VARIATIONS ON STRASSEN-WINOGRAD ALGORITHMS FOR RECTANGULAR MATRIX MULTIPLICATION

Sarah M. Loos, D. S. Wise*, Indiana University, Department of Computer Science, Bloomington, IN 47405, dswise@cs.indiana.edu

The quadrant additions in the classical $\mathrm{O}\big(n^{lg7}\big)$, Strassen-Winograd matrix-multiplication algorithms are not ideal for preserving boundaries in properly rectangular multiplications. Cleaving an order-p vector into one of order $2^{\lceil \lg p \rceil -1}$, and one of order $n-2^{\lceil \lg p \rceil -1}$, suggests a way to split an $m \times n$ rectangular matrix into quadrants. With $p = \max(m, n)$, the northwest is the only necessarily non-empty of the four quadrants. Adding it to other quadrants creates dense sums that, as factors, make more work for six of the seven recursive multiplications.

A very simple modification of the traditional algorithms, however, has only three recursions on factors built from that dense northwest block. The other four products are far more likely to have zero factors and be annihilated, accelerating the speed of multiplication.

Experimental results plot total computations, time, L1 cache misses, L2 cache misses, and TLB misses for matrices of order 192 up to 4096. They compare among traditional Strassen-Winograd, the faster modified Strassen-Winograd, Douglas's benchmark implementation of Strassen-Winograd, classic recursion algorithm, and the BLAS3 dgemm from the manufacturer's ACML dgemm.

In spite of our simple and effective improvements to the $\mathrm{O}\big(n^{lg7}\big)$ algorithms, the race is still won by the classic $\mathrm{O}\big(n^3\big)$ algorithm which uses the memory hierarchy much better.