

Doubly-periodic string comparison

Nikita Gaevoy and Alexander Tiskin

Abstract. The longest common subsequence (LCS) problem is a textbook problem in string algorithms and bioinformatics. Given a pair of strings, the problem asks for the length of the longest string that is a subsequence in both input strings. In previous works, the second author developed a novel powerful approach to the LCS and related problems. This approach is based on the algebraic framework of the Hecke monoid, which can be visualised by manipulating braid-like objects that we call sticky braids. Among the many algorithmic problems that can be solved efficiently by the Hecke monoid approach, there is the natural problem of obtaining the LCS for a pair of strings, one or both of which has periodic structure. The case of one periodic string has been considered before; in this work, we extend the solution to the case where both input strings are periodic. The resulting algorithm for doubly-periodic LCS has been engineered by the first author while developing the content for an ICPC training camp. In numerical experiments, the code shows performance that allows one to process doubly-periodic inputs of sizes far beyond the reach of ordinary and singly-periodic LCS algorithms.

Nikita Gaevoy
St. Petersburg State University
St. Petersburg, Russia
Steklov Institute of Mathematics at St. Petersburg
St. Petersburg, Russia
e-mail: nikgaevoy@gmail.com

Alexander Tiskin
St. Petersburg State University
St. Petersburg, Russia
St. Petersburg Electrotechnical University "LETI"
St. Petersburg, Russia
e-mail: a.tiskin@spbu.ru