindromes (for instance, one third of the factors of  $(aab)^\omega$  of each length are palindromes), or no palindromes after a certain length (like  $(aababb)^\omega$ ).

Assume now that  $\mathbf{u}$  is non eventually periodic. Let  $\text{fac}(n)$  be the subword complexity of  $\mathbf{u}$  (i.e., the number of its factors of length  $n$ ,  $\text{fac}(n) = \#L_n(\mathbf{u})$ ), and  $\text{pal}(n)$  its palindrome complexity (i.e., the number of palindromes of length  $n$  that are factors of  $\mathbf{u}$ ). As  $\text{pal}(n)$  is usually much smaller than  $\text{fac}(n)$ , instead of a proportion it is more interesting to consider the *lower palindrome density*

$$\pi(\mathbf{u}) = \liminf_{n \rightarrow \infty} \frac{n \text{pal}(n)}{\text{fac}(n)}$$



and the *total lower palindrome density*

$$\bar{\pi}(\mathbf{u}) = \liminf_{n \rightarrow \infty} \frac{n \sum_{k=0}^{n-1} \text{pal}(k)}{\sum_{k=0}^{n-1} \text{fac}(k)}.$$

For Sturmian words,  $\pi(\mathbf{u}) = 1$  and  $\bar{\pi}(\mathbf{u}) = 3$ , as follows from the characterization of Sturmian words using palindromes by Droubay and Pirillo [15]. Compare with the Prouhet-Thue-Morse word  $\mathbf{t}$ : as  $\text{pal}(5) = 0$ , the function  $\text{pal}(n)$  vanishes for odd  $n$  and  $\pi(\mathbf{t}) = 0$ ; but there are infinitely many palindromes of even length and one can compute that  $\bar{\pi}(\mathbf{t}) = 20/19$ , see Proposition 6.3.

**Open problem 3.** Does there exist a non eventually periodic infinite word  $\mathbf{u}$  such that  $\pi(\mathbf{u}) > 1$  or  $\bar{\pi}(\mathbf{u}) > 3$ ?

Note that, on the other hand, the *upper palindrome density*  $\limsup_{n \rightarrow \infty} \frac{n \text{pal}(n)}{\text{fac}(n)}$  can be infinite, as shown in [1], Remark 9.

#### 4.2. PALINDROMIC PREFIXES

Let  $n_0 = 0, n_1 = 1, n_2, \dots$  be the lengths, in increasing order, of palindromes that are prefixes of  $\mathbf{u}$ . Define then the *palindromic prefix gap* by

$$\delta(\mathbf{u}) = \limsup_{i \rightarrow \infty} \frac{n_{i+1}}{n_i}$$

with the convention  $\delta(\mathbf{u}) = +\infty$  if  $\mathbf{u}$  has finitely many palindromic prefixes.

If  $\mathbf{u}$  is periodic with palindromic period,  $\delta(\mathbf{u}) = 1$ . For the Fibonacci word,  $n_i = F_{i+3} - 2$  hence  $\delta(\mathbf{u}) = \Phi$ . For any other infinite word,  $\delta(\mathbf{u}) \geq 1 + \sqrt{2}/2 > \Phi$  [17].

#### 4.3. FIRST OCCURRENCE OF A PALINDROME

We conclude with one last open problem.

Let  $\mathbf{u}$  be a non eventually periodic word containing palindromes of each length. Let  $p_1(n)$  be the starting position of the first occurrence of a palindrome of length  $n$  in  $\mathbf{u}$ , and define the *first palindrome occurrence rate* by

$$\psi(\mathbf{u}) = \limsup_{n \rightarrow \infty} \frac{p_1(n)}{n}.$$

The Fibonacci word has  $\psi(\mathbf{f}) = \Phi$ .

**Open problem 4.** What is the minimal value of  $\psi$  for non eventually periodic words, and for which word is it attained (if it is)?

## 5. SUMMARY

**Index**  $\text{ind}(\mathbf{u})$ 

Among all words: infimum 1 (not attained), maximum  $+\infty$ .

