Formal Intercepts of Sturmian words

Caïus Wojcik

Abstract :

We study Sturmian words, and particularly the second parameter describing this class of words.

Sturmian words are infinite words over a 2-letter alphabet. They are defined as the infinite words having lowest unbounded complexity. Namely, a theorem of Morse and Hedlund states that an infinite word is ultimately periodic if and only if it has bounded complexity. Sturmian words are characterised by the properties of not being ultimately periodic, and being balanced, that is, the number of 1 appearing in factors of a given length only takes two values.

The first parameter decribing a Sturmian word is its slope, defined as the asymptotic proportion of 1, and characterises the set of factors of the sturmian word. Through the continued fraction expansion of this irrational number one can construct a distinguished Sturmian word of the corresponding slope, called the characteristic word, which is the only sturmian word of this slope admitting two sturmian extensions on the left.

Sturmian words are obtained geometrically by drawing a line on the plane and coding the crossing through a vertical line by a 1 and the horizontal line by a 0. The slope of the sturmian word being obtained through the slope of the considered line, and the second parameter, the intercept, is usually presented as the real number on the y-axis intersecting the line.

The characteristic word may be described using so-called standard and central words. The first ones are obtained by a concatenation process similar to Euclide's algorithm on integers. The second ones are obtained by removing the last two letters of a standard word, and are palindromic words with

two relatively prime periods, realizing the sharp bound of Fine and Wilf's theorem.

On the other hand, we study the repetition function of sturmian words, defined as the number of distinct factors of a given length appearing at the beginning of an infinite word. This function is related to the diphantine exponent of infinite words, and in the case of sturmian words, it has been shown that the diophantine exponent is bounded if and only if the partial quotients of the continued fraction expansion of the slope are bounded.

We study the repetition function using so-called Rauzy graphs of infinite words, also called factor graphs. Rauzy graphs are defined as a sequence of directed graphs, with vertexes the factors of a fixed infinite word and arrows linking two factors if one is obtained by the other by a one-letter shift. The Rauzy graphs of arbitrary words are in general difficult to compute, but in the case of sturmian words they are particularly simple.

Indeed, the rauzy graphs of sturmian words are constituted of two cycles, patched together by a common part. The length of these two cycles are linked to the denominator of the sequence of convergents in the continued fraction expansion of the slope. Also, the evolution of the Rauzy graphs in the case of sturmian word may be fully described.

Any infinite word defines an infinite path on its Rauzy graph. Since these paths must be coherent with each others, there are a lot of restrictions on the possible paths taken by a given word. The consideration of these paths for different element of the shift orbit of the base word is a natural point of view of combinatorics on words. The repetition function can be read on those path, since it is the length of the longest hamiltonian path at the beginning of the base word's path.

Among the two cycles of the Rauzy graph of a sturmian word there a distinguished one, that we call the referent cycle, defined as the one through whom the characteristic word first passes. Indeed, we can compute the number of times the characteristic word turns around a cycle, and take it as a reference for other sturmian words. We use this idea to compute the length of cycles in Rauzy graphs, using the fact that the first repeated factor at the beginning of the characteristic word is its prefix.

These description of Rauzy graphs give another understanding of the so-called three-gap theorem, stating that an arithmetic sequence on the circle divides it into arcs with at most three possible lengths, corresponding to the frequencies of factors in the sturmian word. Indeed, two factors belonging to the same branch of the Rauzy graph will have same frequencies, and the frequencies of the factors of the common part will be the sum of the two other frequencies.

The dynamical point of view of sturmian words as coding of rotations on the circle gives an analytic description of the intercept. Namely, an intercept can be expressed as an infinite alternating sum of irrational number obtained by shifting the original sequence of partial quotients, pounded with integer coefficients satisfying Ostrowski conditions. However, computations difficulties aside, there remains the problem of combinatorialy define the intercept, in addition to the non-injectivity of this representation.

Ostrowski conditions may be view as a condition of injectivity of the number system associated to the continuants of the continued fraction expansion. They describe the rule that are being applied when one wants to sum two integers written in Ostrowski expansion, an operation that is very far from being understood.

In the dynamical setting given by an infinite word, and given an element in its dynamical subshift, realized as an approximation by suffixes of the base word, one can naturally consider the sequence of natural integers encoding each of the corresponding suffixes. In the context of Sturmian words, this sequence consists of partial expansions of an infinite Ostrowski expansion, linking the chaos of dynamics to the rigidity of combinatorics.

This infinite Ostrowski expansion is obtained by reading the operation of shifting on the Rauzy graphs of Sturmian words. Namely, the partial sums are obtained by considering the case of a length that is the increment of the length of bispecials that are purely central (that is, coming from standard words that are not semi-standard).

This observation allows us to define what we call the *formal intercept* of a sturmian word. They are formally defined as the projective limit of the system of integers with a given maximal number of terms in their Ostrowski expansion. Our main result is that this definition fully describes in a combinatorial and dynamical point of view the set of Sturmian words, in a bijective way.

To prove this result, we use the particularly simple shape of Rauzy graphs of Sturmian words. We use the computations of lengths of cycles previously obtained to check the congruences relations to build a Sturmian words associated to a given formal intercept. For the converse, we characterize the formal intercept as an exact approximation of a given Sturmian word.

We will compute the formal intercept of the two sturmian extensions of the caracteristic word. They have the particularity of having their support in even places and odd places respectively, the support being defined as the set of indices with non-zero Ostrowski coefficients.

We will present the content of a work in progress, that goes as follows.

The action of the shift on Sturmian words allows us to add 1 to a formal intercept, therefore defining the addition of a formal intercept with a natural number. We will say that two formal intercept are equivalent whenever they coincide up to addition by some integers. For Sturmian words that are not equivalent to the characteristic word, this reduces to equality of almost all Ostrowski coefficients.

On the other hand, given a formal intercept one can use product formulas on reversal of standard words to define a Sturmian word. However this process will not reach the sturmian words that are equivalent to the characteristic word without being one of its suffixes. This operation, presented as a product formula, is convenient when considering questions of computations.

The sturmian words obtained by this process have a formal intercept that is complementary of the base intercept. That is, for a sum law defined on the set of classes of equivalence of Sturmian words, this amounts to the consideration of the opposite of a given Sturmian word.

The sum law defined this way would give us an isomorphism between the set of equivalence classes of Sturmian words and the quotient of the real numbers by the subgroup generated by 1 and the slope of the sturmian word. We will present how certain formulas on continuants of continued fraction expansions of quadratic numbers can be interpreted as elements of torsion in the group of equivalence class of Sturmian words, giving an asymptotic realisation of the operation of division by a natural integer in Ostrowski expansion.