Combinatorics on Words

Gabriele Fici

CWI, Amsterdam — April 2024

Part 5: Sturmian words

Sturmian Words

Sturmian words are words with factor complexity function $f(n) = n + 1$ for every n. They are binary words (since they have 2 distinct factors of length 1) and we will consider them over the alphabet $\Sigma_2 = \{0, 1\}$. They are aperiodic word by the theorem of Morse and Hedlund.

Examples of Sturmian words are the Fibonacci word

$$f = 0100101001001010010100100101001001010010100100101001010\cdots$$

and the Pell word

$$pl = 001001000100100010010010001001000100100100010010001001\cdots$$

The Thue–Morse word $t = 0110100110010110\cdots$, instead, is not Sturmian, as it contains 4 distinct factors of length 2.

Sturmian Words

Sturmian words have several equivalent definitions. We will outline some of them in the sequel.

We begin by stating some general properties of Sturmian words.

Theorem 1

The following properties hold:

1. *There is an uncountable number of Sturmian words;*
2. *Every Sturmian word is uniformly recurrent;*
3. *The set of factors of Sturmian words is closed under reversal;*
4. *Every Sturmian word has infinitely many palindromic prefixes;*
5. *Every Sturmian word has infinitely many square prefixes.*

Sturmian Words

Recall that the factor recurrence function of an infinite word x is defined as

$$R''_x(n) = \inf\{|v| \mid v \text{ is a factor and every factor of length } n \text{ occurs in } v\}$$

The first characterization of Sturmian words is the following:

Theorem 2

An infinite word x is Sturmian if and only if $R''_x(n) = 2n$ for every $n \geq 0$.

As a consequence, in a Sturmian word there is a factor of length $2n$ containing all the $n + 1$ distinct factors of length n, for every n.

Exercise 3

Prove Theorem 2.

Balance

Definition 4

A finite or infinite word w over Σ_2 is balanced if and only if for every factors of w of the same length u and v, one has $||u|_0 - |v|_0| \leq 1$.

That is, the difference between the number of 0s (or, equivalently, 1s) in two factors of the same length is at most one.

The following is a characterization of Sturmian words based on the previous definition:

Theorem 5

An infinite word over Σ_2 is Sturmian if and only if it is balanced and aperiodic.

Balance

A natural way to extend the balance property to alphabets of more than two letters is the following:

Definition 6

A finite or infinite word w over Σ_k, $k \geq 2$, is balanced on each letter (or, simply, balanced) if for all pairs u and v of factors of w of the same length, one has $||u|_i - |v|_i| \leq 1$ for every letter i.

Infinite non-periodic words balanced on each letter can be constructed by a modification of Sturmian words. This construction is due to Hubert.

Balance

We need to consider infinite periodic words with constant gaps. An infinite periodic word w^ω has constant gaps if the distance between two successive occurrences of letter a_i of w^ω is constant, for each i.

For example $(abac)^\omega$ is a constant gap word while $(abaac)^\omega$ is not.

Now, take an infinite periodic word with constant gaps $(w_1)^\omega$ over some alphabet A and an another infinite periodic word with constant gaps $(w_2)^\omega$ over some alphabet B disjoint from A.

Let x be a Sturmian word over Σ_2. If one replaces every occurrence of 0 in x by letters of $(w_1)^\omega$ and every occurrence of 1 in x by letters of $(w_2)^\omega$ then one obtains an infinite aperiodic word balanced on each letter.

Balance

As an example, we build a non-periodic word over a five-letter alphabet by modification of the Fibonacci word $f = 0100101001001010010\cdots$.

To do this, we replace periodically the occurrences of the letter 0 by the constant gap word $(cdce)^\omega$:

$$c1dc1e1cd1ce1c1dc1e1cd1ce1c1\cdots$$

By construction, the word is balanced on the letter 1, because the Fibonacci word is balanced. It remains as an exercise to check that the word is balanced on the letters c, d and e.

We can now replace periodically the occurrences of the letter 1 by the constant gap word $(ab)^\omega$:

$$cabdcabeabcdabceabcabdcabeabcdabceabcab\cdots$$

thus obtaining an aperiodic balanced word over five letters.

Balance

Notice that in the previous example, the frequencies of the letters are not all distinct. Actually, we have the following:

Conjecture 7 (Fraenkel)

For every $k > 2$, there is only one infinite balanced word (up to letter permutation) over an alphabet of size k, in which all letters have different frequencies, and this word is periodic.

The only infinite balanced word with different frequencies in the statement of the theorem is the periodic word $(z_k)^\omega$, where z_k is the kth Zimin word (recall that Zimin words are defined recursively by $z_0 = 0$ and $z_n = z_{n-1}(n)z_{n-1}$, for $n > 0$).

The infinite periodic words of the form $(z_k)^\omega$ are sometimes called Fraenkel words.

Balance

The statement of the Fraenkel conjecture is not true for $k = 2$ since there are aperiodic balanced binary words with different letter frequencies, namely the Sturmian words.

The Fraenkel conjecture has been proved true for small alphabet sizes (up to 7), but it remains open for larger alphabet sizes.

Balance

The notion of balanced word can be generalized to the notion of C-balanced word, as follows:

Definition 8

A finite or infinite word w over Σ_2 is C-balanced for an integer $C \geq 1$ if and only if for every factors of w of the same length u and v, one has $||u|_0 - |v|_0| \leq C$, that is, the number of 0s (or, equivalently, 1s) in two factors of the same length differ at most by C.

Of course, a balanced word is just a C-balanced word for $C = 1$.

The Tribonacci word is not 1-balanced (otherwise the Fraenkel conjecture would be false for $k = 3$) but it is 2-balanced.

Cassaigne, Ferenczi and Zamboni proved that there are Arnoux–Rauzy words that are not C-balanced for any C.

Every fixed point of a Pisot morphisms is C-balanced for some $C > 0$.

Palindromes

Another characterization of Sturmian words is related to palindromes.

Unlike the factor complexity, the unboundedness of the palindromic complexity does not characterize aperiodic words. Indeed, we have the following characterization of Sturmian words in terms of the palindromic complexity:

Theorem 9 (Droubay, Pirillo, 1999)

An infinite word is Sturmian if and only if it has palindromic complexity 1 for every even n and 2 for every odd n.

Let us look at the Fibonacci word f. The first factors of f that are palindromes are, in increasing length order: ε, 0, 1, 00, 010, 101, 1001, 01010, 00100, etc.

Palindromes

More generally, we have the following

Theorem 10 (Damanik, Zamboni, 2003)

The palindromic complexity of any Arnoux–Rauzy word over Σ_k is 1 for every even n and k for every odd n.

However, this is no longer a characterization, as there exist words with the same palindromic complexity that are not Arnoux–Rauzy.

Exercise 11

Prove Theorem 10. Hint: by induction on n.

Palindromes

Another property of Sturmian words related to palindromes is that they are rich words.

Theorem 12

Every Sturmian word is rich. That is, any factor of length n of a Sturmian word contains exactly n distinct nonempty palindromes.

