k-Abelian Complexity: A New Complexity Measure for Infinite Words

Juhani Karhumäki*
(in collaboration with J. Cassaigne, A. Saarela and L. Zamboni)

March 21, 2014

k-abelian equivalence of words was recently introduced as an equivalence relation in-between the equality and abelian equality of words. It identifies words which have equally many occurrences of all words of length k (and, as a technicality, share a prefix of length $k-1$). For finite words this provides an infinite sequence approximating the equality of words.

The notion calls for a number of challenging problems. Avoidability was among the first having been considered: Which k-abelian repetitions are avoidable in small alphabets? An in-between nature of the notion was revealed. The size of the smallest alphabet where squares are 2-abelian avoidable is four, like in the case of the abelian equality, while that for cubes is two, like in the case of word equality.

One way of defining the complexity of infinite words is to count inequal factors of different lengths. k-abelian complexity states a natural measure of this type. This is the topic of this talk. Bounds for these complexities are given by the numbers of the equivalence classes for different values of k and the cardinality of the alphabet. The upper bounds are known to be polynomial, however, the degrees of these polynomials grow exponentially in k.

The starting points here are as follows. Clearly, all these complexity functions are bounded for ultimately periodic words. The smallest complexities allowing words to be aperiodic are the function $f_{\mathrm{w}}(n) = n+1$ in the case of equality (over a binary alphabet), the function $f_{\mathrm{a}}(n) = 2$ in the abelian case, and the function $f_{k-\mathrm{a}}(n) = \min(n+1, 2k)$ in the k-abelian case. Moreover, each of these extreme cases characterizes Sturmian words among the aperiodic words.

It follows from the first result, known as Morse-Hedlund theorem, that there exists a complexity gap from the bounded complexity to the linear one, when it is measured as the standard factor complexity. No similar gaps exist in the other cases. Indeed, the k-abelian complexity of the famous Thue-Morse word alternates from a fixed constant to an upper bound of order $\Theta(\log n)$. Actually, one can even construct binary words having arbitrarily low unbounded k-abelian complexity.

These results constitute a background for our considerations. We concentrate on the following two questions: (i) How much the complexity can grow when we move from level $k-1$ to level k; (ii) How much the k-abelian complexity can fluctuate, i.e., how

*Department of Mathematics and Statistics, University of Turku, FI-20014 Turku, Finland, karhumak@utu.fi

much the growth rates of the "minimal" and "maximal" values of the complexity can deviate?

Our results give more or less exhaustive answers to these questions. In the first problem the jump can be from a constant to $\Theta(\mathrm{Max}_k(n)/\mathrm{Max}_{k-1}(n))$, where $\mathrm{Max}_k(n)$ denotes the maximal k-abelian complexity. The answer to the question (ii) is more subtle. The fluctuation cannot be maximal, but indeed can be close to that. Namely, we construct examples where the fluctuation is from $\Theta(1)$ to $o(\mathrm{Max}_k(n))$, as well as from $\Theta(n)$ to $\Theta(\mathrm{Max}_k(n))$. On the other hand, the fluctuation all the way from $o(n)$ to $\Theta(\mathrm{Max}_k(n))$ is not possible.