Sturmian Words

Keith Li

November 2025

Contents

Abstract

Sturmian words are infinite binary sequences with the smallest possible subword complexity (a measure of the diversity of a sequence in terms of patterns) among all aperiodic words. They can be described by their complexity, by a balance condition, or as mechanical words. In this paper, we introduce the basic definitions for infinite binary sequences, show that mechanical words with irrational slope are aperiodic and balanced, and use the balanced characterization to conclude that they are Sturmian. Examples such as the Fibonacci word illustrate how these sequences come from simple algebraic rules. Our goal is to provide an accessible introduction to Sturmian words and to highlight the connection between their algebraic and combinatorial structure.

1 Introduction

Infinite binary sequences appear in many parts of mathematics, such as combinatorics on words and symbolic dynamics. One of the main questions while studying these sequences is determining whether a sequence eventually becomes periodic or continues to show new behavior forever. To assist in answering this question, mathematicians use tools such as the complexity function, which counts how many different blocks of length n appear. Fundamental results about complexity and periodicity, including the classical Morse–Hedlund theorem, can be found in Lothaire's *Algebraic Combinatorics on Words* [1].

Among aperiodic sequences, Sturmian words play a critical role because they have the smallest possible complexity. They can be described in several ways, by their minimal complexity, by a balance property, and by a construction called a mechanical word. Mechanical words and their role in describing Sturmian sequences are explained in detail by Berstel and Séébold [2].

In this paper, we introduce these ideas and focus on the mechanical viewpoint. We prove that mechanical words with irrational slope are aperiodic and balanced, and by applying the Balanced Characterization of Sturmian Words, we conclude that they are Sturmian. Examples such as the Fibonacci word are included to illustrate how these constructions appear in practice.

2 Preliminaries

In this paper, we will be working over the binary alphabet $\{0, 1\}$. We start by defining a *finite word* to be a sequence made up of 0s and 1s of finite length. Sequences like 101001, 0101, 1, and 000 would be classified as finite words. Let $|w|$ be the length of a finite word. An *infinite word* is a sequence $w = w_0w_1w_2\dots$ indexed by the natural numbers. Throughout this paper, we will focus on infinite words made up of the binary alphabet that are aperiodic, meaning that they do not repeat.

Suppose we had an word $w = w_0w_1w_2\dots$ finite or infinite. A *subword* of length n of this word would be any consecutive block $w_iw_{i+1}\dots w_{i+n-1}$ where i and $i + n - 1$ are within the bounds of this word. For example, in the finite word 0100101, the subwords of length 3 are 010, 100, 001, and 101. Subwords can prove useful when studying patterns in a sequence. The question we are interested in studying is how many distinct subwords of a given length can occur in a word.

Lucky for us, there is already a function designed for this purpose. The complexity function $p(n)$ outputs the number of distinct subwords of length n that occur in a word w. If we go back to the example earlier where we found all the subwords of length 3 of the finite word 0100101, we can see that $p(3) = 4$ in this case since there are 4 subwords of length 3. The complexity function quantifies the richness of a word by counting the distinct subwords it contains. Periodic words generally have a low complexity due to repeated structure, whereas aperiodic words display higher complexity.

A classic result called the Morse-Hedlund theorem[1] states that an infinite word is periodic if and only if there exists some n such that $p(n) \leq n$. This means that every aperiodic word satisfies $p(n) \geq n + 1$ for all n. The proof of this theorem is outside the scope of this paper. An infinite word made up of the binary alphabet is called *Sturmian* if it satisfies

$$p(n) = n + 1 \quad \text{for all } n \geq 1$$

This means that Sturmian words must have the least complexity among all aperiodic infinite words.

3 Results

For any finite word u, define $|u|_1$ to be the number of 1s in the word. An infinite word is *balanced* if for any two subwords u and v where $|u| = |v|$, $||u|_1 - |v|_1| \leq 1$. This means that infinite words that are balanced have their 1s distributed as evenly as possible across all the subwords.

Theorem 3.1 (Balanced Characterization of Sturmian Words[1])**.** *A binary infinite word is Sturmian if and only if it is both aperiodic and balanced. The proof of this characterization is technical and is not included here..*

Mechanical words provide an algebraic method to describe Sturmian words. Suppose we had a slope α where $0 < \alpha < 1$ and an intercept ρ. The mechanical word with these two parameters is the infinite sequence $s = s_0 s_1 s_2 \dots$ defined by

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

Each term in this sequence is either 0 or 1, because the difference between the two terms without the floor is $((n+1)\alpha + \rho) - (n\alpha + \rho) = \alpha$, and since $0 < \alpha < 1$, increasing a real number by α changes the floor by at most 1. Thus, $s_n \in \{0, 1\}$.

Mechanical words[2] are fundamental because every Sturmian word arises from a mechanical word with an irrational slope, and conversely any mechanical word with an irrational slope is Sturmian. This perspective makes key properties such as aperiodicity and balance much easier to establish.

