On Periodicities in Strings

(*Abstract*)

Frantisek Franek

Department of Computing and Software
McMaster University, Hamilton, Canada
franek@mcmaster.ca

Periodicities in strings, in particular tandem repetitions, have been of interest to researchers from the beginning. The pioneering work of Corchemore in 1981 showed that the optimal bound for the number of maximal repetitions in a string of length n is of $O(n \log(n))$ complexity and attained by Fibonacci strings, followed closely by the seminal work of Apostolico and Preparat. In 1989, Main introduced an $O(n \log(n))$ algorithm for detection of maximal repetitions where the $\log(n)$ factor represented the size of the alphabet, so for a constant size alphabet, it was a linear algorithm and a linear number of repetitions. From these beginnings, two subsequent lines of research crystallized over several years. The first line of research, dealing with the generalization of maximal repetitions in the form of runs, focused on determining and computing the maximum number of runs, the second line of research focused on determining the maximum number of distinct squares in a string. These two lines of research culminated in respective conjectures: the ***runs conjecture:*** *the maximum number of runs in string is bounded by the length of the string*, and the ***distinct squares conjecture:****the maximum number of distinct squares in string is bounded by the length of the string.*

Intense research on these problems were initiated by the pioneering work of Kolpakov & Kutcherov (1999) for runs and by Fraenkel & Simpson (1998) for distinct squares. Before both conjectures were settled, Deza, myself, and our graduate students Jiang and Baker strengthen both conjecture to hypothesize the bound to be the length of the string less the size of the alphabet of the string. The bounds $n - d$ were based on the d-step approach where the d stands for the size of the alphabet and n the length of the string.

The runs conjecture was settled by Banai et al. in 2015 (published in 2017). The d-step conjecture for runs was proven by Deza et al. in 2017. The d-step conjecture for distinct squares (and hence the distinct squares conjecture) was recently proven by Brlek & Li. The Brlek & Li approach shows why the size of the alphabet naturally occurs in the bound. The d-step aspects of both problems investigated by Deza et al. have some additional consequences beyond the size of the bound, and in both cases they are shown to be optimal as strings of length n with $n-d$ runs or distinct squares are shown.

Acknowledgements

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant program number RGPIN-5504-2018.

Frantisek Franek: On Periodicities in Strings, p. 1.
Proceedings of PSC 2025, Jan Holub and Jan Žďárek (Eds.), ISBN 978-80-01-07461-9