Asymptotic Behaviour of the Maximal Number of Squares in Standard Sturmian Words
(Extended Abstract)

Marcin Piątkowski[2⋆] and Wojciech Rytter[1,2⋆⋆]

[1] Department of Mathematics, Computer Science and Mechanics,
University of Warsaw, Warsaw, Poland
rytter@mimuw.edu.pl
[2] Faculty of Mathematics and Informatics,
Nicolaus Copernicus University, Toruń, Poland
martinp@mat.umk.pl

Abstract. Denote by $sq(w)$ the number of distinct squares in a string w and let $\mathcal{S}$ be the class of standard Sturmian words. They are generalizations of Fibonacci words and are important in combinatorics on words. For Fibonacci words the asymptotic behaviour of the number of runs and the number of squares is the same. We show that for Sturmian words the situation is quite different. The tight bound $\frac{8}{10}|w|$ for the number of runs was given in [3]. In this paper we show that the tight bound for the maximal number of squares is $\frac{9}{10}|w|$. We use the results of [11] where exact (but not closed) complicated formulas were given for $sq(w)$ for $w \in \mathcal{S}$ and we show:

(1) for all $w \in \mathcal{S} \quad sq(w) \leq \frac{9}{10}\,|w| + 4$,

(2) there is an infinite sequence of words $w_k \in \mathcal{S}$ such that

$$\lim_{k\to\infty} |w_k| \ = \ \infty \qquad \text{and} \qquad \lim_{k\to\infty} \frac{sq(w_k)}{|w_k|} \ = \ \frac{9}{10}.$$

Surprisingly the maximal number of runs is reached by the words with recurrences of length only 5. This contrasts with the situation of Fibbonaci words, though standard Sturmian words are natural extension of Fibonacci words. If this length drops to 4, the asymtotic behaviour of the maximal number of squares falls down significantly below $\frac{9}{10}|w|$. The structure of Sturmian words rich in squares has been discovered by us experimentally and verified theoretically. The upper bound is much harder, its proof is not a matter of simple calculations. The summation formulas for the number of squares are complicated, no closed formula is known. Some nontrivial reductions were necessary.

1 Introduction

A square in a string is a subword of the form ww, where w is nonempty. The squares are a simplest form of repetitions, despite the simple formulation many combinatorial problems related to squares are not well understood. The subject of computing maximal number of squares and repetitions in words is one of the fundamental topics in combinatorics on words [18,22] initiated by A. Thue [28], as well as it is important in other areas: lossless compression, word representation, computational biology, etc.

⋆ The research supported by Ministry of Science and Higher Education of Poland, grant N N206 258035.

⋆⋆ Supported by grant N206 004 32/0806 of the Polish Ministry of Science and Higher Education.

Let $sq(w)$ be the number of distinct squares in the word w and $sq(n)$ be the maximal number of distinct squares in the word of length n. The behaviour of the function $sq(n)$ is not well understood, though the subject of squares was studied by many authors, see [9,10,17]. The best known results related to the value of $sq(n)$ are, see [13,15,16]:

$$n - o(n) \ \leq\ sq(n) \ \leq\ 2n - O(\log n).$$

In this paper we concentrate on the asymptotic behaviour of the maximal number of squares in class of standard Sturmian words $\mathcal{S}$. We show: for all $w \in \mathcal{S}\ \ sq(w) \leq \frac{9}{10}\,|w|$ and there is an infinite sequence of strictly growing words $\{w_k\} \in \mathcal{S}$ such that

$$\lim_{k\to\infty} \frac{sq(w_k)}{|w_k|} \ =\ \frac{9}{10}\,.$$

There are known efficient algorithms for the computation of integer powers in words, see [2,6,11,23,24]. The powers in words are related to maximal repetitions, also called *runs*. It is surprising that the known bounds for the number of runs are much tighter than for squares, this is due to the work of many people [3,7,8,14,19,20,25,26,27].

One of interesting questions related to squares is the relation of their number to the number of runs. In case of Fibonacci words the number of squares and runs differ only by 1.

The results of this paper show that the maximal number of squares and the maximal number of runs are possibly not closely related, since in case of well structured words (Sturmian words) the *density ratio* of squares (the asymptotic quotient of the maximal number of squares by the size of the string) is $\frac{9}{10}$ and for runs it is $\frac{8}{10}$.

