On the conjectures of Rauzy and Shallit for infinite words

Jean-Paul Allouche, Mireille Bousquet-Mélou

Abstract. We show a connection between a recent conjecture of Shallit and an older conjecture of Rauzy for infinite words on a finite alphabet. More precisely we show that a Rauzy-like conjecture is equivalent to Shallit's. In passing we correct a misprint in Rauzy's conjecture.

Keywords: combinatorics on words, recurrence function, Sturmian sequences

Classification: 11B05, 68R15

In a paper of Shallit and Breitbart (see [9], see also [7]) Shallit proposed a conjecture on the finite factors (subwords) of the non-ultimately periodic infinite words on a finite alphabet: *let $u = u_0u_1u_2\cdots$ be an infinite word over a finite alphabet that is not ultimately periodic. Define $S(n)$ to be the length of the longest suffix of $u_0u_1\cdots u_n$ that is also a factor of $u_0u_1\cdots u_{n-1}$. Then*

$$\liminf_{n\to\infty} \frac{S(n)}{n} \leq C,$$

where $C = \frac{3-\sqrt{5}}{2}$.

Furthermore the conjecture is proved optimal (if it is true) by considering the Fibonacci word

$$0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ \cdots$$

which is the fixed point of the morphism $0 \to 01$, $1 \to 0$.

This conjecture resembles a conjecture by Rauzy [8] on the recurrence function R of a non-ultimately periodic primitive infinite word on a finite alphabet: *let $u = u_0u_1u_2\cdots$ be an infinite word over a finite alphabet that is primitive and not ultimately periodic. Define $R(n)$ to be the smallest integer such that every factor of u of length $R(n)$ contains at least one copy of every factor of u of length n, (the hypothesis that u is primitive means that $R(n)$ exists for each n). Then*

$$\limsup_{n\to+\infty} \frac{R(n)}{n} \geq D,$$

where $D = \frac{3+\sqrt{5}}{2}$.

The way the conjecture is given in [8] (and this was confirmed recently to us by G. Rauzy) seems to consider that the constant D is the same as the (best possible) constant when restricting to Sturmian sequences, (but there is a misprint, the best constant for Sturmian sequences is $1 + D$, as stated in [5]).

Of course the number $\sqrt{5}$ often enters the picture when studying infinite words on a finite alphabet, because the Fibonacci word which has many "optimal" properties is related to the golden ratio. But we were intrigued by these two conjectures (note in particular that $CD = 1$) and we tried to find a link between them.

We first defined a function R' resembling the recurrence function R of an infinite word. In particular we noticed that Rauzy's conjecture (as stated with the constant D) is implied by the following Rauzy-like conjecture:

$$\limsup_{n\to+\infty} \frac{R'(n)}{n} \geq D.$$

We then proved that this Rauzy-like conjecture is equivalent to Shallit's. Unfortunately Shallit's conjecture does not imply Rauzy's because of the misprint in the latter: the constant D in Rauzy's conjecture has to be replaced by $1 + D$ to keep the Sturmian sequences optimal.

In this paper we state our Rauzy-like conjecture and we prove that it is equivalent to Shallit's.

1. The recurrence function of an infinite word and the Rauzy conjecture

Definition. *Let u be an infinite word on a finite alphabet. The recurrence function of the word u is the function $n \to R(n)$ (which might be infinite), where $R(n)$ is the smallest integer such that every factor of u of length $R(n)$ contains at least one copy of every factor of u of length n.*

Remark. An infinite word for which $\forall n \geq 1,\ R(n) < +\infty$ is called primitive. It is not hard to see that an infinite word is primitive if and only if each factor of it occurs infinitely often and "with bounded gaps" (i.e. the distance between two consecutive occurrences is bounded). A primitive word is sometimes called minimal, as the dynamical system it generates is minimal: the dynamical system generated by an infinite word u is defined as (Ω, T) where T is the shift and Ω the closure of the orbit of u under the shift; the system is said minimal if there is no closed subset of Ω stable under the shift except $\emptyset$ and Ω. It is easy to see that this means exactly that the word u is primitive.

The reader wanting to know more about the recurrence function of infinite words can read for instance [4], [5] and also [6].

We can now state the Rauzy conjecture.