Among binary words: minimum  $\text{ind}(\mathbf{t}) = 2$ , maximum  $+\infty$ .

Among Sturmian words: minimum  $\text{ind}(\mathbf{f}) = \Phi + 2$ , maximum  $+\infty$  (with gaps, *e.g.*  $(\Phi + 2, 11/3)$ ).

**Asymptotic index**  $\text{ind}^*(\mathbf{u})$ 

Among all words, or binary words: minimum 1 (see Th. 2.4), maximum  $+\infty$ .

Among Sturmian words: minimum  $\text{ind}^*(\mathbf{f}) = \Phi + 2$ , maximum  $+\infty$ .

**Initial critical exponent**  $\text{ice}(\mathbf{u})$ 

Among all words: minimum  $\text{ice}(ab^\omega) = 1$ , maximum  $\text{ice}(a^\omega) = +\infty$ . ( $\text{ice}(\mathbf{f}) = \Phi + 1$ ).

**Asymptotic initial critical exponent**  $\text{ice}^*(\mathbf{u})$ 

Among Sturmian words: minimum 2 (see Prop. 2.7), maximum  $+\infty$ .

**Minimal asymptotic initial critical exponent in a subshift**  $I(\mathbf{u})$ 

Among all words: minimum  $I(ab^\omega) = 1$ , maximum  $I(a^\omega) = +\infty$ .

Among non (purely) periodic words: minimum  $I(ab^\omega) = 1$ , maximum  $I(\mathbf{f}) = \Phi + 1$ .

Among Sturmian words: minimum 2 (see Prop. 2.7), maximum  $I(\mathbf{f}) = \Phi + 1$ .

**Recurrence quotient**  $\rho^*(\mathbf{u})$ 

Among Sturmian words: minimum  $\rho^*(\mathbf{f}) = \Phi + 2$ , maximum  $+\infty$  (with gaps, *e.g.*  $(\Phi + 2, 3 + \sqrt{2})$ ).

**First occurrence quotient**  $\rho'^*(\mathbf{u})$ 

Among all words: minimum  $\rho'^*(a^\omega) = 1$ , maximum  $+\infty$ .

Among non eventually periodic words: minimum  $\frac{29-2\sqrt{10}}{9}$  (see Th. 3.4), maximum  $+\infty$  ( $\rho'^*(\mathbf{f}) = \Phi + 1$ ).

**Lower palindrome density**  $\pi(\mathbf{u})$ 

Among non eventually periodic words: minimum 0, maximum unknown.

Among Sturmian words: constant 1.

**Total lower palindrome density**  $\bar{\pi}(\mathbf{u})$ 

Among non eventually periodic words: minimum 0, maximum unknown.

Among Sturmian words: constant 3.

**Palindromic prefix gap**  $\delta(\mathbf{u})$ 

Among all words: minimum 1, maximum  $+\infty$ .

Among non periodic words: minimum  $\delta(\mathbf{f}) = \Phi$ , maximum  $+\infty$  (with gaps, *e.g.*  $(\Phi, 1 + \sqrt{2}/2)$ ).

**First palindrome occurrence rate**  $\psi(\mathbf{u})$ 

Among non eventually periodic words: minimum unknown, maximum  $+\infty$

( $\psi(\mathbf{f}) = \Phi$ ).



*Proof.* The set of singular factors of  $\mathbf{t}$  is:

$$S = \{a, b, aa, ab, ba, bb, aba, bab\} \cup \{x\theta^k(z)y : x, y, z \in B \text{ and } k \geq 1\}.$$

The return words of small singular factors are listed below, up to symmetries (mirror  $w \mapsto \tilde{w}$  and alphabet permutation  $w \mapsto E(w)$ ):

$$\begin{aligned} r(a) &= \{a, ab, abb\}; \\ r(ab) &= \{ab, aba, abb, abba\}; \\ r(aa) &= \{aabb, aababb, aabbab, aababbab\}; \\ r(aba) &= \{aba, ababb, abaabba, ababbaabb\}; \\ r(aabb) &= \{aabbab, aabbaababb, aabbabaababbab, abbaababbabaababb\}; \\ r(aba) &= \{abaababb, abaabbaababb, abaababbaabb, abaabbaabbaababb\}; \\ r(abab) &= \{ababbaba, ababbaabbaba, ababbabaabba, ababbaabbabaabba\}. \end{aligned}$$