2 Standard Sturmian words

The *standard Sturmian words* (*standard words*, in short) are generalization of Fibonacci words and have a very simple *grammar-based* representation which has some algorithmic consequences.

Let $\mathcal{S}$ denote the set of all standard Sturmian words. These words are defined over a binary alphabet $\Sigma = \{a, b\}$ and are described by recurrences (or grammar-based representation) corresponding to so called *directive sequences*: integer sequences

$$\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_n),$$

where $\gamma_0 \geq 0$, $\gamma_i > 0$ for $0 < i \leq n$.

The word x_{n+1} corresponding to γ, denoted by $\mathrm{Sw}(\gamma)$, is defined by recurrences:

$$\begin{array}{ll} x_{-1} = b, & x_0 = a, \\ x_1 = x_0^{\gamma_0} x_{-1}, & x_2 = x_1^{\gamma_1} x_0, \\ \ldots & \ldots \\ x_n = x_{n-1}^{\gamma_{n-1}} x_{n-2}, & x_{n+1} = x_n^{\gamma_n} x_{n-1}. \end{array} \tag{1}$$

Fibonacci words are standard Sturmian words given by the directive sequences of the form $\gamma = (1, 1, \ldots, 1)$ (n-th Fibonacci word F_n corresponds to a sequence of

n ones). We consider here standard words starting with the letter a, hence assume $\gamma_0 > 0$. The case $\gamma_0 = 0$ can be considered similarly.

For even $n > 0$ a standard word x_n has the suffix ba, and for odd $n > 0$ it has the suffix ab. The number $N = |x_{n+1}|$ is the (real) size, while $n+1$ can be thought as the compressed size.

Example 1.
Consider directive sequence $\gamma = (1, 2, 1, 3, 1)$. We have:

$$\mathrm{Sw}(1,2,1,3,1) \;=\; abababababaababababaabababaababaab$$

$$x_{-1} = b, \quad x_0 = a, \quad x_1 = x_0^1 x_{-1} = a\ b, \quad x_2 = x_1^2 x_0 = ab\ ab\ a,$$

$$x_3 = x_2^1 x_1 = ababa\ ab, \quad x_4 = x_3^3 x_2 = ababaab\ ababaab\ ababaab\ ababa,$$

$$x_5 = x_4^1 x_3 = ababaabababaabababaababababa\ ababaab$$

The *grammar-based compression* consists in describing a given word by a context-free grammar G generating this (single) word. The size of the grammar G is the total length of all productions of G. In particular each directive sequence of a standard Sturmian word corresponds to such a compression – the sequence of recurrences corresponding to the directive sequence. In this case the size of the grammar is proportional to the length of the directive sequence.

For some lexicographic properties and structure of repetitions of standard Sturmian words see [5,3,1,4].

3 Summation formulas for the number of squares

The exact formulas for the number of squares in standard Sturmian words were given by Damanik and Lenz in [11]. In this section we reformulate their formulas to have compact version more suitable for the asymptotic analysis. The formulas are rather complicated and such an analysis is nontrivial. It will be done in the section 5.

Denote $q_i = |x_i|$, where x_i are as in equation (1). The following lemma charecterize the possible lengths of periods of squares in Sturmian words.

Lemma 2. ([11])
Let $w = \mathrm{Sw}(\gamma_0, \gamma_1, \ldots, \gamma_n)$ be a standard Sturmian word. Each primitive period of a square in w has the length kq_i for $1 \le k \le \gamma_i$ or $kq_i + q_{i-1}$ for $1 \le k < \gamma_i$.

The squares in standard Sturmian word w with period of the length kq_i for $1 \le k \le \gamma_i$ or $kq_i + q_{i-1}$ for $1 \le k < \gamma_i$ are said to be of type i.

Example 3.
Consider the word from Example 1:

$$\mathrm{Sw}(1,2,1,3,1) \;=\; abababababaababababaabababaababaab.$$

We have:

one square of type 0: $a{\cdot}a$,

three squares of type 1 (period 2, 3): $ab{\cdot}ab$, $ba{\cdot}ba$, $aba{\cdot}aba$,

three squares of type 2 (period 5): $ababa{\cdot}ababa$, $babaab{\cdot}babaab$, $abaab{\cdot}abaab$,

and eleven squares of type 3 (with periods 7,14):

$ababaab \cdot ababaab$, $babaaba \cdot babaaba$, $abaabab \cdot abaabab$, $baababa \cdot baababa$,

$aababab \cdot aababab$, $abababa \cdot abababa$, $bababaa \cdot bababaa$,

$ababaabababaab \cdot ababaabababaab$, $babaababababaaba \cdot babaababababaaba$,

$abaababababaabab \cdot abaababababaabab$, $baabababaababa \cdot baabababaababa$.