Rauzy's conjecture ([8]). *Let u be an infinite word on a finite alphabet that is primitive and non-ultimately periodic. Then:*

$$\limsup_{n\to+\infty} \frac{R(n)}{n} \geq \frac{5+\sqrt{5}}{2} = 3.61803\cdots$$

and the constant is optimal.

Remark. As we said in the introduction, the original text has the constant $\frac{3+\sqrt{5}}{2}$, but there is a misprint and this constant should be replaced by $\frac{5+\sqrt{5}}{2}$.

2. The Shallit conjecture

Conjecture ([9] and [7]). *Let $u = u_0u_1u_2\cdots$ be an infinite word over a finite alphabet that is not ultimately periodic. Define $S(n)$ to be the length of the longest suffix of $u_0u_1\cdots u_n$ that is also a factor of $u_0u_1\cdots u_{n-1}$. Then*

$$\liminf_{n\to\infty} \frac{S(n)}{n} \leq \frac{3-\sqrt{5}}{2} = .381966\cdots$$

and the constant is optimal.

To know more about this conjecture and related topics the reader is referred to [9], [10] and [7].

3. A function resembling the recurrence function and a Rauzy-like conjecture

We introduce here a function resembling the recurrence function. Instead of looking for a length $R(n)$ such that every window having this length contains all factors of length n of the infinite word, we just look for a length $R'(n)$ such that the window of length $R'(n)$ at the beginning of the infinite word contains all factors of length n.

Definition. *Let u be an infinite word on a finite alphabet. We define the function R' by: $R'(n)$ is the length of the shortest prefix of the infinite word u that contains at least one occurrence of each factor of u of length n.*

Remark. For any $n \geq 1$, one has $R'(n) < +\infty$ as the alphabet is finite. Furthermore the following inequality is clear:

$$\forall n \geq 1, \;\; R'(n) \leq R(n).$$

Indeed as any window of length $R(n)$ contains all the factors of length n, this is the case in particular for the prefix of u of length $R(n)$.

A Rauzy-like conjecture. *Let u be an infinite word on a finite alphabet that is non-ultimately periodic. Then:*

$$\limsup_{n\to+\infty} \frac{R'(n)}{n} \geq \frac{3+\sqrt{5}}{2} = 2.61803\cdots$$

and the constant is optimal.

4. The Rauzy-like conjecture and the Shallit conjecture are equivalent

The aim of this paragraph is to prove the following theorem.

Theorem. *For any infinite word u with values in a finite alphabet, one has:*

$$\frac{1}{\liminf\limits_{n\to+\infty} \frac{S(n)}{n}} = \limsup_{n\to+\infty} \frac{R'(n)}{n}.$$

Hence the Rauzy-like conjecture is equivalent to Shallit's.

We will prove this theorem in six steps:

(i) $R'(S(n)+1) \geq n+1$,

(ii) $S(R'(n)) < n+1$,

(iii) $S(R'(n)) \geq n$,

(iv) $S(R'(n)) = n$,

(v) $\liminf\limits_{n\to+\infty} \frac{S(n)}{n} \leq \frac{1}{\limsup\limits_{n\to+\infty} \frac{R'(n)}{n}}$,

(vi) $\frac{1}{\liminf\limits_{n\to+\infty} \frac{S(n)}{n}} \leq \limsup\limits_{n\to+\infty} \frac{R'(n)}{n}$.

$*$ Step (i)

Consider the word $u_0 \cdots u_n$ and its suffix $u_{n-S(n)} u_{n-S(n)+1} \cdots u_n$ of length $S(n)+1$. This suffix is not a factor of $u_0 \cdots u_{n-1}$ by definition of $S(n)$. Hence the length of the prefix of the infinite word u with minimal length that contains all the factors of u of length $S(n)+1$ is certainly larger than or equal to the length of $u_0 \cdots u_n$, i.e. $n+1$.

$*$ Step (ii)

Consider the word $u_0 \cdots u_{R'(n)}$ and its suffix $u_{R'(n)-n} u_{R'(n)-n+1} \cdots u_{R'(n)}$ of length $n+1$. The factor $u_{R'(n)-n} u_{R'(n)-n+1} \cdots u_{R'(n)-1}$, of length n, is not a factor of $u_0 \cdots u_{R'(n)-2}$ from the definition of R', (if it were, the word $u_0 \cdots u_{R'(n)-2}$ would then contain all factors of length n of the infinite word u, although it has length $R'(n)-1$). This implies in turn that the word $u_{R'(n)-n} u_{R'(n)-n+1} \cdots u_{R'(n)}$ is not a factor of $u_0 \cdots u_{R'(n)-1}$, hence $S(R'(n)) < n+1$.