Palindromes

Given a finite word w, one can also consider the minimum number of palindromes in which it is possible to decompose w. This measure is called palindromic length.

For example, 00101100 can be decomposed in $00 \cdot 101 \cdot 1 \cdot 00$, but it is not possible to decompose it in less than 4 palindromes, so its palindromic length is 4.

The palindromic length complexity of an infinite word x is the function that counts, for each n, the maximum of the palindromic lengths among the factors of x of length n.

It has been proved that the palindromic length of Sturmian words is unbounded.

The following conjecture is still open:

Conjecture 13 (Frid, Puzynina, Zamboni, 2013)

An infinite word is aperiodic if and only if it has unbounded palindromic length complexity.

Return Words

Another characterization of Sturmian words is related to the notion of return words.

Theorem 14 (Vuillon, 2001)

An infinite word is Sturmian if and only if every of its factors has exactly two returns.

This is once again tight, because an infinite recurrent word x is (purely) periodic if and only if there is a factor of x that has only one return in x.

Indeed, if $x = w^\omega$, and w is primitive, then w has itself as the only return. Conversely, if v is a factor of x with only one return, say w, then $x = pw^\omega$ for some p, i.e., x is purely periodic or ultimately periodic.

But since x is supposed to be recurrent, we have that x is purely periodic.

Remark 15

As a consequence, the derived word of a Sturmian word is again a Sturmian word.

Slope

Given a nonempty word w over Σ_2, we define the slope of w as the number $|w|_1/|w|$, that is, the frequency of the letter 1 in the word.

For an infinite word x over Σ_2, the slope is defined as the limit of the slopes of its prefixes, when this limit exists.

Every Sturmian word has a slope, that is an irrational number α between 0 and 1.

Conversely, for any irrational α between 0 and 1 there exist (uncountable many) Sturmian words with slope α (we will see this more in detail later).

Slope

For example, the slope of the Fibonacci word f is $1/\varphi^2 = 2 - \varphi \approx 0.382$, where $\varphi = (1 + \sqrt{5})/2$ is the golden ratio.

Indeed, f is the limit of the sequence of the Fibonacci finite words, and it is easy to see that for every n the Fibonacci finite word f_n, of length F_n — the nth Fibonacci number — contains exactly F_{n-1} 0's and F_{n-2} 1's. So the slope of f is $\lim_{n\to\infty} F_{n-2}/F_n = 1/\varphi^2 = 2 - \varphi$.

The slope of the complement of the Fibonacci word

$$\bar{f} = 1011010110110101 10\cdots$$

is $\lim_{n\to\infty} F_{n-1}/F_n = 1/\varphi = \varphi - 1 \approx 0.618$.

The slope of the Pell word is $1 - \sqrt{2}/2 \approx 0.292893$.

Slope

Theorem 16

Two Sturmian words have the same slope if and only if they have the same factors.

Therefore, the shift orbit closure of a Sturmian word is made by all Sturmian words that have the same slope α.

Characteristic Sturmian Words

An important class of Sturmian words is that of characteristic (or standard) Sturmian words.

Definition 17

A Sturmian word x is characteristic if $0x$ and $1x$ are Sturmian words.

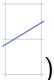
For any irrational α between 0 and 1 there exists exactly one characteristic Sturmian word of slope α, whence the name characteristic.

Remark 18

A Sturmian word is characteristic if and only if the set of its left special factors coincides with the set of its prefixes.

Characteristic Sturmian Words

What is the best digital approximation of a straight line of equation $y = \alpha x + \rho$? (We can suppose α between 0 and 1.)

All we need to know is, for every n, whether the portion of the straight line having x-coordinate between n and $n+1$ crosses a horizontal line of the grid (like this: ) or not (like this: ). We then code this information writing 1 for cross, 0 for not cross.

A problem can arise when the straight line intersects the grid, but for our purposes we can restrict α to be an irrational number (so that the grid is intersected for exactly one value of ρ) and ρ to be such that the grid is not intersected.

We have defined the mechanical word $s_{\alpha,\rho}$ of slope α and intercept ρ.

Characteristic Sturmian Words

By definition $s_{\alpha,\rho} = s_{\alpha,\rho+k}$ for any integer k, so we can suppose $\rho \in [0, 1)$. The case $\rho = 0$ is also degenerate. Actually, there are two possible choices for a mechanical word with intercept 0, one beginning with 0 and the other beginning with 1. We consider both of them as mechanical words. With this precaution, the following result holds:

Theorem 19

An infinite word over Σ_2 is a Sturmian word if and only if it is a mechanical word of irrational slope.

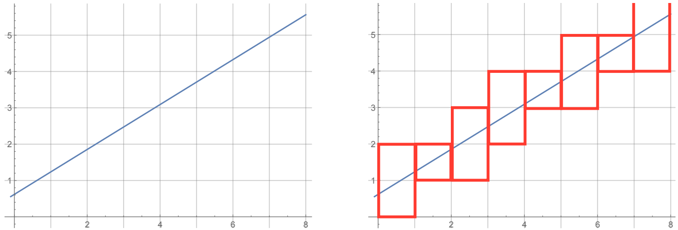


Figure: The complement of the Fibonacci word $1011010110110\cdots$ as a mechanical word of slope $1/\varphi = \varphi - 1$ and intercept $1/\varphi = \varphi - 1$.

Characteristic Sturmian Words

The characteristic Sturmian word of slope α is the Sturmian word $s_{\alpha,\alpha}$. Indeed, given a characteristic Sturmian word $s_{\alpha,\alpha}$, the words $0s_{\alpha,\alpha}$ and $1s_{\alpha,\alpha}$ are the two mechanical words of slope α and intercept 0.

For example, when $\alpha = 2 - \varphi = 1/\varphi^2 \approx 0.382$, we get the Fibonacci word f; when $\alpha = 1 - \sqrt{2}/2 \approx 0.292893$, we get the Pell word.

Remark 20

The characteristic Sturmian word $s_{1-\alpha,1-\alpha}$ is obtained from $s_{\alpha,\alpha}$ by applying the automorphism of Σ_2 that exchanges 0 and 1.

For example, if $\alpha = 2 - \varphi = 1/\varphi^2$, one has $1 - \alpha = \varphi - 1 = 1/\varphi$. So, the characteristic word of slope $1/\varphi = \varphi - 1 \approx 0.618$ is the complement of the Fibonacci word:

$$\bar{f} = 1011010110110101 10\cdots$$

Characteristic Sturmian Words

Theorem 21

An aperiodic infinite word x over Σ_2 is a characteristic Sturmian word if and only if $0x$ is lexicographically smaller than any infinite suffix of x and $1x$ is lexicographically greater than any infinite suffix of x.

Equivalently, take an aperiodic infinite word $x = x_0x_1x_2\cdots$ and define another infinite word $x' = x'_0x'_1x'_2\cdots$ by

$$x'_n = \begin{cases} 0 \text{ if } x < x_{n+1}x_{n+2}\cdots; \\ 1 \text{ otherwise.} \end{cases}$$

The previous theorem states that the word x' is equal to x if and only if x is a characteristic Sturmian word.

