Report on the PhD thesis of Elzbieta Krawczyk
by
Fabien Durand,
Professeur des Universités at Université de Picardie Jules Verne

Elzbieta Krawczyk's thesis concerns the symbolic dynamics and combinatorics on words. It focuses on the arithmetic, combinatorial and dynamic properties of substitutive sequences, with a majority of results concerning the family of automatic sequences.

Automatic sequences have been the subject of a great deal of work since the 1970s, from the point of view of dynamics, automata theory, logic and even number theory. In 1972, A. Cobham characterized p-automatic sequences $y \in B^{\mathbb{N}}$, defined by automata, using fixed points $x \in A^{\mathbb{N}}$ of free monoid endomorphisms $\sigma : A^* \to A^*$ of constant length p and morphisms $\phi : A^* \to B^*$ such that $y = \phi(x)$ and $\sigma(x) = x$; the sets A and B being finite and called alphabets. The constant length of σ means that the number of letters in $\sigma(a)$ is independent of a and equal to p. When the endomorphism (substitution) is not of constant length, y is said to be a substitutive sequence.

Other equivalent definitions have subsequently been given. For example, p-automatic sequences "correspond" to the characteristic sequences of sets of integers that Presburger arithmetic can define. Or the algebraicity of power series on $\mathbb{F}_p$ (Christol, 1979).

Earlier, in 1969, when the sequence x is not ultimately periodic, A. Cobham showed that x is p and q-automatic if, and only if, there exist integers $k, l \geq 2$ such that $p^k = q^l$.

This last result is now known as Cobham's Theorem and was considered by Samuel Eilenberg, in his 1974 book Automata Theory, to be a cornerstone of automata theory. Eilenberg commented that, since the proof was a long and complex succession of many technical and combinatorial lemmas, it was a challenge to find more elegant proofs.

Until very recently, new proofs of this theorem were published using methods from logic, dynamical systems, combinatorics or automatic theory.

It's a very rich subject, both in terms of the methods used and the fields of investigation to be developed.

All this is well documented in Elzbieta Krawczyk's thesis, right up to very recent algebraic developments.

This thesis has its origins in Cobham's 1969 theorem and its developments.

1. Introduction

The introduction of this document is particularly well written. It will give the reader unfamiliar with the field an idea of the issues at stake and the variety of problems and openings offered by the objects manipulated.

She presents her five main results. Two are extensions of Cobham's theorem, Theorem A and Theorem B. Theorem C is a characterization of automatic substitution sequences. This answers a question posed by J.-P. Allouche and J. Shallit in 2003 in their book "Automatic sequences". Theorem D states that transitive dynamical subsystems of substitutive dynamical systems are substitutive.

Theorem E tends to specify the structure of return times in cylinders within p-automatic dynamical systems.

2. Preliminaries

In this section, all the classical definitions of the field and the context of this work are given. E. Krawczyk takes the approach of considering only increasing endomorphisms, i.e. such that the length of the iterates tends to infinity. This is a very standard assumption, which avoids considerations that are often too technical and that cloud the understanding of the problems considered. In the case of automatic sequences, the endomorphisms are necessarily increasing. In the substitutive case, it is possible to consider erasing endomorphisms, for example. This manuscript therefore does not consider these cases.

In my opinion, it's much better that way.

Several classical results that are useful for the rest of the paper are recalled and proved with statements adapted to the context of this thesis. I'm thinking of Prop. 1.2.16, Corollary 1.2.17, Lemma 1.2.21, Corollary 1.2.27, Theorem 1.3.19.

3. Chapter 2

Theorem D is proven in this chapter.

E. Krawczyk introduced the useful terminology of quasi-fixed point of substitutions. It has been considered by Holton and Zamboni in 2001. They showed that in the primitive framework (the composition matrix of the substitution is primitive) quasi-fixed points are substitutive. Proposition 2.1.8 extends this result avoiding the primitivity assumption.

Then, E. Krawczyk considers quasi-fixed points of automatic sequences and show that this property is preserved by factor maps.