The return words of other singular factors are obtained recursively, using the following lemma: if  $x, y, z \in B$  and  $k \geq 2$ , then

$$r(x\theta^k(z)y) = E(x)^{-1}\theta(r(E(x)\theta^{k-1}(z)y))E(x)$$

(recall that  $E(a) = b$  and  $E(b) = a$ , so that  $\theta(E(x)) = E(x)x$ ). As a consequence,  $\ell(x\theta^k(z)y) = 2\ell(E(x)\theta^{k-1}(z)y)$ , and we get  $\ell(x\theta^k(z)y) = 2^k c$  with  $c = 9$  (if  $k$  is odd and  $x = z \neq y$ , or if  $k$  is even and  $x = y \neq z$ ) or  $c = 8$  (otherwise). We deduce that  $R(n) = n - 1 + 9 \cdot 2^k$  for  $2^k + 2 \leq n < 2^{k+1} + 2$ , for all  $k \in \mathbb{N}$ . As a direct consequence,  $\rho^*(\mathbf{t}) = 10$ .  $\square$

Palindromes in  $\mathbf{t}$  can be described recursively. It is easier to simultaneously describe *antipalindromes*, *i.e.*, words  $w$  such that  $\tilde{w} = E(w)$ . Let  $\text{ap}(n)$  denote the number of antipalindromes of length  $n$  in  $\mathbf{t}$ . Let also PAL denote the set of all palindromes in  $\mathbf{t}$  and AP the set of all antipalindromes in  $\mathbf{t}$ . If  $w$  is a word of length at least 2, let  $\gamma(w)$  be the word obtained by deleting the first and last letter in  $w$ .

**Lemma 6.2.** *The sets PAL and AP satisfy*

$$\begin{aligned} \text{PAL} &= \{a, b, aba, bab\} \cup \theta(\text{AP}) \cup \gamma(\theta(\text{AP})) \\ \text{AP} &= \theta(\text{PAL}) \cup \gamma(\theta(\text{PAL})) \end{aligned}$$

and the functions  $\text{pal}(n)$  and  $\text{ap}(n)$  satisfy

$$\begin{aligned} \text{pal}(4n) &= \text{ap}(2n) \\ \text{pal}(4n + 2) &= \text{ap}(2n + 2) \\ \text{pal}(1) &= 2 \\ \text{pal}(3) &= 2 \\ \text{pal}(2n + 1) &= 0 && (\text{if } n \geq 2) \\ \text{ap}(0) &= 1 \\ \text{ap}(2) &= 2 \\ \text{ap}(2n) &= \text{pal}(n) + \text{pal}(n + 1) && (\text{if } n \geq 2) \\ \text{ap}(2n + 1) &= 0 \end{aligned}$$

for all  $n \in \mathbb{N}$  (except when otherwise noted). They are given by

$$\text{pal}(n) = \begin{cases} 1 & \text{if } n = 0 \\ 2 & \text{if } 1 \leq n \leq 4 \\ 0 & \text{if } n \geq 5 \text{ and } n \text{ is odd} \\ 2 & \text{if } 3 \cdot 4^k < n \leq 4^{k+1} \text{ and } n \text{ is even} \\ 4 & \text{if } 4^{k+1} < n \leq 3 \cdot 4^{k+1} \text{ and } n \text{ is even} \end{cases}$$

$$\text{ap}(n) = \begin{cases} 0 & \text{if } n \text{ is odd} \\ 1 & \text{if } n = 0 \\ 2 & \text{if } n = 2 \\ 4 & \text{if } 2 \cdot 4^k < n \leq 6 \cdot 4^k \text{ and } n \text{ is even} \\ 2 & \text{if } 6 \cdot 4^k < n \leq 2 \cdot 4^{k+1} \text{ and } n \text{ is even} \end{cases}$$

for all  $k \in \mathbb{N}$ .