For example, let $x = 010010100100101\cdots$ be the Fibonacci word. Then $x'_0 = 0$ since x is smaller than $10010100100101\cdots$, $x'_1 = 1$ since x is greater than $0010100100101\cdots$, $x'_2 = 0$ since x is smaller than $010100100101\cdots$, $x'_3 = 0$ since x is smaller than $10100100101\cdots$, etc.

Beatty Sequences

A Beatty sequence is the sequence of integers found by taking the floor of the positive multiples of a positive irrational number r. Note that if $r > 1$ the sequence has no repeated elements and forms a subset of the set of positive integers.

Formally, an irrational number r generates the Beatty sequence

$$\mathcal{B}_r = \lfloor r \rfloor, \lfloor 2r \rfloor, \lfloor 3r \rfloor, \ldots$$

If $r > 1$, the number $s = r/(r-1)$ is also an irrational number greater than 1. These two numbers naturally satisfy the equation $1/r + 1/s = 1$.

The two Beatty sequences they generate, $\mathcal{B}_r = (\lfloor nr \rfloor)_{n \geq 1}$ and $\mathcal{B}_s = (\lfloor ns \rfloor)_{n \geq 1}$ form a pair of complementary Beatty sequences, in the sense that every positive integer belongs to exactly one of these two sequences. (This is in fact a famous theorem due to Rayleigh.)

For example, let us take $r = \varphi$, so that $s = \varphi/(\varphi - 1) = \varphi + 1$. In this case, the complementary Beatty sequences

$$\mathcal{B}_\varphi = 1, 3, 4, 6, 8, 9, 11, 12, 14, 16, 17, 19, 21, 22, 24, 25, 27, 29, \ldots$$

and

$$\mathcal{B}_{\varphi+1} = 2, 5, 7, 10, 13, 15, 18, 20, 23, 26, 28, 31, 34, 36, 39, 41, 44, \ldots$$

are called respectively the lower and the upper Wythoff sequences.

Remark 22

If we color the positive integers with two colors: 0 and 1, according to the fact that the integer belongs to the lower or to the upper Wythoff sequence, respectively, then we get the sequence $0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, \ldots$, which is in fact the Fibonacci word f.

As another example, take $r = \sqrt{2}$, so that $s = \sqrt{2}/(\sqrt{2} - 1) = 2 + \sqrt{2}$. In this case, the complementary Beatty sequences are:

$$\mathcal{B}_{\sqrt{2}} = 1, 2, 4, 5, 7, 8, 9, 11, 12, 14, 15, 16, 18, 19, 21, 22, 24, 25, 26, \ldots$$

and

$$\mathcal{B}_{2+\sqrt{2}} = 3, 6, 10, 13, 17, 20, 23, 27, 30, 34, 37, 40, 44, 47, 51, 54, 58, 61, \ldots$$

Remark 23

If we color the positive integers with two colors: 0 *and* 1*, according to the fact that the integer belongs to* $\mathcal{B}_{\sqrt{2}}$ *or to* $\mathcal{B}_{2+\sqrt{2}}$*, respectively, then we get the sequence* $0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, \ldots$*, which is in fact the Pell word* pl.

Beatty Sequences

Let $s_{\alpha,\rho} = x_0 x_1 \cdots$ be the Sturmian word of slope α and intercept ρ.

From the definition of mechanical word, the letter x_n is 1 if if we cross a horizontal line, i.e., when $\lfloor \rho + (n+1)\alpha \rfloor = \lfloor \rho + n\alpha \rfloor + 1$ and 0 otherwise, i.e., when $\lfloor \rho + (n+1)\alpha \rfloor = \lfloor \rho + n\alpha \rfloor$.

Therefore, we have this definition of Sturmian word: For every $n \geq 0$, the letter x_n is given by

$$x_n = \lfloor \rho + (n+1)\alpha \rfloor - \lfloor \rho + n\alpha \rfloor.$$

This becomes the sequence of consecutive differences of a Beatty sequence when $\rho = \alpha$, that is, when the Sturmian word is characteristic.

Beatty Sequences

Actually, for an irrational number $r > 1$ one can define the word $b_r = b_1 b_2 b_3 \cdots$ by

$$b_r(n) = \begin{cases} 1 & \text{if } n = \lfloor kr \rfloor \text{ for some integer } k, \\ 0 & \text{otherwise} \end{cases}$$

that is, the characteristic function of the subset of positive integers defined by the Beatty sequence $\mathcal{B}_r$.

The characteristic Sturmian word $s_{\alpha,\alpha}$ then coincides with the word $b_{1/\alpha}$.

Coding of Rotations

Equivalently, x_n is 1 if the fractional part of $\rho + (n+1)\alpha$ is smaller than the fractional part of $\rho + n\alpha$ and 0 otherwise.

Hence, if we consider the sequence $(\{\rho + n\alpha\})_{n\geq 0}$, where we write $\{\beta\} = \beta - \lfloor \beta \rfloor$ the fractional part of the real number β, then from the definition of mechanical word we have [1]

$$x_n = \begin{cases} 0 & \text{if } \{\rho + n\alpha\} \in [0, 1-\alpha), \\ 1 & \text{if } \{\rho + n\alpha\} \in [1-\alpha, 1). \end{cases}$$

The latter is the definition of Sturmian word as coding of a rotation on the unit circle $I = [0,1) \subset \mathbb{R}$ (sometimes also called one-dimensional torus).

[1] There is a technical point with this definition that we skipped for the sake of simplicity. Indeed, we know that both $0s_{\alpha,\alpha}$ and $1s_{\alpha,\alpha}$ are Sturmian words, but with the definition above we can only construct the word $0s_{\alpha,\alpha}$.

Indeed, consider a point initially in position ρ. Then rotate this point on the circle (clockwise) by the angles α, 2α, 3α, etc. and write consecutively the letters associated with the intervals the rotated points fall into. The sequence of letters obtained is the Sturmian word $s_{\alpha,\rho}$.

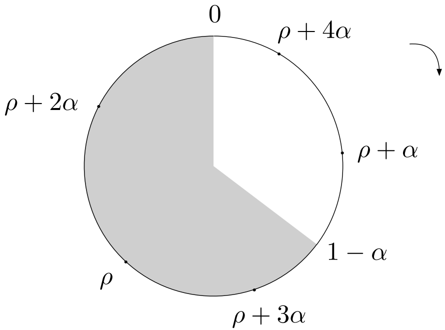


Figure: The rotation of the initial point $\rho = 1/\varphi = \varphi - 1 \approx 0.618$ by the angle $\alpha = 1/\varphi = \varphi - 1$ generating the complement of the Fibonacci word $\bar{f} = 1011010110110110 \cdots$.