Let $sq_i(\gamma_0, \gamma_1, \ldots, \gamma_n)$, for $1 \le i \le n$, be the number of squares of the type i and let $sq_0(\gamma_0, \gamma_1, \ldots, \gamma_n)$ be the number of squares with period of the form a^+ in the word $\mathrm{Sw}(\gamma_0, \gamma_1, \ldots, \gamma_n)$.
We slightly abuse the notation and denote $sq(\gamma_0, \gamma_1, ..., \gamma_n) = sq\big(\mathrm{Sw}(\gamma_0, \gamma_1, ..., \gamma_n)\big)$.

Denote $d(0) = \left\lfloor \frac{\gamma_0+1}{2} \right\rfloor$ and for $1 \le i \le n$ and $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_n)$:

$$d_1(i) = \begin{cases} \frac{\gamma_i}{2} q_i + q_{i-1} - 1 & \text{for even } \gamma_i \\ \frac{\gamma_i}{2} q_i + \frac{1}{2}\, q_i & \text{for odd } \gamma_i \end{cases}$$

$$d(i) = d_1(i) + \gamma_i\, q_i - q_i - \gamma_i + 1.$$

Let $\mathrm{Sw}(\gamma_0, \gamma_1, \ldots, \gamma_n)$ be a standard Sturmian word. Then $sq(\gamma_0, \gamma_1, \ldots, \gamma_n)$ is determined as follows, see [11]:

Summation formulas:

(1) $sq(\gamma_0, \gamma_1, \ldots, \gamma_n) = \sum_{i=0}^{n} sq_i(\gamma_0, \gamma_1, \ldots, \gamma_n)$.

(2) $(0 \le i \le n-3)$ or $(i = n-2 \ \& \ \gamma_n \ge 2) \Rightarrow sq_i(\gamma) = d(i)$.

(3) $\gamma_n = 1 \Rightarrow sq_{n-2}(\gamma) = \begin{cases} d(n-2) - q_{n-3} + 1 & \text{for even } \gamma_{n-2} \\ d(n-2) - q_{n-2} + q_{n-3} + 1 & \text{otherwise} \end{cases}$

(4) $\gamma_n = 1 \Rightarrow sq_{n-1}(\gamma) = \begin{cases} d_1(n-1) - q_{n-2} + 1 & \text{for even } \gamma_{n-1} \\ d_1(n-1) - q_{n-1} + q_{n-2} - 1 & \text{otherwise} \end{cases}$

(5) $\gamma_n > 1 \Rightarrow sq_{n-1}(\gamma) = \begin{cases} d(n-1) - q_{n-2} + 1 & \text{for even } \gamma_{n-1} \\ d(n-1) - q_{n-1} + q_{n-2} - 1 & \text{otherwise} \end{cases}$

(6) $sq_n(\gamma) = \begin{cases} d_1(n) - q_n + 2 & \text{for even } \gamma_n \\ d_1(n) - q_n & \text{otherwise} \end{cases}$

4 Sturmian words with many squares

In this section we present and analyse the sequence $\{w_k\}$ of Sturmian words achieving asymptotically maximal ratio:

$$\lim_{k\to\infty} |w_k| = \infty \qquad \text{and} \qquad \lim_{k\to\infty} \frac{squares(w_k)}{|w_k|} = \frac{9}{10}\,.$$

Recall that the squares with periods kq_i for $1 \le k \le \gamma_i$ or $kq_i + q_{i-1}$ are said to be of the type i.

Consider the words

$$w_k \;=\; \mathrm{Sw}(k, k, 2, 1, 1).$$

Example 4.

$$\begin{aligned}
w_1 &= \mathrm{Sw}(1,1,2,1,1) &&= (aba)^2ab(aba)^3ab,\\
w_2 &= \mathrm{Sw}(2,2,2,1,1) &&= \Big((aab)^2a\Big)^2 aab\Big((aab)^2a\Big)^3 aab,\\
w_3 &= \mathrm{Sw}(3,3,2,1,1) &&= \Big((aaab)^3a\Big)^2 aaab\Big((aaab)^3a\Big)^3 aaab.
\end{aligned}$$

$\mathrm{Sw}(3,3,2,1,1)$ is illustrated in Figure 1.