Theorem 3.2. *If α is irrational, the mechanical word of slope α is aperiodic.*

Proof. Assume for the sake of contradiction that s has a period of $T > 0$. This means that $s_{n+T} = s_n$. Let $a_n = \lfloor n\alpha + \rho \rfloor$, so $s_n = a_{n+1} - a_n$. The condition for periodicity $s_{n+T} = s_n$ can now be written as

$$a_{n+T+1} - a_{n+T} = a_{n+1} - a_n$$

Rearranging, we get

$$a_{n+T+1} - a_{n+1} = a_{n+T} - a_n$$

Now, let $b_n = a_{n+T} - a_n$. We can write the equation above as

$$b_{n+1} = b_n$$

This means that b is constant throughout, which also means that $a_{n+T} - a_n$ is constant for all n. Substituting the original expression back in, we get that

$$\lfloor (n+T)\alpha + \rho \rfloor - \lfloor n\alpha + \rho \rfloor$$

is also constant. However, since α is irrational, the fractional parts of $n\alpha + \rho$ are dense in $[0, 1)$, which means that the difference of the floored terms cannot remain constant unless $T\alpha$ is an integer. This forces α to be rational, which is a contradiction. Therefore, the mechanical word with slope α where α is irrational is aperiodic. □

Theorem 3.3. *If α is irrational, the mechanical word of slope α is balanced.*

Proof. Let u and v be the subwords of length k starting at positions i and j. To calculate the number of 1s in the subwords, we can simply sum each element of the word since it is either 0 or 1.

$$|u|_1 = \sum_{n=i}^{i+k-1} s_n = \sum_{n=i}^{i+k-1} \lfloor (n+1)\alpha + \rho \rfloor - \lfloor n\alpha + \rho \rfloor$$

$$|v|_1 = \sum_{n=j}^{j+k-1} s_n = \sum_{n=j}^{j+k-1} \lfloor (n+1)\alpha + \rho \rfloor - \lfloor n\alpha + \rho \rfloor$$

Notice that these sums telescope into a difference of two terms.

$$|u|_1 = \lfloor (i+k)\alpha + \rho \rfloor - \lfloor i\alpha + \rho \rfloor$$

$$|v|_1 = \lfloor (j+k)\alpha + \rho \rfloor - \lfloor j\alpha + \rho \rfloor$$

Notice that without the floors on the terms, the differences for both would be $k\alpha$. And since putting the floor can only change the value by at most 1, we get that

$$||u|_1 - |v|_1| \leq 1$$

which means that it is balanced. □

Corollary 3.3.1. *If α is irrational, then the mechanical word of slope α is a Sturmian.*

Proof. By Theorem 3.2 and Theorem 3.3, we get that the word is aperiodic and balanced. Therefore, by Theorem 3.1, we get that any aperiodic and balanced word must be Sturmian. □

4 Examples

One of the most well known Sturmian words is the Fibonacci word. We can generate the word by using the morphism $0 \mapsto 01$ and $1 \mapsto 0$. By starting at 0 and applying the morphism infinitely, we get the Fibonacci word.

$$0, 01, 010, 01001, 01001010, 0100101001001, \ldots$$

The Fibonacci word corresponds to a mechanical word with slope $\frac{1}{\phi^2}$ where $\phi = \frac{1+\sqrt{5}}{2}$, the golden ratio. Since the slope is irrational, by Corollary 3.3.1, the word must be Sturmian.

Another example is when the slope $\alpha = \sqrt{2} - 1 = \frac{1}{\delta}$ where $\delta = 1 + \sqrt{2}$, the silver ratio. The word constructed from this begins with

$$0, 1, 0, 1, 0, 0, 1, 0, 1$$

Since the slope is irrational, the word is Sturmian. This example is useful because it demonstrates how changing the irrational slope produces different Sturmian words with slightly different distributions of 0s and 1s. Unlike the Fibonacci case, this slope does not come from a simple substitution rule, so it highlights how mechanical words are able to generate Sturmian sequences even when no morphism is available. In general, any mechanical word with irrational slope is Sturmian.

5 Discussion

The mechanical description of Sturmian words offers a useful alternative to understanding them through subword complexity. Instead of counting the number of different blocks of length n in a word, the mechanical definition allows us to describe Sturmian sequences using a simple algebraic formula. This makes several important properties, such as aperiodicity and balance, much easier to verify directly.

The examples considered, including the Fibonacci word and the word generated from the slope $\sqrt{2}-1$, show that different irrational slopes produce distinct Sturmian sequences while always preserving these same crucial features. In this way, mechanical words provide a accessible framework for studying Sturmian words that complements the more abstract complexity definition.

6 Conclusion

In this paper, we presented the main definitions and ideas needed to study Sturmian words, beginning with subwords and complexity and then introducing mechanical words. We showed that mechanical words with irrational slope are necessarily aperiodic and balanced, and by applying the Balanced Characterization of Sturmian Words, we obtained a direct criterion for recognizing Sturmian sequences. Our examples, including the Fibonacci word and the sequence generated using the slope $\sqrt{2} - 1$, demonstrated how these constructions work in specific cases. Taken together, the results give a clear path from the algebraic definition of a mechanical word to the identification of Sturmian structure, providing a straightforward overall framework for working with these sequences.

References

[1] Lothaire, M. (2002). *Algebraic Combinatorics on Words.* Cambridge University Press.

[2] Berstel, J., and Séébold, P. (2005). Sturmian words. In *Algebraic Combinatorics on Words* (pp. 45–110). Springer.