Jonas Ellert, DIENS, ENS Paris

Faster two-dimensional pattern matching with k mismatches

Joint work with Paweł Gawrychowski, Adam Górkiewicz, and Tatiana Starikovskaya

The classical pattern matching asks for locating all occurrences of one string, called the pattern, in another, called the text, where a string is simply a sequence of characters. Due to the potential practical applications, it is desirable to seek approximate occurrences, for example by bounding the number of mismatches. This problem has been extensively studied, and by now we have a good understanding of the best possible time complexity as a function of *n* (length of the text), *m* (length of the pattern), and *k* (number of mismatches). In particular, we know that for $k = O(\sqrt{m})$, we can achieve quasi-linear time complexity [Gawrychowski and Uznański, ICALP 2018].

We consider a natural generalisation of the approximate pattern matching problem to two-dimensional strings, which are simply square arrays of characters. The exact version of this problem has been extensively studied in the early 90s, motivated by the potential applications in image processing. While periodicity, which is the basic tool for one-dimensional pattern matching, admits a natural extension to two dimensions, it turns out to become significantly more challenging to work with, and it took some time until an alphabet-independent linear-time algorithm has been obtained by Galil and Park [SICOMP 1996].

In the approximate two-dimensional pattern matching, we are given a pattern of size $m \ x \ m$ and a text of size $n \ x$ n, and ask for all locations in the text where the pattern matches with at most k mismatches. The asymptotically fastest algorithm for this algorithm works in $O(k \ n^2)$ time [Amir and Landau, TCS 1991]. We provide a new insight into two-dimensional periodicity to improve on these 30-years old bounds. Our algorithm works in $\tilde{O}((m^2+mk^{5/4})n^2/m^2)$ time, which is $\tilde{O}(n^2)$ for $k = O(m^{4/5})$.