Coding of Rotations

An equivalent view is to fix the point and rotate the intervals backwards. Let us write $I_0 = [0, 1-\alpha)$ and $I_1 = [1-\alpha, 1)$.

The interval $I_0 = I_0^0$ is rotated at each step, so that after i rotations it is transformed into $I_0^{-n} = I(\{-n\alpha\}, \{-(n+1)\alpha\})$, while $I_1^{-n} = I \setminus I_0^{-n}$.

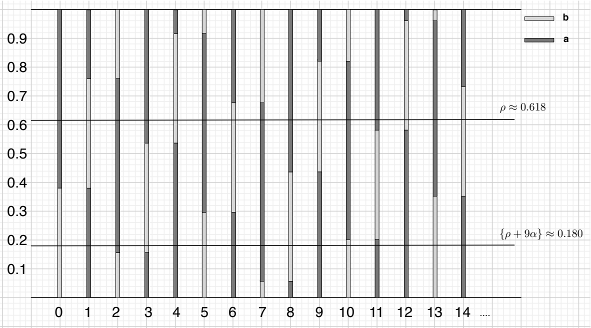


Figure: The interval $I_0 = I_0^0 = [0, 1-\alpha)$ is rotated at each step, defining the intervals $I_0^{-n} = [\{-n\alpha\}, \{-(n+1)\alpha\})$ (light gray) and I_1^{-n} (dark gray). The word $\bar{f}$ can be obtained by looking at the horizontal line of height $\rho = \alpha$.

Coding of Rotations

This representation is convenient since one can read within it not only a Sturmian word but also any of its finite factors. More precisely, for every positive integer m, the factor of length m of $s_{\alpha,\rho}$ starting at position n is determined only by the value of $\{\rho + n\alpha\}$.

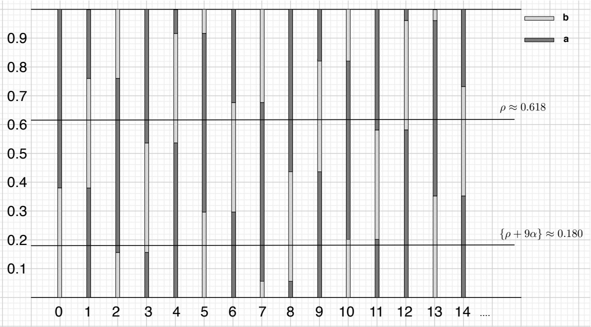


Figure: The factor of length 15 starting at position 9 of the word $\bar{f}$, namely 011010110101101, can be obtained by looking at the horizontal line of height $\{\rho + 9\alpha\}$.

Proposition 24

Let $s_{\alpha,\rho} = x_0x_1x_2\cdots$ be a Sturmian word. Then, for every n and i, we have:

$$x_{n+i} = \begin{cases} 0 & \textit{if } \{\rho + n\alpha\} \in I_0^{-i}; \\ 1 & \textit{if } \{\rho + n\alpha\} \in I_1^{-i}. \end{cases}$$

For example, suppose we want to know the factor of length 15 starting at position 9 in the word $\bar{f}$, the complement of the Fibonacci word. We have $\{\rho + 9\alpha\} = \{\varphi - 1 + 9(\varphi - 1)\} \approx 0.180$. The first few terms of the sequence $(\{-i\alpha\})_{i\geq 0}$ are, approximately, 0, 0.382, 0.764, 0.146, 0.528, 0.910, 0.292, 0.674, 0.056, 0.438, 0.820, 0.202, 0.584, 0.966, 0.348, 0.729. So we get $x_9x_{10}\cdots x_{23} = 011010110101101$.

Coding of Rotations

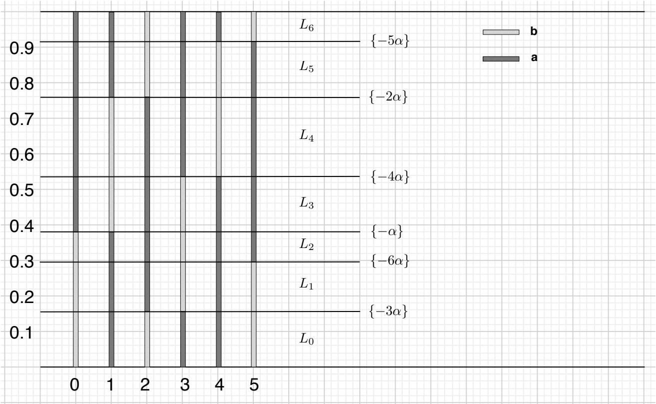
A remarkable consequence of Proposition 24 is the following: Given a Sturmian word $s_{\alpha,\rho}$ and a positive integer m, the $m+1$ different factors of $s_{\alpha,\rho}$ of length m are completely determined by the intervals $I_0^0, I_0^{-1}, \ldots, I_0^{-(m-1)}$, that is, only by the points $\{-i\alpha\}$ for $0 \leq i < m$.

In particular, they do not depend on the initial point ρ, so the set of factors of $s_{\alpha,\rho}$ is the same as the set of factors of $s_{\alpha,\rho'}$ for any ρ and ρ'.

Hence, from now on, we let s_α denote any Sturmian word of angle α.

Coding of Rotations

If we arrange the $m+2$ points $0, 1, \{-\alpha\}, \{-2\alpha\}, \ldots, \{-m\alpha\}$ in increasing order, we determine a partition of I in $m+1$ half-open subintervals $L_0(m), L_1(m), \ldots, L_m(m)$. Each of these subintervals is in bijection with a factor of length m of any Sturmian word of angle α. Moreover, the factors associated with these intervals are lexicographically ordered.



Coding of Rotations

The Three-Distance Theorem says that these intervals have at most 3 possible lengths, and if there are three lengths, then the longest one is the sum of the two others.

As a consequence of the Three-Distance Theorem, we have:

Theorem 25

Let s be a Sturmian word. For each n, the frequencies of factors of length n take at most three values. If they take three values, then one is the sum of the two others.

Recall that a Rote word is a word with factor complexity $2n$ for every n.

Examples of Rote words are the Stewart words and the fixed point of $0 \mapsto 001,\ 1 \mapsto 111$:

$$0010011110010011111111111110010011110010011111111111111\cdots$$

Theorem 26 (Rote, 1994)

Let $0 < \alpha < 1$ be irrational, ρ a real number and $\beta < 1$ such that $\alpha < \min(\beta, 1-\beta)$ and α and β are rationally independent (i.e., $n\alpha \mod 1 \neq \beta \mod 1$ for every $n > 0$). Define the word $x = x_0 x_1 \cdots$ coding of rotation of angle α, intercept ρ and intervals $[0, 1-\beta)$ and $[1-\beta, 1)$; that is, the word:

$$x_n = \begin{cases} 0 & \text{if } \{\rho + n\alpha\} \in [0, 1-\beta), \\ 1 & \text{if } \{\rho + n\alpha\} \in [1-\beta, 1). \end{cases}$$

Then x is a Rote word.