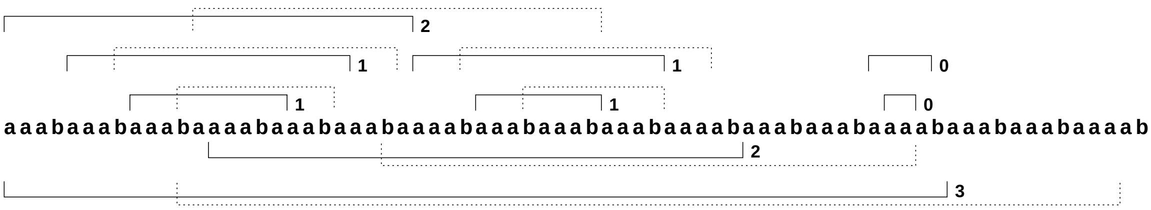

Figure 1. The squares in word $\mathrm{Sw}(3,3,2,1,1)$ with their *shifts* and types.

Theorem 5.
We have $sq(k,k,2,1,1) \longrightarrow \frac{9}{10}\cdot\Big|\mathrm{Sw}(\gamma_0,\gamma_1,\ldots,\gamma_n)\Big|$ *for* $k \longrightarrow \infty$.

Proof.
Let $\gamma = (k,k,2,1,1)$. We have:

$$\mathrm{Sw}(\gamma) \;=\; \Big((a^kb)^ka\Big)^2 a^kb\Big((a^kb)^ka\Big)^3 a^kb$$

and

$$|\mathrm{Sw}(\gamma)| \;=\; 5k^2 + 7k + 7.$$

We compute separately the number of squares for each type $0 \le i \le 4$.

There are two cases depending on the parity of k and we can assume that $k > 1$.

Case 1: k is odd.
We have (according to our formulas):

$$\begin{aligned} sq_0(\gamma) &= \frac{1}{2}\Big(k+1\Big), \\ sq_1(\gamma) &= \frac{1}{2}\Big(3k^2+1\Big), \\ sq_2(\gamma) &= 2k^2+2k+1, \\ sq_3(\gamma) &= k^2+k, \\ sq_4(\gamma) &= 0, \\ sq(\gamma) &= \frac{1}{2}\Big(9k^2+7k+4\Big). \end{aligned}$$

Finally

$$\lim_{k\to\infty} \frac{sq(\gamma)}{|\mathrm{Sw}(\gamma)|} = \lim_{k\to\infty} \frac{9k^2+7k+4}{10k^2+14k+14} = 0.9\,.$$

Case 2: k is even.
We have (according to our formulas):

$$\begin{aligned} sq_0(\gamma) &= \frac{1}{2}k, \\ sq_1(\gamma) &= \frac{1}{2}\Big(3k^2-k\Big), \\ sq_2(\gamma) &= 2k^2+2k+1, \\ sq_3(\gamma) &= k^2+k, \\ sq_4(\gamma) &= 0, \\ sq(\gamma) &= \frac{1}{2}\Big(9k^2+6k+2\Big). \end{aligned}$$

Finally

$$\lim_{k\to\infty} \frac{sq(\gamma)}{|\mathrm{Sw}(\gamma)|} = \lim_{k\to\infty} \frac{9k^2+6k+2}{10k^2+14k+14} = 0.9.$$

This concludes the proof.

5 Asymptotic behaviour of the maximal number of squares

The formulas (1-6) from the section 3 give together the value of $sq(\gamma)$, however there is no close simple formula. Therefore tight asymptotic estimations are nontrivial. We start with an estimation for short γ. The proof of the following simple lemma is omitted in this version.

Lemma 6. [Short γ]
$sq(\gamma_0,\gamma_1,\gamma_2) \le \frac{7}{3}|\mathrm{Sw}(\gamma_0,\gamma_1,\gamma_2)|$ *and* $sq(\gamma_0,\gamma_1,\gamma_2) \le |\mathrm{Sw}(\gamma_0,\gamma_1,\gamma_2)| - 4$.