$*$ Step (iii)

Consider the word $u_0 \cdots u_{R'(n)}$ and its suffix $u_{R'(n)-n+1} u_{R'(n)-n+2} \cdots u_{R'(n)}$ of length n. By definition of $R'(n)$ this last word is a factor of $u_0 \cdots u_{R'(n)-1}$, hence $S(R'(n)) \geq n$.

$*$ Step (iv)

This step results from Steps (ii) and (iii).

$*$ Step (v)

Define $a = \liminf_{n\to+\infty} \frac{S(n)}{n}$ and $b = \limsup_{n\to+\infty} \frac{R'(n)}{n}$.

We first notice that the sequence R' is strictly increasing: indeed it results from the proof of Step (ii) that the word $u_{R'(n)-n}u_{R'(n)-n+1}\cdots u_{R'(n)}$ which has length $n+1$ is not a factor of $u_0\cdots u_{R'(n)-1}$. This implies $R'(n+1) > R'(n)$.

Now, replacing the sequence of integers n by the sequence $R'(n)$, we can write:

$$a = \liminf_{n\to+\infty} \frac{S(n)}{n} \leq \liminf_{n\to+\infty} \frac{S(R'(n))}{R'(n)}.$$

But using Step (iv) we have:

$$\liminf_{n\to+\infty} \frac{S(R'(n))}{R'(n)} = \liminf_{n\to+\infty} \frac{n}{R'(n)} = \frac{1}{\limsup\limits_{n\to+\infty} \frac{R'(n)}{n}}.$$

Hence

$$a \leq \frac{1}{b}.$$

Let us take now a strictly increasing sequence of integers n_k such that

$$\lim_{k\to+\infty} \frac{S(n_k)}{n_k} = a.$$

We claim that the sequence $S(n_k)$ is unbounded: indeed, if $S(n_k)$ were bounded, then $R'(S(n_k)+1)$ would be bounded as R' is strictly increasing, but one has $R'(S(n_k)+1) \geq n_k+1$ from Step (i). Then, after replacing if necessary the sequence n_k by a subsequence, one can assume that the sequence $S(n_k)$ is strictly increasing. Of course the sequence $S(n_k)+1$ will be also strictly increasing. Hence

$$b = \limsup_{n\to+\infty} \frac{R'(n)}{n} \geq \limsup_{k\to+\infty} \frac{R'(S(n_k)+1)}{S(n_k)+1} \geq \limsup_{k\to+\infty} \frac{n_k+1}{S(n_k)+1},$$

the last inequality coming from Step (i). By the definition of the sequence n_k one has

$$\limsup_{k\to+\infty} \frac{n_k+1}{S(n_k)+1} = \lim_{k\to+\infty} \frac{n_k+1}{S(n_k)+1} = \frac{1}{a},$$

hence

$$\frac{1}{a} \leq b,$$

and finally

$$\frac{1}{a} = b.$$

5. Miscellanea

We make here some remarks on the questions discussed above.

- For any infinite word one has the inequality $R'(n) \geq p(n)+n-1$, where $p(n)$ is the number of blocks of length n that occur in this word, (the function $n \to p(n)$ is called the complexity of the infinite word, see [1] for a survey). Indeed, in the "worst" case, the $p(n)$ blocks of length n that occur in the infinite word are respectively the blocks $u_0 \cdots u_{n-1}$, $u_1 \cdots u_n$, $\cdots$. Note that the lower bound $R(n) \geq p(n) + n - 1$ is well known. It is implied by this remark. For more precise results see [4] and [5].

- From the above remark we can deduce a simpler way than in [9] and [7] of obtaining the bound $\frac{1}{2} = .5$ instead of $.381966\cdots$ in Shallit's conjecture, or the bound 2 instead of $2.61803\cdots$ in our Rauzy-like conjecture. Indeed for a non-ultimately periodic word, one has (see [4] for instance, see also [2]): $p(n) \geq n+1$, hence $R'(n) \geq 2n$.