Sturmian Words and Rote Words

The converse is not true in general, as there are Rote words that cannot be obtained as coding of rotations. For example, those that are not uniformly recurrent.

Example 27

Let $\alpha = \rho = 1 - \sqrt{2}/2$ and $\beta = 1/\varphi^2$. We obtain the Rote word

$$x = 001001001100100110010010001001001100100100010010 0\cdots$$

Rote words closed under complement can be defined by means of a coding of a rotation with intervals of size $1/2$, i.e., a Rote word x is closed under complement if and only if it there exist ρ and α such that

$$x_n = \begin{cases} 0 & \text{if } \{\rho + n\alpha\} \in [0, 1/2), \\ 1 & \text{if } \{\rho + n\alpha\} \in [1/2, 1). \end{cases}$$

For example, for $\rho = \alpha = 1/\varphi^2$ we get the word

$$0101101001011010010100101101001011010110100101101001 01 \cdots$$

Note that for these words the frequencies of letters, if they exist, must be equal to $1/2$.

However, there are uniformly recurrent Rote words with frequencies of letters equal to $1/2$ that are not closed under complement, e.g., the Stewart choral word.

Sturmian Words and Rote Words

Given an infinite word x over Σ_2 we define $\sigma(x)$ as the word such that $\sigma(x)_i = (x_i + x_{i+1}) \mod 2$.

Theorem 28 (Rote, 1994)

Let x be a Rote word, i.e., a word with factor complexity $2n$ for every $n \geq 0$. Then x is closed under complement if and only if $\sigma(x)$ is a Sturmian word.

For example, the Rote word that gives the Fibonacci word is

$$\sigma^{-1}(f) = 0011100111000110\cdots$$

and it is, according to Theorem 28, closed under complement.

Standard Sequence

Characteristic Sturmian words can also be defined in terms of standard sequences. We first recall some classical results from elementary number theory.

Every irrational number α can be uniquely written as a (simple) continued fraction as follows:

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \dots}}$$

where $a_0 = \lfloor \alpha \rfloor$, the integer part of α, and the infinite sequence $(a_i)_{i \geq 0}$ is called the sequence of partial quotients of α.

The continued fraction expansion of α is usually denoted by its sequence of partial quotients as follows: $\alpha = [a_0; a_1, a_2, \dots]$, and each its finite truncation $[a_0; a_1, a_2, \dots, a_i]$ is a rational number p_i/q_i (we take p_i and q_i coprime) called the i-th convergent to α.

The sequence $(q_i)_{i \geq 0}$ can be defined by: $q_{-1} = 0$, $q_0 = 1$ and $q_n = a_n q_{n-1} + q_{n-2}$ for $n \geq 1$.

Standard Sequence

We say that an irrational $\alpha = [a_0; a_1, a_2, \ldots]$ has bounded partial quotients if the sequence $(a_i)_{i\geq 0}$ is bounded.

For example, we have $\varphi = [1; 1, 1, 1, \ldots]$. Indeed, this follows immediately from the equation $\varphi = 1 + \frac{1}{\varphi}$ that defines the golden ratio.

The convergents of φ are: 1, 2, $3/2$, $5/3$, $8/5$, $13/8$, $21/13$, $34/21$, etc., i.e., the numbers of the form F_i/F_{i+1}, where (F_n) is the sequence of positive Fibonacci numbers $1, 1, 2, 3, 5, 8, \ldots$

By the way, we have $1/\varphi = \varphi - 1 = [0; 1, 1, 1, \ldots]$ and $1/\varphi^2 = 2 - \varphi = [0; 2, 1, 1, \ldots]$.

Standard Sequence

As another example, we have $\sqrt{2} = [1; 2, 2, 2, 2, \dots]$ (this is why $1 + \sqrt{2} = [2; 2, 2, 2, 2, \dots]$ is called the silver ratio).

The convergents of $\sqrt{2}$ are: 1, $3/2$, $7/5$, $17/12$, $41/29$, $99/70$, $239/169$, $577/408$, $1393/985$, etc., i.e., the numbers of the form $(P_{k-1} + P_k)/P_k$, where (P_i) is the sequence of positive Pell numbers $1, 2, 5, 12, 29, 70, 169, 408, 985, 2378, 5741, \dots$ defined by $P_1 = 1$, $P_2 = 2$ and $P_n = 2P_{n-1} + P_{n-2}$ for $n > 2$.

An example of number whose continued fraction expansion is not ultimately periodic is

$$e = [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \dots, 1, 1, 2n, \dots]$$

Standard Sequence

Theorem 29

Let α be an irrational number. The continued fraction expansion of α is ultimately periodic if and only if α is quadratic, i.e., a solution of a quadratic equation with integer coefficients.

Remark 30

A long-standing conjecture of Khintchine (1949) states that the continued fraction expansion of an irrational algebraic number α is either ultimately periodic (and we know that this happens if and only if α is a quadratic irrational) or it contains arbitrarily large partial quotients.

Equivalently, an irrational number whose continued fraction expansion has bounded partial quotients is either quadratic or transcendental.

Standard Sequence

The characteristic Sturmian word $s_{\alpha,\alpha}$ of slope α, $0 < \alpha < 1$, is the limit of the sequence of words $(s_n)_{n\geq 0}$ defined recursively as follows:

Let $[0; d_0 + 1, d_1, d_2, \ldots]$ be the continued fraction expansion of α, and define $s_{-1} = 1$, $s_0 = 0$ and

$$s_{n+1} = s_n^{d_n} s_{n-1}$$

for every $n \geq 0$.

The sequence s_n is called the standard sequence of $s_{\alpha,\alpha}$.

Standard Sequence

Note that $s_{\alpha,\alpha}$ starts with letter 1 if and only if $\alpha > 1/2$, i.e., if and only if $d_0 = 0$. In this case, $[0; d_1 + 1, d_2, \ldots]$ is the continued fraction expansion of $1 - \alpha$, and $s_{1-\alpha,1-\alpha}$ is the word obtained from $s_{\alpha,\alpha}$ by exchanging 0's and 1's.

For example, the slope of the complement of the Fibonacci word $\overline{f}$ is $1/\varphi = \varphi - 1 = [0; 1, 1, 1, \ldots]$. Hence, $d_0 = 0$ and $d_i = 1$ for every $i > 0$. We therefore get $s_1 = 1$, $s_2 = 10$, $s_3 = 101$, $s_4 = 10110$, etc. That is, we obtain the sequence of the complements of the Fibonacci finite words.

On the other hand, the slope of the Fibonacci word f is $1/\varphi^2 = 2 - \varphi = [0; 2, 1, 1, \ldots]$. Hence, $d_i = 1$ for every $i \geq 0$. We therefore get $s_{-1} = 1$, $s_0 = 0$, $s_1 = 01$, $s_2 = 010$, $s_3 = 01001$, $s_4 = 01001010$, etc. That is, we obtain the sequence of the Fibonacci finite words.

Standard Sequence

The slope of the Pell word pl is $1 - \sqrt{2}/2 = [0; 3, 2, 2, 2, \ldots]$. Hence, $d_i = 2$ for every $i \geq 0$.