*Proof.* Observe first that  $\widetilde{\theta(w)} = E(\theta(\tilde{w}))$ . As a consequence,  $\theta(\text{PAL}) \subseteq \text{AP}$  and  $\theta(\text{AP}) \subseteq \text{PAL}$ . It is also clear that PAL and AP are stable under  $\gamma$ . This proves inclusions in one direction.

To prove the reverse inclusions, consider first palindromes of odd length. It is easy to check that among the four palindromes of length 3, only two occur in  $\mathbf{t}$ , and that none of the eight palindromes of length 5 occurs in  $\mathbf{t}$ ; therefore no longer palindrome of odd length occurs in  $\mathbf{t}$ . Obviously there are no antipalindromes of odd length.

Consider now factors of even length. A factor  $w$  of  $\mathbf{t}$  of even length is always either of the form  $\theta(w')$  (if it occurs at an even position) or  $\gamma(\theta(w'))$  (if it occurs at an odd position). If  $w$  is a palindrome, then  $w'$  is an antipalindrome, and if  $w$  is an antipalindrome, then  $w'$  is a palindrome. This proves the language equalities.

To get the recurrence relations, one has to pay attention to the fact that the language equalities may be ambiguous; for instance,  $ab$  is both in  $\theta(\text{PAL})$  and  $\gamma(\theta(\text{PAL}))$ . One checks that  $\theta(A^*) \cap \gamma(\theta(A^*)) = \{ab\}^* \cup \{ba\}^*$ , and consequently  $\theta(L(\mathbf{t})) \cap \gamma(\theta(L(\mathbf{t}))) = \{\varepsilon, ab, ba\}$  so ambiguity affects only words of length up to 2.

The last formulas are easily deduced from the recurrence relations.  $\square$

**Proposition 6.3.** *The total lower palindrome density of the Prouhet-Thue-Morse word is  $\bar{\pi}(\mathbf{t}) = 20/19$ .*

*Proof.* Recall that

$$\bar{\pi}(\mathbf{t}) = \liminf_{n \rightarrow \infty} g(n)$$

where

$$g(n) = \frac{n \sum_{k=0}^{n-1} \text{pal}(k)}{\sum_{k=0}^{n-1} \text{fac}(k)}.$$

From the formulas for  $\text{pal}(n)$  and  $\text{fac}(n)$ , a long but elementary computation produces a formula for  $g(n)$ , with eight different cases. For instance, if  $n$  is odd and  $4.4^k < n \leq 6.4^k$ , one has

$$g(n) = \frac{n(6n + 1 - 4.4^k)}{6n^2 - 12.4^k n + 28.4^{2k} - 18n + 18.4^k + 23}.$$

One finds that  $\liminf_{n \rightarrow \infty} g(n) = \lim_{k \rightarrow \infty} g(4^k + 1) = \frac{20}{19}$  and  $\limsup_{n \rightarrow \infty} g(n) = \frac{39 + \sqrt{1554}}{66}$ .  $\square$