For a word w of length at least 2 denote by $exch_2(w)$ the word resulting from w by exchanging the last two letters.

Lemma 7. $sq(w) \leq sq\big(exch_2(w)\big) + 4$.

Proof.
It is known, see [13], that there are at most two *last* occurrences of different squares at a single position in a string. If we reverse the word then this corresponds to the end-positions of the *first* occurrences. Hence at the last two positions at most 4 different squares can end which do not appear earlier in the same word with the last two letters removed. This completes the proof.

The next two lemmas allows us to restrict the values of last two elements of the directive sequence in the asymptotic estimation of $sq(\gamma)$.

Lemma 8. [Reduction of γ_n]
Let $\mathrm{Sw}(\gamma_0, \gamma_1, \ldots, \gamma_n)$ *be a standard Sturmian word. If* $\gamma_n > 1$ *then*

$$sq(\gamma_0, \ldots, \gamma_{n-1}, \gamma_n) \leq sq(\gamma_0, \ldots, \gamma_{n-1}, \gamma_n - 1, 1) + 4.$$

Proof.
The words $\mathrm{Sw}(\gamma_0, \ldots, \gamma_n - 1, 1)$ and $\mathrm{Sw}(\gamma_0, \ldots, \gamma_n)$ differ only on the last two letters, see [18]. Hence

$$\mathrm{Sw}(\gamma_0, \ldots, \gamma_n - 1, 1) \;=\; exch_2\big(\mathrm{Sw}(\gamma_0, \ldots, \gamma_n)\big).$$

Now the thesis follows from the Lemma 7.

Lemma 9. [Reduction of γ_{n-1}]
Let $x \;=\; SW(\gamma_0, \ldots, \gamma_{n-2}, \gamma_{n-1}, 1)$, $x' \;=\; SW(\gamma_0, \ldots, \gamma_{n-2}, 1, 1)$,
$x'' \;=\; SW(\gamma_0, \ldots, \gamma_{n-2}, 2, 1)$.
Then

$$\left(sq(x') \;\leq\; \frac{9}{10}\,|x'| \text{ and } sq(x'') < \frac{9}{10}\,|x''| \right) \;\Rightarrow\; sq(x) \;\leq\; \frac{9}{10}\,|x|.$$

Proof.
If γ_{n-1} is odd then let $\Delta = \gamma_{n-1} - 1$ otherwise let $\Delta = \gamma_{n-1} - 2$.
Consider what happens when we change γ_{n-2} by the quantity Δ.
The increase of the number of squares is $\frac{\Delta}{2}\, q_{n-1}$, while the increase in the length of the word is $\Delta\, q_{n-1}$. The increase of squares is amortized by half of the increase of the length. Therefore we can subtract Δ from γ_{n-1}.

Observation

$$d(i) \;\leq\; \begin{cases} \left(\dfrac{3}{2}\,\gamma_i - 1\right) q_i + q_{i-1} - 1 & \text{for even } \gamma_i \\[2ex] \left(\dfrac{3}{2}\,\gamma_i - \dfrac{1}{2}\right) q_i & \text{for odd } \gamma_i \end{cases}$$

Lemma 10.
For $2 \leq r \leq n-3$ we have

$$\sum_{i=0}^{r} d(i) \;<\; \frac{3}{2}\, q_{r+1} + q_r.$$

Proof.
According to the observation above and implication

$$\gamma_i \geq 2 \quad \Rightarrow \quad q_{i-1} - q_i < -\,\frac{1}{2}\, q_i,$$

we have:

$$d(i) \leq \frac{3}{2}\, \gamma_i\, q_i - \frac{1}{2}\, q_i.$$

Observe now that $\gamma_i\, q_i = q_{i+1} - q_{i-1}$. Hence for $r \geq 2$:

$$\sum_{i=1}^{r} \gamma_i\, q_i \;=\; q_{r+1} + q_r - q_0 - q_1.$$

Consequently

$$\begin{aligned}
\sum_{i=0}^{r} d(i) \;&<\; d(0) + \frac{3}{2} \sum_{i=1}^{r} \gamma_i\, q_i \;-\frac{1}{2}\, q_r \\
&\leq\; d(0) + \frac{3}{2}\, \Big(q_{r+1} + q_r - q_0 - q_1\Big) - \frac{1}{2}\, q_r \\
&\leq\; \frac{3}{2}\, q_{r+1} + q_r.
\end{aligned}$$

This completes the proof.

