
Fast and accurate sub-cubic matrix multiplication

Clément Pernet^{*1}

¹Grenoble INP - UGA, LJK – Institut polytechnique de Grenoble (Grenoble INP) – France

Abstract

After the introduction by Strassen of the first sub-cubic matrix multiplication algorithm, a series of new algorithms have been proposed improving the asymptotic exponent for this problem. While several have demonstrated their relevance in practice in terms of computation efficiency, their weaker numerical accuracy has always been a concern and was therefore studied by many authors (Brent, Bini and Lotti, Higham, Demmel et al., Ballart et al. Schwartz et al.) with forward error bounds and measurements in practice.

We will present recent work on the topic, which allowed us to produce new algorithms improving on the numerical accuracy of existing ones in the same class while preserving the best known cost estimates in terms of exponent and leading constant.

We first propose a unification and generalization of the state of the art accuracy bounds to work over any pair of input/output norms. This allows us to express the stability growth factor in the Frobenius norm, more amenable to optimization, which we then conduct by exploiting the geometry of the tensor representation of the considered product.

Combining the conversion to alternative basis representations of Schwartz et al. and common sub-expression elimination heuristics, we lastly produce algorithms with the best known complexity for the schemes, and with improved stability, as demonstrated by experiments. This is illustrated on two classes of algorithms, based on either Strassen's 2x2x2 or Smirnov's 3x3x6 schemes.

^{*}Speaker