Efficient Algorithms for Evaluating High-Degree Matrix Polynomials

Niv Hoffman[1], Oded Schwartz[2], Sivan Toledo[1]

[1] *Blavatnik School of Computer Science, Tel-Aviv University, Israel,* stoledo@tau.ac.il

[2] *School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel*

In the early 1970s, Patterson and Stockmeyer discovered a surprising, elegant, and very efficient algorithm to evaluate a matrix polynomial. Later in the 1970s, Van Loan showed how to reduce the memory consumption of their algorithm, addressing an issue that was important back then. There has not been any significant progress in this area since, in spite of dramatic changes in computer architecture and in closely-related algorithmic problems.

We revisit the problem and apply to it both cache-miss reduction methods and new algorithmic tools. Our main contributions are:

- We develop a new block variant of Van-Loan's algorithm, which is usually almost as memory-efficient as Van-Loan's original variant, but much faster.
- We develop two algorithms that reduce the matrix to its Schur form, to speed up the computation relative to both Patterson and Stockmeyer's original algorithm and Van Loan's variants, including the new block variant. One algorithm exploits the fact that multiplying triangular matrices is faster (by up to a factor of 6) than multiplying dense square matrices. The other algorithm partitions the problem into a collection of smaller ones using a relatively recent algorithm due to Davies and Higham.
- We analyze the number of cache misses that the main variants generate, thereby addressing a major cost on modern architecture. The analysis is theoretical and it explains our experimental results, discussed below.
- We evaluate the performance of the direct algorithms (the ones that do not reduce the matrix to Schur form), both existing and new, pinpointing algorithms that are particularly effective.
- We predict the performance of algorithms that reduce the matrix to Schur form using an empirically-based performance model of the performance of their building blocks.