It is also proved in this chapter that transitive substitutive subshifts are substitutive : this is Theorem D. The proof of this result also provides that subsystems of p-automatic systems are also p-automatic. This is an important observation when we wish to move from the primitive to the non-primitive case. It has already been used in the literature, but here it has the merit of being written down and proved explicitly. The argument is a classic one, involving the use of the primitive normal form of the incidence matrix and the dominant eigenvalues of the diagonal blocks.

Then, E. Krawczyk considers quasi-fixed points of automatic sequences and show that this property is preserved by factor maps. It is a consequence of the two previous results of this chapter.

4. Chapter 3

Theorem E is proven in this chapter.

Return times on words w are considered for k-automatic sequences x. Let $N(x, w)$ be the set of occurrences of w in x :

$$x_i x_{i+1} \cdots x_{i+|w|-1} = w.$$

E. Krawczyk shows that there exists an integer M such that all elements of this set have the same residue modulo $k^{|w|-M}$. It is explained when x is uniformly recurrent this result is implicit in two papers about recognizability but never stated nor proven. E. Krawczyk proposes its own proof that follows the ideas of these two papers. The arguments are classical, they uses the recognizability of primitive substitution.

But the global constant length case (or automatic case) is also treated and much more tricky. In such a framework minimal subsystems always exists. They can be either periodic or not. In both case they are automatic. Given a word w in x, E. Krawczyk uses these two types of subsystems as "markers" in w in order to determine the hitting times to w along x. In this way, and using results from the previous chapter, she obtained there exists a finite set F and an integer M, not depending on w such that $N(x, w) - N(x, w)$ is included in $F \mod k^{|w|-M}$.

5. Chapter 4

Theorem A and Theorem B are proven in this chapter.

This chapter is a refinement of Cobham's theorem and is, with Chapter 5, the heart of this thesis. It is proven that for $k, l \in \mathbb{N}$ multiplicatively independent, x a k-automatic sequence and y a l-automatic sequence, if a sufficiently long word w (whose size is effectively bounded) appears in x and y then the sequence w^∞ appears in both subshifts generated by x and y. This is Theorem B.

This is an interesting extension of Cobham's theorem. It is used to obtain what E. Krawczyk call, with reason, a finitary version of Cobham's theorem. It characterizes the intersection of languages generated by multiplicatively independent automatic sequences. Moreover the proof as the advantage to provide an algorithm that computes the common words of any two given multiplicatively independent automatic sequences. This is a very nice result and I think it is the best of this thesis. The proof is tricky, technical but well written.

6. Chapter 5

It is easy to construct automatic sequences that are given by non constant length substitutive sequences. It is less obvious to determine if a non constant length substitutive sequence is automatic.

Such situation is called in the litterature, hidden automatic sequences. M. Dekking gave in 1978 a necessary criterion for a substitutive sequence to be automatic.

In this work, E. Krawczyk provides a necessary and sufficient condition for purely substitutive sequences and a simpler one for the primitive case. Both are algorithmic and thus provide a decidability result for automaticity.

E. Krawczyk precises that a general necessary and sufficient condition is missing. She obtains a statement for substitutive sequences (instead of fixed point of substitutions) whose subshifts does not contain periodic point. Hence it remains to show the result for those having periodic points in their subshifts in order to completely solve this problem. This is known to be the most difficult situation when dealing with substitutions. Nevertheless it is a very important result she obtained.

Many interesting examples, or more specific situations, are given and treated.

The main part of the proof consists in the control of the length of the iterates of the underlying substitutions applied to some very well chosen words. The proof is long and well seen.

7. Conclusion

This thesis is very well written despite many combinatorial and technical developments. Throughout the reading, there are many comments, remarks and observations that help to understand the global context and see future work to be done. This is an invaluable aid.

The sum of all the results she obtained is very impressive. I'm convinced that all these results will soon become classic in this field. I believe they already have. They perfectly complement the theory of automatic sequences.

To sum up, E. Krawczyk makes an in-depth study of automatic sequence theory from many angles: combinatorics, dynamics, decidability.

This thesis clearly meets and exceeds the requirements of a doctorate.

Paris le 15/6/2023