Complexity of infinite words

Svetlana Puzynina

Saint Petersburg State University s.puzynina@gmail.com

For an infinite word $x = x_0x_1x_2\cdots \in \Sigma^{\mathbb{N}}$ over a finite alphabet Σ, the *(factor) complexity* function $p_x : \mathbb{N} \to \mathbb{N}$ counts the number of distinct blocks (or factors) of each length n occurring in x. First introduced by G.A. Hedlund and M. Morse in their 1938 seminal paper on Symbolic Dynamics under the name of *block growth*, the factor complexity provides a useful measure of the extent of randomness of x and more generally of the subshift it generates. They proved that every aperiodic infinite word contains at least $n+1$ distinct factors of each length n. They further showed that an infinite word x has exactly $n+1$ distinct factors of each length n if and only if x is binary, aperiodic and balanced, i.e., x is a *Sturmian word*. Thus Sturmian words are those aperiodic words of the lowest factor complexity. They arise naturally in many different areas of mathematics including combinatorics, number theory and dynamical systems. Sturmian words also have implications in theoretical physics as simple models of quasi-crystals, and in theoretical computer science where they are used in computer graphics as digital approximation of straight lines.

There exists many generalizations of the notion of words complexity and associated extensions of the Morse-Hedlund theorem. For instance, abelian complexity, which counts the number of distinct abelian classes of words of each length n occurring in x; palindrome complexity, which counts the number of distinct palindromes of each length n occurring in x; cyclic complexity, which counts the number of conjugacy classes of factors of each length n occurring in x; arithmetic and maximal patterns complexities. In most cases, these different notions of complexity may be used to detect (and in some cases characterize) ultimately periodic words. Generally, amongst all aperiodic words, Sturmian words have the lowest possible complexity, although in some cases they are not the only ones (for instance, a restricted class of Toeplitz words is found to have the same maximal pattern complexity as Sturmian words).

In this talk we introduce a unified framework via group actions for constructing complexities of infinite words. Factor complexity, abelian complexity and cyclic complexity turn out to be particular cases of this general construction. Jointly with E. Charlier and L. Q. Zamboni, we introduced a concept of *group complexity* of infinite words: We consider infinite sequences of permutation groups $\omega = (G_n)_{n\geq 1}$ with each $G_n \subseteq S_n$. Associated with every such sequence and with every infinite word $x \in \Sigma^{\mathbb{N}}$, a group complexity function $p_{\omega,x} : \mathbb{N} \to \mathbb{N}$ counts, for each n, the number of equivalence classes of factors of x of length n under the action of G_n on Σ^n given by $g * (u_1u_2\cdots u_n) = u_{g^{-1}(1)}u_{g^{-1}(2)}\cdots u_{g^{-1}(n)}$. We show that in this general framework, Sturmian words can also be characterised as aperiodic words of minimal complexity. We further discuss some open questions concerning complexity of infinite words.