- Almost all infinite words on the finite alphabet $\mathcal{A}$ are "normal", i.e. every possible block of length k occurs with frequency $(\text{Card } \mathcal{A})^{-k}$. In particular the complexity of almost all infinite words is given by: $p(n) = (\text{Card } \mathcal{A})^n$, hence
$$R'(n) \geq (\text{Card } \mathcal{A})^n + n - 1$$
and $\limsup_{n \to +\infty} \frac{R'(n)}{n} = +\infty$. Of course the normal words are not primitive, hence this says nothing for Rauzy's conjecture.

- Using again our first remark, one sees that
$$\limsup_{n \to +\infty} \frac{R'(n)}{n} \geq 1 + \limsup_{n \to +\infty} \frac{p(n)}{n},$$
hence it "suffices" to show the Shallit conjecture or the Rauzy-like conjecture for infinite words such that $\limsup_{n \to +\infty} \frac{p(n)}{n} \leq \frac{1+\sqrt{5}}{2}$. Among them the case of the Sturmian words (for which $p(n) = n+1$) can certainly be addressed by adapting the arguments of [5] for the computation of $\limsup_{n \to +\infty} \frac{R(n)}{n}$, but we have not written the details. Furthermore the fact that the conjecture is true both for the "high" complexities and for the "minimal" complexity (i.e. the Sturmian case) is a strong indication it may be true for all infinite words.

- We would like to finish with another conjecture which looks like the previous ones. J. Currie, trying to find an easier version of Shallit's conjecture proposed the following one: *if all sufficiently long prefixes of an infinite word over* *a finite alphabet have a non-trivial square suffix, then this word is ultimately periodic.* Shallit showed the conjecture is false (for our friend the Fibonacci word), but he also modified the conjecture as follows: *if all sufficiently long prefixes of an infinite word over a finite alphabet have a non-trivial suffix which is a* $\frac{3+\sqrt{5}}{2}$*-power, then this infinite word is ultimately periodic, (with an optimal exponent* $\frac{3+\sqrt{5}}{2}$ *obtained for the Fibonacci word).* Meanwhile Currie and Vandeth proved a weak version of the conjecture *with the exponent* 3. Recently, Mignosi, Restivo and Salemi proved the modified Currie conjecture *with the correct exponent* $\frac{3+\sqrt{5}}{2}$, see [3].

Acknowledgments. The authors want to thank very warmly J. Shallit, F. Mignosi and A. Restivo for interesting discussions and references.

References

[1] Allouche J.-P., *Sur la complexité des suites infinies*, Bull. Belg. Math. Soc. **1** (1994), 133–143.

[2] Coven E.M., Hedlund G.A., *Sequences with minimal block growth*, Math. Systems Theory **7** (1973), 138–153.

[3] Mignosi F., Restivo A., Salemi S., *A periodicity theorem on words and applications*, Math. Foundations of Computer Science (1995).

[4] Morse M., Hedlund G.A., *Symbolic dynamics*, Amer. J. Math. **60** (1938), 815–866.

[5] Morse M., Hedlund G.A., *Symbolic dynamics II, Sturmian trajectories*, Amer. J. Math. **62** (1940), 1–42.

[6] Mouline J., *Contribution à l'étude de la complexité des suites substitutives*, Thèse, Université de Provence (1989).

[7] Pomerance C., Robson J.M., Shallit J., *Automaticity II: descriptional complexity in the unary case*, preprint, 1994.

[8] Rauzy G., *Suites à termes dans un alphabet fini*, Sém. de Théorie des Nombres de Bordeaux (1982-1983), 25-01–25-16.

[9] Shallit J., Breitbart Y., *Automaticity I: properties of a measure of decriptional complexity*, in: P. Enjalbert, E.W. Mayr, K.W. Wagner editors, STACS 94: 11th Annual Symposium on Theoretical Aspects of Computer Science, Lecture Notes in Computer Science (Springer-Verlag) **775** (1994), 619–630.

[10] Shallit J., Breitbart Y., *Automaticity I: properties of a measure of decriptional complexity*, preprint, 1994.

CNRS, LMD, Luminy, Case 930, F-13288 Marseille Cedex 9, France

CNRS, LaBRI, 351 cours de la Libération, F-33405 Talence Cedex, France

E-mail: allouche@lmd.univ-mrs.fr; bousquet@labri.u-bordeaux.fr