Now we are ready to prove the tight bound for the number of squares in standard Sturmian words.

Theorem 11.
Let $\mathrm{Sw}(\gamma_0, \gamma_1, \ldots, \gamma_n)$ *be a standard Sturmian word. Then*

$$sq(\gamma_0, \gamma_1, \ldots, \gamma_n) \;\leq\; \frac{9}{10} \cdot \Big|\mathrm{Sw}(\gamma_0, \gamma_1, \ldots, \gamma_n)\Big| + 4.$$

Proof.
First assume that:

$$\boxed{\gamma_n = 1 \quad \text{and} \quad \gamma_{n-1} \in \{1, 2\}}$$

Let us shorten the notation and denote:

$$A = q_{n-2}, \quad B = q_{n-3}, \quad \alpha = \gamma_{n-2}.$$

We have, due to Lemma 10, the following fact (in terms of A and B):

Claim 1.

$$\sum_{i=0}^{n-3} sq_i(\gamma) \;=\; \sum_{i=0}^{n-3} d(i) \;\leq\; \frac{3}{2}\,A+B.$$

This, together with the fact that $sq_n(\gamma_0,\gamma_1,\ldots,\gamma_{n-1},1)=0$, implies:

Claim 2.

$$\boxed{sq(\gamma) \;\leq\; \Phi(\gamma) \;\stackrel{def}{=}\; \frac{3}{2}\,A+B+sq_{n-1}(\gamma)+sq_{n-2}(\gamma).}$$

Our goal is to prove the inequality

$$\Phi(\gamma) \leq \frac{9}{10}\,|w|.$$

Using our terminology we can write:

$$\begin{array}{|l|}\hline
\textbf{(a)}\ \left|\mathrm{Sw}(\gamma)\right| \;=\; \begin{cases} 2\,\alpha\,A+A+2B & \text{for}\ \ \gamma_{n-1}=1\\ 3\,\alpha\,A+A+3B & \text{for}\ \ \gamma_{n-1}=2\end{cases}\\
\textbf{(b)}\ sq_{n-2}(\gamma) \;\leq\; \begin{cases} \frac{3}{2}\,\alpha\,A-A & \text{for even}\ \gamma_{n-2}\\ \frac{3}{2}\,\alpha\,A-\frac{3}{2}A+B+1 & \text{for odd}\ \gamma_{n-2}\end{cases}\\
\textbf{(c)}\ sq_{n-1}(\gamma) \;\leq\; \begin{cases} \alpha\,A+B & \text{for}\ \ \gamma_{n-1}=2\\ A-1 & \text{for}\ \ \gamma_{n-1}=1\end{cases}\\ \hline
\end{array}$$

There are 4 cases depending on $\gamma_{n-1}\in\{1,2\}$ and the parity of α.

Case 1: ($\gamma_{n-1}=1$, α is even)

In this case inequality $\Phi(\gamma)\leq\frac{9}{10}\,|w|$ reduces to:

$$\boxed{\frac{3}{2}\Big(\alpha+1\Big)\ A+B \;\leq\; \frac{9}{10}\ \Big((2\,\alpha+1)\ A+2\ B\Big).}$$

This reduces to:

$$\frac{3}{2}\Big(\alpha+1\Big) \leq \frac{9}{10}\ \Big(2\,\alpha+1\Big).$$

The last inequality is reduced to $0.6 \leq 0.3\,\alpha$, which obviously holds for $\alpha\geq 2$.

This completes the proof of this case.

Case 2: ($\gamma_{n-1} = 1$, α is odd)

In this case the inequality $\Phi(\gamma) \leq \frac{9}{10}\,|w|$ reduces to:

$$\boxed{\left(\frac{3}{2}\,\alpha{+}1\right)\,A{+}2\,B \leq \frac{9}{10}\,\Big((2\,\alpha{+}1)\,A{+}2\,B\Big),}$$

which holds for $\alpha \geq 1,\; A > B > 0$.

Case 3: ($\gamma_{n-1} = 2$, α is even)

In this case

$$\Phi(\gamma) \leq \left(\frac{5}{2}\,\alpha + \frac{1}{2}\right)\,A + 2\,B.$$

Consequently the inequality $\Phi(\gamma) \leq \frac{9}{10}\,|w|$ reduces to:

$$\boxed{\left(\frac{5}{2}\,\alpha{+}\frac{1}{2}\right)\,A{+}2\,B \;\leq\; \frac{9}{10}\,\Big(3\,\alpha\,A{+}A{+}3B\Big).}$$

This holds since $\alpha \geq 2,\; A > B > 0$.