## REFERENCES

- [1] J.-P. Allouche, M. Baake, J. Cassaigne and D. Damanik, Palindrome complexity. *Theoret. Comput. Sci.* **292** (2003) 9–31.
- [2] J.-P. Allouche, J.L. Davison, M. Queffélec and L.Q. Zamboni, Transcendence of Sturmian or morphic continued fractions. *J. Number Theory* **91** (2001) 39–66.
- [3] J. Berstel, On the index of Sturmian words, in *Jewels are Forever*. Springer, Berlin (1999) 287–294.
- [4] V. Berthé, C. Holton and L.Q. Zamboni, Initial powers of Sturmian sequences. *Acta Arith.* **122** (2006) 315–347.
- [5] S. Brlek, Enumeration of factors in the Thue-Morse word. *Discrete Appl. Math.* **24** (1989) 83–96.
- [6] A. Carpi, On Dejean’s conjecture over large alphabets. *Theoret. Comput. Sci.* **385** (2007) 137–151.
- [7] A. Carpi and A. de Luca, Special factors, periodicity, an application to Sturmian words. *Acta Inform.* **36** (2000) 983–1006.
- [8] J. Cassaigne, Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin* **4** (1997) 67–88.
- [9] J. Cassaigne, On a conjecture of J. Shallit, in *Automata, languages and programming (ICALP 1997)*, Springer, Berlin. *Lect. Notes Comput. Sci.* **1256** (1997) 693–704.
- [10] J. Cassaigne, Limit values of the recurrence quotient of Sturmian sequences. *Theoret. Comput. Sci.* **218** (1999) 3–12.
- [11] J. Cassaigne and N. Chekhova, Fonctions de récurrence des suites d’Arnoux-Rauzy et réponse à une question de Morse et Hedlund. *Ann. Inst. Fourier (Grenoble)* **56** (2006) 2249–2270.
- [12] J.D. Currie and M. Mohammad-Noori, Dejean’s conjecture and Sturmian words. *Eur. J. Combin.* **28** (2007) 876–890. Also in Morteza Mohammad-Noori, PhD. thesis, Université Paris-Sud (2005).

- [13] D. Damanik and D. Lenz, The index of Sturmian sequences. *Eur. J. Combin.* **23** (2002) 23–29.
- [14] F. Dejean, Sur un théorème de Thue. *J. Comb. Theory A* **13** (1972) 90–99.
- [15] X. Droubay and G. Pirillo, Palindromes and Sturmian words. *Theoret. Comput. Sci.* **223** (1999) 73–85.
- [16] R.C. Entringer, D.E. Jackson and J.A. Schatz, On nonrepetitive sequences. *J. Comb. Theory A* **16** (1974) 159–164.
- [17] S. Fischler, Palindromic prefixes and episturmian words. *J. Comb. Theory A* **113** (2006) 1281–1304.
- [18] M. Lothaire, *Algebraic combinatorics on words*, *Encyclopedia of Mathematics and its Applications* **90**. Cambridge University Press, Cambridge (2002).
- [19] F. Mignosi and G. Pirillo, Repetitions in the Fibonacci infinite word. *RAIRO-Theor. Inf. Appl.* **26** (1992) 199–204.
- [20] F. Mignosi, A. Restivo and S. Salemi, Periodicity and the golden ratio. *Theoret. Comput. Sci.* **204** (1998) 153–167.
- [21] M. Morse, Recurrent geodesics on a surface of negative curvature. *Trans. Amer. Math. Soc.* **22** (1921) 84–100.
- [22] M. Morse and G.A. Hedlund, Symbolic dynamics. *Amer. J. Math.* **60** (1938) 815–866.
- [23] M. Morse and G.A. Hedlund, Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.* **62** (1940) 1–42.
- [24] J. Moulin-Ollagnier, Proof of Dejean’s conjecture for alphabets with 5, 6, 7, 8, 9, 10 and 11 letters. *Theoret. Comput. Sci.* **95** (1992) 187–205.
- [25] J.-J. Pansiot, À propos d’une conjecture de F. Dejean sur les répétitions dans les mots. *Discrete Appl. Math.* **7** (1984) 297–311.
- [26] É. Prouhet, Mémoire sur quelques relations entre les puissances des nombres. *C. R. Acad. Sci. Paris Sér. I* **33** (1851) 225.
- [27] G. Rauzy, Suites à termes dans un alphabet fini, in *Seminaire de théorie des nombres 1982–1983*, Univ. Bordeaux I, 1983. Exposé 25.
- [28] A. Thue, Über unendliche Zeichenreihen. *Norske Vid. Selsk. Skr., I. Mat. Nat. Kl., Christiania* **7** (1906) 1–22.
- [29] D. Vandeth, Sturmian words and words with a critical exponent. *Theoret. Comput. Sci.* **242** (2000) 283–300.