We therefore get $s_{-1} = 1$, $s_0 = 0$, $s_1 = 0^2 1$, $s_2 = 0^2 1 0^2 1 0$, $s_3 = 0^2 1 0^2 1 0 \cdot 0^2 1 0^2 1 0 \cdot 0^2 1$, etc.

$$pl = 001001000100100010010010001001000100100100010010 \cdots$$

Standard Sequence

Definition 31

A standard word is a word that appears in some standard sequence.

Proposition 32

A word is standard if and only if it is of the form $s_n = p_n ab$, where p_n is a central word (hence a palindrome), and a and b are different letters.

Hence, central words are precisely the palindromic prefixes of Sturmian words.

Ostrowski Numeration Systems

Fibonacci numbers can be used as a basis for representing integers.

Indeed, we have the following theorem, known as Zeckendorf theorem, but published earlier by Lekkerkerker and which, in fact, is a special case of an older and more general result due to Ostrowski (1922).

Theorem 33

Every positive integer can be expressed uniquely as the sum of one or more distinct non-consecutive Fibonacci numbers F_n, $n > 1$.

For example, $17 = 13 + 3 + 1 = F_7 + F_4 + F_2$ and there is no other way to write 17 as the sum of non-consecutive Fibonacci numbers (assuming the convention that F_1 is not used in the representation).

Ostrowski Numeration Systems

Thus, one can represent natural numbers as sequences of 0-1 bits, where the i-th bit (from the right) encodes the presence/absence of the $(i+1)$-th Fibonacci number in the representation given by Theorem 33.

So for example the number 17 is represented by 100101. We call this representation of natural numbers the Zeckendorf representation.

Zeck.	decimal	Zeck.	decimal	Zeck.	decimal
000000	0	010000	8	100100	16
000001	1	010001	9	100101	17
000010	2	010010	10	101000	18
000100	3	010100	11	101001	19
000101	4	010101	12	101010	20
001000	5	100000	13		
001001	6	100001	14		
001010	7	100010	15		

Let us define $f(n)$, for every $n \geq 0$, as the rightmost digit of the Zeckendorf representation of n.

For every $n > 1$ the word $f_n = f(0)f(1)\cdots f(F_n - 1)$ is indeed the nth Fibonacci word.

Hence, the Fibonacci word f can be obtained by concatenating the last digits of the Zeckendorf representations of integers.

Let us now see how this generalizes to any characteristic Sturmian word.

Ostrowski Numeration Systems

Take an irrational α between 0 and $1/2$ [2]. Instead of Fibonacci numbers, we consider the denominators q_i of the convergents of α. In fact, the Fibonacci numbers are the denominators of the convergents of $1/\varphi^2$, which is the slope of the Fibonacci word.

For example, let $\alpha = \frac{1}{e+1} \approx 0.269 = [0; 3, 1, 2, 1, 1, 4, 1, 1, 6, \ldots]$. The first few convergents are $\frac{1}{3}, \frac{1}{4}, \frac{3}{11}, \frac{4}{15}, \frac{7}{26}, \frac{32}{119}, \frac{39}{145}$, etc.

Recall that we can define the sequence $(q_i)_{i\geq 0}$, of the denominators of the convergents to α, by: $q_{-1} = 0$, $q_0 = 1$ and $q_n = a_n q_{n-1} + q_{n-2}$ for $n \geq 1$.

The sequence $(q_i)_{i\geq 0} = 1, 3, 4, 11, 15, 26, 119, 145, \ldots$ can be used as a basis for representing integers.

[2] We suppose here α between 0 and $1/2$ in order to obtain a word beginning with 0, otherwise the procedure is analogous but we obtain a word beginning with 1.

Indeed, any integer n can be represented in a unique way as $n = \sum_{i=1}^{t} c_i q_i$ by taking the largest c_i in the division of the remainder by q_i, starting from the largest q_i smaller than or equal to n.

Equivalently, the same representation can be obtained by subtracting the largest possible q_i from n and repeating, until one has $n = \sum_{i=1}^{t} c_i q_i$.

For example, $23 = 1 \cdot 15 + 2 \cdot 4$, so its representation in basis α is 10200.

This representation is usually called the Ostrowski representation of integers with basis (q_i).

Ostrowski Numeration Systems

An example of the representation of the first few integers using as a basis the denominators of the convergents of $\alpha = \frac{1}{e+1}$ is given below.

Ostrowski	decimal	Ostrowski	decimal	Ostrowski	decimal
000000	0	000200	8	010001	16
000001	1	000201	9	010002	17
000002	2	000202	10	010010	18
000010	3	001000	11	010100	19
000100	4	001001	12	010101	20
000101	5	001002	13		
000102	6	001010	14		
000110	7	010000	15		

Table: The Ostrowski representations in basis $\alpha = \frac{1}{e+1} \approx 0.269$ of the first few natural numbers coded with 6 bits.

Note that the last digit in the Ostrowski representation of n ranges from 0 to $q_1 - 1$.

Now, we can obtain the characteristic Sturmian word (and its standard sequence) of slope α in the following way:

Let π be the map that sends $q_1 - 1$ to 1 and all other letters to 0. Apply π to the sequence of last digits of the Ostrowski representations of integers.

In our example, we have $q_1 - 1 = 3 - 1 = 2$, so $\pi : 0, 1 \mapsto 0, 2 \mapsto 1$, whence

$$s_{\frac{1}{e+1}, \frac{1}{e+1}} = \pi(01200120012012001200 1 \cdots) = 001000100010010001000 \cdots$$

Ostrowski Numeration Systems

As another example, with $\alpha > 1/2$, let
$\alpha = \frac{e+1}{e+2} \approx 0.788 = [0; 1, 3, 1, 2, 1, 1, 4, 1, 1, \ldots]$.

The first few convergents are $1, \frac{3}{4}, \frac{4}{5}, \frac{11}{14}, \frac{15}{19}, \frac{26}{33}, \frac{119}{151}$, etc.

The sequence $(q_i)_{i \geq 0} = 1, 4, 5, 14, 19, 33, 151, \ldots$ of denominators of the convergents can be used as a basis for representing any integer.

For example, $23 = 1 \cdot 19 + 3 \cdot 1$, so its Ostrowski representation is 10003.

Ostr.	decimal	Ostr.	decimal	Ostr.	decimal
000000	0	000103	8	001002	16
000001	1	000110	9	001003	17
000002	2	000200	10	001010	18
000003	3	000201	11	010000	19
000010	4	000202	12	010001	20
000100	5	000203	13		
000101	6	001000	14		
000102	7	001001	15		

Ostrowski Numeration Systems

In this case, we can obtain the characteristic Sturmian word (and its standard sequence) of slope α in the following way:

Let π' be the coding that maps $q_1 - 1$ to 0 and all the numbers smaller than $q_1 - 1$ to 1 (we use π' when $\alpha > 1/2$ and π when $\alpha < 1/2$).

