SUBSTITUTIVE SYSTEMS AND A FINITARY VERSION OF COBHAM'S THEOREM

JAKUB BYSZEWSKI

The talk is based on joint work with Jakub Konieczny and Elżbieta Krawczyk [1].

In the talk, we study substitutive systems generated by nonprimitive substitutions and show that transitive subsystems of substitutive systems are substitutive. As an application we obtain a complete characterisation of the sets of words that can appear as common factors of two automatic sequences defined over multiplicatively independent bases. This generalises the famous theorem of Cobham.

Let $\mathscr{A}$ be a finite alphabet, let $\mathscr{A}^*$ be the set of finite words over $\mathscr{A}$ and let $\mathscr{A}^\omega$ be the set of sequences $x = (a_n)_{n \geqslant 0}$ with values in $\mathscr{A}$. A sequence in $\mathscr{A}^\omega$ is called *purely substitutive* if it is a fixed point of some substitution $\varphi\colon \mathscr{A} \to \mathscr{A}^*$, assumed to be growing, meaning that the length of the words $\varphi^n(a)$ tends to infinity for all letters $a \in \mathscr{A}$. A sequence in $\mathscr{A}^\omega$ is called *substitutive* if it arises from a purely substitutive sequence over some alphabet $\mathscr{B}$ after applying a (possibly non-injective) map $\pi\colon \mathscr{B} \to \mathscr{A}$. We say that a dynamical system $X \subseteq \mathscr{A}^\omega$ is *substitutive* if it arises as the orbit closure of a substitutive sequence x, meaning that it contains all sequences whose factors are also factors of x. If the substitution is of constant length k, we call the system k-automatic. Substitutive systems were extensively studied in the context of primitive substitutions, necessarily restricting such studies to minimal systems. In the recent years there has been growing interest in the study of nonminimal substitutive systems, e.g. with connection to Bratteli diagrams and tiling spaces. Nevertheless, some basic questions seem not to have been studied in this generality.

Our first main result gives the following description of transitive subsystems of substitutive systems.

Theorem A. *Every transitive subsystem of a substitutive system is substitutive. Every transitive subsystem of a k-automatic system is k-automatic.*

In fact, we obtain a much more precise description of substitutive (resp., k-automatic) sequences generating such subsystems. A simplified version of this result is as follows.

Theorem B. *Let x be a purely substitutive sequence produced by a substitution $\varphi\colon \mathscr{A} \to \mathscr{A}^*$, and let X be the orbit closure of x. There exists a power $\tau = \varphi^m$ of φ and a finite set of words $W \subset \mathscr{A}^*$ such that every transitive subsystem $Y \subset X$ can be generated by a sequence $y \in X$ that is a suffix of a biinfinite sequence of the form*

$$\cdots \tau^2(v)\tau(v)vabw\tau(w)\tau^2(w)\cdots \tag{1}$$

for some $v \in W$, $w \in W \setminus \{\epsilon\}$, and $a, b \in \mathscr{A} \cup \{\epsilon\}$.

One of the most fundamental results about automatic sequences is Cobham's theorem, which states that a sequence is simultaneously automatic with respect to two multiplicatively independent bases if and only if it is ultimately periodic. This result has sparked a lot of research and has been generalised to a variety of different settings, including substitutive systems, nonstandard numeration systems, iterated function systems, regular sequences and Mahler functions. We apply the above results above to obtain a strengthening of Cobham's theorem. This extends earlier results by Fagnot [2], which says that if two sequences x and y defined with respect to two multiplicatively independent bases share the same language (that is, the set of their factors), then they are both ultimately periodic, and Mol, Rampersad, Shallit, and Stipulanti [3], who give an explicit bound on the length of a common prefix of x and y that depends on the number of states in the automata generating x and y. Denoting by $\mathscr{L}(z)$ the language of a sequence z, we obtain the following characterisation of the set of common factors of x and y.

Theorem C. *Let $k, l \geqslant 2$ be multiplicatively independent integers, let $\mathscr{A}$ be an alphabet, and let $U \subset \mathscr{A}^*$. The following conditions are equivalent:*

- (i) *there exist a k-automatic sequence x and an l-automatic sequence y such that U is the set of common factors of x and y;*
- (ii) *the set U is a finite nonempty union of sets of the form $\mathscr{L}({}^{\omega}vuw^{\omega})$, where u, v, w are (possibly empty) words over $\mathscr{A}$ and ${}^{\omega}vuw^{\omega} = \cdots vvvuwww \cdots$.*

Note that Cobham's theorem follows immediately from Theorem C. One of the crucial ingredients in the proof of Theorem C is Theorem A applied to k-automatic systems.

References

1. Jakub Byszewski, Jakub Konieczny, Elżbieta Krawczyk, *Substitutive systems and a finitary version of Cobham's theorem*, Combinatorica **41** (2021), no. 6, 765–801.
2. Isabelle Fagnot, *Sur les facteurs des mots automatiques*, Theoret. Comput. Sci. **172** (1997), no. 1-2, 67–89.
3. Lucas Mol, Narad Rampersad, Jeffrey Shallit, and Manon Stipulanti, *Cobham's theorem and automaticity*, Internat. J. Found. Comput. Sci. 30 (2019), no. 8, 1363–1379.