Case 4: ($\gamma_{n-1} = 2$, α is odd)

In this case

$$\Phi(\gamma) \;\leq\; \frac{5}{2}\,\alpha\,A + 3\,B + 1.$$

Now the inequality $\Phi(\gamma) \leq \frac{9}{10}\,|w|$ reduces to:

$$\boxed{\frac{5}{2}\,\alpha\,A + 3\,B + 1 \;\leq\; \frac{9}{10}\,\Big(3\,\alpha\,A + A + 3B\Big).}$$

This holds for $\alpha \geq 1,\; A > B > 0$.

We proved that

$$sq(\gamma_0, \gamma_1, \ldots, \gamma_{n-2}, 1, 1) \;\leq\; \frac{9}{10}\,|\mathrm{Sw}(\gamma_0, \gamma_1, \ldots, \gamma_{n-2}, 1, 1)|$$

and

$$sq(\gamma_0, \gamma_1, \ldots, \gamma_{n-2}, 2, 1) \;\leq\; \frac{9}{10}\,|\mathrm{Sw}(\gamma_0, \gamma_1, \ldots, \gamma_{n-2}, 2, 1)|.$$

This implies, that in general case, due to Lemma 8 and Lemma 9, we have

$$sq(\gamma_0, \gamma_1, \ldots, \gamma_n) \;\leq\; \frac{9}{10}\,|\mathrm{Sw}(\gamma_0, \gamma_1, \ldots, \gamma_n)| + 4,$$

which completes the proof of the theorem.

6 Final remarks

The *maximal repetition* (the *run*, in short) in a word w is a nonempty subword $w[i..j] = u^k v$ ($k \geq 2$), where u is of the minimal length and v is proper prefix (possibly empty) of u, that can not be extended (neither $x[i-1..j]$ nor $x[i..j+1]$ is a run with period $|u|$).

Let $\rho(w)$ be the number of runs in the word w. For n-th Fibonacci word F_n we have:

$$\begin{aligned} sq(F_n) &= 2|F_{n-2}| - 2, \\ \rho(F_n) &= 2|F_{n-2}| - 3, \end{aligned}$$

hence $sq(F_n) = \rho(f_n) + 1$, see [12,21].

For standard Sturmian words the situation is different. We have:

$$\frac{\rho(w)}{|w|} \longrightarrow 0.8 \qquad \text{and} \qquad \frac{sq(w)}{|w|} \longrightarrow 0.9,$$

see [3] for more details.

The maximal number of runs is reached for the standard Sturmian words of the form $v_k = \mathrm{Sw}(1, 2, k, k)$. Using the formulas (1-6) from the section 3 we have:

$$sq(v_k) = \begin{cases} \frac{5}{2}k^2 + \frac{5}{2}k + 4 & \text{for even } k \\ \frac{5}{2}k^2 + 5k - \frac{5}{2} & \text{for odd } k \end{cases}$$

and

$$|v_k| = 5k^2 + 2k + 5,$$

consequently

$$\frac{sq(v_k)}{|v_k|} \longrightarrow \frac{1}{2}.$$

We have shown in the section 4 that the maximal number of squares is achieved for the Sturmian words of the form $w_k = \mathrm{Sw}(k, k, 2, 1, 1)$. Now we compute the number of runs for w_k using formulas from [3]. We have:

$$\rho(w_k) = 9k + 7$$

and

$$|w_k| = 5k^2 + 7k + 7,$$

hence

$$\frac{\rho(w_k)}{|w_k|} \longrightarrow 0.$$

The results above show that the maximal number of squares and the maximal number of runs for standard Sturmian words are not closely related. The asymptotical limits are close, but both are reached for different type of words.