Apply the coding π' to the sequence of last digits of the Ostrowski representations of integers.

In our example, we have $q_1 - 1 = 3$ so $\pi' : 0, 1, 2 \mapsto 1, 3 \mapsto 0$, whence

$$s_{\frac{e+1}{e+2}, \frac{e+1}{e+2}} = \pi'(012300123001230123\cdots) = 111011101111101110\cdots$$

Exercise 34

Do the same with $\alpha = 1/\sqrt{2}$. What is the corresponding Sturmian word?

Critical Exponent of Sturmian Words

Recall that the critical exponent of an infinite word x is the supremum of the real number β such that v^β is a factor of x for some nonempty v. The critical exponent of the Fibonacci word is $2 + \varphi$.

In general, the critical exponent of a Sturmian word can be finite or infinite. The following theorem gives a characterization of Sturmian words with finite critical exponent.

Theorem 35

Let s_α be a Sturmian word of slope α. The following are equivalent:

1. *s_α is β-free for some β;*
2. *s_α is linearly recurrent;*
3. *α has bounded partial quotients.*

Critical Exponent of Sturmian Words

Let $\alpha = [0; a_1, a_2, \ldots]$ and suppose that the sequence (a_i) of partial quotients of α is bounded. Let $p_i/q_i = [0; a_1, a_2, \ldots, a_i]$ be the sequence of convergents of α. Then the critical exponent $E(s_\alpha)$ of s_α is given by

$$E(s_\alpha) = \max\left\{a_1, 2 + \sup_{i \geq 2}\{a_i + (q_{i-1} - 2)/q_i\}\right\}$$

Thus, the critical exponent of the Fibonacci word is the least critical exponent a Sturmian word can have.

As an application, let $\alpha = 1 - \sqrt{2}/2 = [0; 3, 2, 2, \ldots]$. The sequence q_i of denominators of the convergents to α is the sequence of positive Pell numbers P_i: $1, 2, 5, 12, 29, 70, 169, 408, 985, 2378$, etc. Since the sequence $(P_{i-1} - 2)/P_i$ converges to $\sqrt{2} - 1$, we have that the critical exponent of the Pell word is $3 + \sqrt{2} \approx 4.4142$.

Recurrence Quotient of Sturmian Words

Recall that the recurrence quotient of x is defined as

$$\rho_x = \limsup_{n \to \infty} \frac{R_x(n)}{n}$$

$R_x(n) = \inf\{m \mid$ every factor of length n occurs in every factor of length $m\}$

is the recurrence function of x.

Theorem 36 (Morse, Hedlund, 1940)

Let s_α be a Sturmian word of slope $\alpha = [0; a_1, a_2, \ldots]$. The recurrence quotient of s_α is

$$\rho_{s_\alpha} = 2 + \limsup_i [a_i; a_{i-1}, \ldots, a_1].$$

For example, if s_α is the Fibonacci word f, then we have $\alpha = [0; 2, 1, 1, 1 \ldots]$ and therefore the recurrence quotient of the Fibonacci word is $\rho_f = 2 + \limsup [1; 1, 1 \ldots, 1] = 2 + \varphi$, which is therefore the smallest possible value for a Sturmian word.

Pure Morphic Sturmian Words

Not all Sturmian words are pure morphic. If a Sturmian word is pure morphic, then it is characteristic. The following theorem characterizes pure morphic Sturmian words in terms of their slope.

Theorem 37

A characteristic Sturmian word $s_{\alpha,\alpha}$ is a fixed point of a morphism (different from the identity) if and only if α is a Sturm number:

1. $\alpha = [0; a_0 + 1, \overline{a_1, \ldots, a_n}]$, *with* $a_n \geq a_0 \geq 1$ *(Sturm number* $< 1/2$*);*
2. $\alpha = [0; 1, a_0, \overline{a_1, \ldots, a_n}]$, *with* $a_n \geq a_0$ *(Sturm number* $> 1/2$*).*

Equivalently, $s_{\alpha,\alpha}$ is a fixed point of a morphism (different from the identity) if and only if α is an algebraic number with minimal polynomial of degree 2 (a quadratic irrational) and the other root α' of the polynomial does not belong to $(0, 1)$ (that is, $1/\alpha < 1$).

Special Factors of Sturmian Words

We saw that Sturmian words have exactly one right special factor for each length. We have the following

Proposition 38

Let $s_{\alpha,\rho}$ be a Sturmian word of slope α. The right special factors of $s_{\alpha,\rho}$ are the reversals of the prefixes of the characteristic Sturmian word $s_{\alpha,\alpha}$.

Since the set of factors of a Sturmian word is closed by reversal, the left special factors of a Sturmian word are precisely the prefixes of the characteristic Sturmian word of the same slope.

Moreover, this also implies that the bispecial factors of a Sturmian word are the right (equivalently, left) special factors that are palindromes.

Special Factors of Sturmian Words

Proposition 39

Let $s_{\alpha,\rho}$ be a Sturmian word of slope α, and v a factor of $s_{\alpha,\rho}$. The following are equivalent:

1. *v is a bispecial factor of $s_{\alpha,\rho}$;*
2. *v is a right special factor of $s_{\alpha,\rho}$ and it is a palindrome;*
3. *v is a left special factor of $s_{\alpha,\rho}$ and it is a palindrome;*
4. *v is a palindromic prefix of the characteristic Sturmian word $s_{\alpha,\alpha}$;*
5. *v is obtained by removing the last two letters from an element of the standard sequence of $s_{\alpha,\alpha}$.*

So, central words are words that are bispecial factors of some Sturmian word.

Iterated Palindromic Closure

Recall that the right palindromic closure of a finite word w is the (unique) shortest palindrome $w^{(+)}$ such that w is a prefix of $w^{(+)}$.

Note that if $w = uv$ and v is the longest palindromic suffix of w, then $w^{(+)} = w\tilde{u}$, where $\tilde{u}$ is the reversal of the word u.

For example, the right palindromic closure of 0 is 0 itself; the right palindromic closure of 01001 is 010010; the right palindromic closure of 010011 is 0100110010.

Iterated Palindromic Closure

It is possible to construct infinite words iterating the right palindromic closure operator. Indeed, given an infinite word $b = b_0 b_1 b_2 \cdots$, called directive word, one constructs the associated infinite word x as the limit of the sequence of words x_n defined by $x_0 = b_0$ and $x_{n+1} = (x_n b_{n+1})^{(+)}$ for every $n > 0$.

For example, take $b = (01)^\omega$ We have $x_1 = (01)^{(+)} = 010$, $x_2 = (0100)^{(+)} = 010010$, $x_3 = (0100101)^{(+)} = 01001010010$, etc. This sequence converges to the Fibonacci word f.

Note that the sequence x_n is the sequence of palindromic prefixes of x.

Iterated Palindromic Closure

We have the following characterization of characteristic Sturmian words:

Theorem 40 (de Luca, 1997)

Every characteristic Sturmian word can be obtained from a right iterated palindromic closure with an infinite directive word containing infinitely many occurrences of 0 *and* 1*, and vice versa.*

Exercise 41

What is the directive sequence of the right iterated palindromic closure that generates the Pell word?

Iterated Palindromic Closure

The construction of words by iterated palindromic closure is not restricted to the binary alphabet.

As another example, the word obtained by iterating the right palindromic closure with directive word $b = (012)^{\omega}$ is the Tribonacci word

$$tr = 0102010010201010201001 02 \cdots$$

and more generally, each m-bonacci word is obtained by iterating the right palindromic closure with directive word $b = (012 \cdots (k-1))^{\omega}$.

Squares in Sturmian Words

Every Sturmian word has infinitely many square prefixes[3].

A square is called minimal if it does not have square prefixes.

For example, 0101 is a minimal square, while 001001 is not.

Theorem 42

Any aperiodic word contains at least 6 minimal squares.

A Sturmian word contains *exactly* 6 minimal squares. However, there are aperiodic words with exactly 6 minimal squares that are not Sturmian, so this is not a characterization of Sturmian words.

[3] The same is true for Tribonacci (and for every standard episturmian word, while it is only conjectured for all suffixes of the Tribonacci word), while surprisingly there is a suffix of the 4-bonacci word that does not have this property.

Squares in Sturmian Words

The minimal squares of the Fibonacci word are: 00, 0101, 1010, 010010, 100100 and 1001010010.

Moreover, in each position of the Fibonacci word one of these squares occurs.

Thus, one can also consider the decomposition of a Sturmian word s in these minimal squares. By deleting half of each square one obtains a new infinite word $\sqrt{s}$, and this word is again a Sturmian word and has the same slope of s (i.e., in the same shift orbit closure).

For example, in the case of the Fibonacci word

$$f = 010010 \cdot 100100 \cdot 1010 \cdot 0101 \cdot 00 \cdot 1001010010 \cdot 0101 \cdot 00 \cdot 1010 \cdots$$

one obtains the Sturmian word

$$\sqrt{f} = 0101001001010010010100101001\cdots$$

Factorizations of the Fibonacci Word

A factorization of an infinite word w is a sequence $(x_n)_{n \geq 1}$ of finite words such that w can be expressed as the concatenation of the elements of the sequence, i.e., $w = \prod_{n \geq 1} x_n$.

We will now show a number of factorizations of the Fibonacci infinite word f that make use of the Fibonacci finite words and other related words.

Recall that the sequence of Fibonacci finite words $(f_n)_{n \geq 1}$ is defined by $f_1 = 1$, $f_2 = 0$ and for every $n > 2$, $f_n = f_{n-1} f_{n-2}$.

Factorizations of the Fibonacci Word

Proposition 43

The Fibonacci infinite word can be obtained by concatenating 0 *and the Fibonacci words:*

$$\begin{aligned} f &= 0\prod_{n\geq 1} f_n \\ &= 0\cdot 1\cdot 0\cdot 01\cdot 010\cdot 01001\cdot 01001010\cdots \end{aligned} \qquad (1)$$

The previous factorization is also the natural factorization of the Fibonacci morphism $\varphi : 0 \mapsto 01,\ 1 \mapsto 0$:

$$\varphi^\omega(0) = 0\cdot 1\cdot \varphi(1)\cdot \varphi^2(1)\cdots$$

Factorizations of the Fibonacci Word

The singular words $\hat{f}_n$ are defined by complementing the first letter in the left rotations of the Fibonacci words f_n. The first few singular words are displayed below. Note that for every $n \geq 1$, one has $\hat{f}_{2n+1} = 0p_{2n+1}0$ and $\hat{f}_{2n+2} = 1p_{2n+2}1$.

$$\hat{f}_1 = 0$$
$$\hat{f}_2 = 1$$
$$\hat{f}_3 = 00$$
$$\hat{f}_4 = 101$$
$$\hat{f}_5 = 00100$$
$$\hat{f}_6 = 10100101$$

Their name comes from the fact that among the $F_n + 1$ factors of f of length F_n, there are F_n of them that are conjugates and one, the singular word, whose left (or equivalently right) rotation is not a factor of f.

Proposition 44

The Fibonacci infinite word is the concatenation of the singular words:

$$\begin{aligned} f &= \prod_{n\geq 1} \hat{f}_n \qquad (2) \\ &= 0 \cdot 1 \cdot 00 \cdot 101 \cdot 00100 \cdot 10100101 \cdots \end{aligned}$$

The factorization (2) is in fact the Lempel–Ziv factorization of f.

The Lempel–Ziv factorization of a word w is $w = w_1 w_2 \cdots$ where w_1 is the first letter of w and for every $i \geq 2$, w_i is the shortest prefix of $w_i w_{i+1} \cdots$ that occurs only once in the word $w_1 w_2 \cdots w_i$.

Roughly speaking, at each step one takes the shortest factor that did not appear before.

Factorizations of the Fibonacci Word

Proposition 45

The Fibonacci word can be obtained also by concatenating the reversals of the Fibonacci words:

$$\begin{aligned} f &= \prod_{n\geq 2} \widetilde{f_n} \\ &= 0 \cdot 10 \cdot 010 \cdot 10010 \cdot 01010010 \cdots \end{aligned} \tag{3}$$

The factorization (3) is basically the Crochemore factorization of f (the only difference is that the Crochemore factorization starts with 0, 1, 0 and then coincides with the one above).

The Crochemore factorization of w is $w = c_1 c_2 \cdots$ where c_1 is the first letter of w and for every $i > 1$, c_i is either a fresh letter or the longest prefix of $c_i c_{i+1} \cdots$ occurring twice in $c_1 c_2 \cdots c_i$.

For example, the Crochemore factorization of the word $w = 0101001$ is $0 \cdot 1 \cdot 010 \cdot 01$, since 010 occurs twice in 01010.

Proposition 46

The Fibonacci word can be obtained by concatenating the reversals of the even Fibonacci words.

$$\begin{aligned} f &= \prod_{n\geq 2} \widetilde{f_{2n}} \\ &= 010 \cdot 01010010 \cdot 0101001010010010100 10 \cdots \end{aligned} \quad (4)$$

Proposition 47

The Fibonacci word can be obtained by concatenating 0 *and the reversals of the odd Fibonacci words:*

$$\begin{aligned} f &= 0 \prod_{n\geq 2} \widetilde{f_{2n+1}} \\ &= 0 \cdot 10010 \cdot 1001001010010 \cdots \end{aligned} \quad (5)$$

Factorizations of the Fibonacci Word

Proposition 48

The Fibonacci word can be obtained by concatenating 01 *and the squares of the reversals of the Fibonacci words:*

$$\begin{aligned} f &= 01 \prod_{n\geq 2} (\widetilde{f_n})^2 \\ &= 01 \cdot (0 \cdot 0)(10 \cdot 10)(010 \cdot 010)(10010 \cdot 10010) \cdots \end{aligned} \qquad (6)$$