References

1. J. Allouche and J. Shallit: *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge University Press, 2003.
2. A. Apostolico and F. P. Preparata: *Optimal off-line detection of repetitions in a string.* Theor. Comput. Sci., 22 1983, pp. 297–315.
3. P. Baturo, M. Piątkowski, and W. Rytter: *The number of runs in Sturmian words*, in CIAA 2008, 2008, pp. 252–261.
4. P. Baturo, M. Piątkowski, and W. Rytter: *Usefulness of directed acyclic subword graphs in problems related to standard Sturmian words*, in PSC 2008, 2008, pp. 193–207.
5. P. Baturo and W. Rytter: *Occurrence and lexicographic properties of standard Sturmian words*, LATA, 2007.
6. M. Crochemore: *An optimal algorithm for computing the repetitions in a word.* Inf. Process. Lett., 12(5) 1981, pp. 244–250.
7. M. Crochemore and L. Ilie: *Maximal repetitions in strings.* J. Comput. Syst. Sci., 74(5) 2008, pp. 796–807.
8. M. Crochemore, L. Ilie, and L. Tinta: *Towards a solution to the "runs" conjecture*, in CPM, P. Ferragina and G. M. Landau, eds., vol. 5029 of Lecture Notes in Computer Science, Springer, 2008, pp. 290–302.
9. M. Crochemore and W. Rytter: *Squares, cubes, and time-space efficient string searching.* Algorithmica, 13(5) 1995, pp. 405–425.
10. M. Crochemore and W. Rytter: *Jewels of Stringology*, World Scientific, 2003.
11. D. Damanik and D. Lenz: *Powers in Sturmian sequences.* Eur. J. Comb., 24(4) 2003, pp. 377–390.
12. A. S. Fraenkel and J. Simpson: *The exact number of squares in Fibonacci words.* Theor. Comput. Sci., 218(1) 1999, pp. 95–106.
13. A. S. Freankel and J. Simpson: *How many squares can a string contain?* J. of Combinatorial Theory Series A, 82 1998, pp. 112–120.
14. M. Giraud: *Not so many runs in strings*, in LATA, C. Martín-Vide, F. Otto, and H. Fernau, eds., vol. 5196 of Lecture Notes in Computer Science, Springer, 2008, pp. 232–239.
15. L. Ilie: *A simple proof that a word of length n has at most $2n$ distinct squares.* J. of Combinatorial Theory Series A, 112 2005, pp. 163–164.
16. L. Ilie: *A note on the number of squares in a word.* Theoretical Computer Science, 380 2007, pp. 373–376.
17. C. S. Iliopoulos, D. Moore, and W. F. Smyth: *A characterization of the squares in a Fibonacci string.* Theor. Comput. Sci., 172(1-2) 1997, pp. 281–291.
18. J. Karhumaki: *Combinatorics on words*, Notes in PDF, `http://www.math.utu.fi/en/home/karhumak/`.
19. R. M. Kolpakov and G. Kucherov: *Finding maximal repetitions in a word in linear time*, in FOCS, 1999, pp. 596–604.
20. R. M. Kolpakov and G. Kucherov: *On maximal repetitions in words*, in FCT, G. Ciobanu and G. Paun, eds., vol. 1684 of Lecture Notes in Computer Science, Springer, 1999, pp. 374–385.
21. R. M. Kolpakov and G. Kucherov: *On maximal repetitions in words*, in FCT '99: Proceedings of the 12th International Symposium on Fundamentals of Computation Theory, London, UK, 1999, Springer-Verlag, pp. 374–385.
22. M. Lothaire: *Applied Combinatorics on Words*, Cambridge University Press, Cambridge, UK, 2005.
23. M. G. Main: *Detecting leftmost maximal periodicities.* Discrete Applied Mathematics, 25(1–2) 1989, pp. 145–153.
24. M. G. Main and R. J. Lorentz: *An $O(n \log n)$ algorithm for finding all repetitions in a string.* J. Algorithms, 5(3) 1984, pp. 422–432.
25. S. J. Puglisi, J. Simpson, and W. F. Smyth: *How many runs can a string contain?* Theor. Comput. Sci., 401(1-3) 2008, pp. 165–171.
26. W. Rytter: *The number of runs in a string: Improved analysis of the linear upper bound*, in STACS, B. Durand and W. Thomas, eds., vol. 3884 of Lecture Notes in Computer Science, Springer, 2006, pp. 184–195.
27. W. Rytter: *The number of runs in a string.* Inf. Comput., 205(9) 2007, pp. 1459–1469.
28. A. Thue: *Über unendliche Zeichenreihen.* Norske Vid. Selsk. Skr. I Math-Nat., 7 1906, pp. 1–22.