30

Gaussian Elimination is Not Optimal (1969)

Volker Strassen

Matrix multiplication is such a simple operation that it is hard to imagine there is anything left to learn about it. To multiply two $n \times n$ matrices A and B and get an $n \times n$ product matrix C, compute the n^2 dot products of rows of A with columns of B. Each of those dot products involves n multiplications of numbers and $n - 1$ additions, for a total of n^3 number multiplications and $n^2(n - 1)$ additions. What else could there be to say?

A great deal, it turns out. The German mathematician Volker Strassen (b. 1936) may have been trying to prove a lower bound, that n^3 multiplications are necessary as well as sufficient, when he discovered this algorithm. The paper entails two remarkable ideas. The first is that a divide-and-conquer, recursive algorithm might beat the conventional algorithm, if there is a way to compute the product of 2×2 matrices with fewer than 8 multiplications. Even after seeing this proved, it still seems surprising that the overhead of implementing the recursion is asymptotically repaid. The other amazing discovery is that two 2×2 matrices can be multiplied with only 7 multiplications. Any high school student might have figured that out scribbling on a pad of paper between classes; in the centuries that people have been multiplying matrices, nobody noticed because nobody had a reason to try. (An analogous algorithm for integer multiplication, due to Karatsuba and Ofman (1962), was already known. It recursively computes the product of two $2n$-bit numbers by three multiplications of n-bit numbers, thus yielding a $O(n^{\log_2 3}) \approx n^{1.58}$ time algorithm for n-bit multiplications, better than the conventional $\Theta(n^2)$ algorithm.)

Strassen's algorithm is tricky to implement both correctly and efficiently, but its utility under a good implementation is not merely theoretical. The discovery that $n \times n$ matrices can be multiplied using $n^{\log_2 7} \approx n^{2.8}$ multiplications led to the still unsolved problem of how much smaller the exponent can be. As of this writing, the answer is no more than 2.373, but no lower bound greater than 2 is known; these more exotic algorithms are not practically useful, however.

This paper, alongside Karatsuba and Ofman (1962), established the divide-and-conquer technique as a tool for a variety of algorithmic problems. The implications for efficiently solving systems of linear equations—which give the paper its title—are remarkable in their own right.

30.1

BELOW we will give an algorithm which computes the coefficients of the product of two square matrices A and B of order n from the coefficients of A and B with less than $4.7 \cdot n^{\log 7}$ arithmetical operations (all logarithms in this paper are for base 2, thus $\log 7 \approx 2.8$; the usual method requires approximately $2n^3$ arithmetical operations). The algorithm induces algorithms for inverting a matrix of order n, solving a system of n linear equations in n unknowns, computing a determinant of order n etc. all requiring less than const $n^{\log 7}$ arithmetical operations.

This fact should be compared with the result of Klyuev and Kokovkin-Shcherbak (1965) that Gaussian elimination for solving a system of linear equations is optimal if one restricts oneself to operations upon rows and columns as a whole. We also note that Winograd (1968) modifies the usual algorithms for matrix multiplication and inversion and for solving systems of linear equations, trading roughly half of the multiplications for additions and subtractions. It is a pleasure to thank D. Brillinger for inspiring discussions about the present subject and S. Cook and B. Parlett for encouraging me to write this paper.

30.2

We define algorithms $\alpha_{m,k}$ which multiply matrices of order $m2^k$, by induction on k: $\alpha_{m,0}$ is the usual algorithm for matrix multiplication (requiring m^3 multiplications and $m^2(m-1)$ additions). $\alpha_{m,k}$ already being known, define $\alpha_{m,k+1}$ as follows:

If A, B are matrices of order $m2^{k+1}$ to be multiplied, write

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},\ B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},\ C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix},$$

where the A_{ik}, B_{ik}, C_{ik} are matrices of order $m2^k$. Then compute

$$\begin{aligned}
\mathrm{I} &= (A_{11} + A_{22})(B_{11} + B_{22}), \\
\mathrm{II} &= (A_{21} + A_{22})B_{11}, \\
\mathrm{III} &= A_{11}(B_{12} - B_{22}), \\
\mathrm{IV} &= A_{22}(-B_{11} + B_{21}), \\
\mathrm{V} &= (A_{11} + A_{12})B_{22}, \\
\mathrm{VI} &= (-A_{11} + A_{21})(B_{11} + B_{12}), \\
\mathrm{VII} &= (A_{12} - A_{22})(B_{21} + B_{22}), \\
C_{11} &= \mathrm{I} + \mathrm{IV} - \mathrm{V} + \mathrm{VII}, \\
C_{21} &= \mathrm{II} + \mathrm{IV}, \\
C_{12} &= \mathrm{III} + \mathrm{V}, \\
C_{22} &= \mathrm{I} + \mathrm{III} - \mathrm{II} + \mathrm{VI},
\end{aligned}$$

using α_{mk} for multiplication and the usual algorithm for addition and subtraction of matrices

of order $m2^k$.

By induction on k one easily sees

Fact 1. $\alpha_{m,k}$ computes the product of two matrices of order $m2^k$ with $m^3 7^k$ multiplications and $(5 + m)m^2 7^k - 6(m2^k)^2$ additions and subtractions of numbers.

Thus one may multiply two matrices of order 2^k with 7^k number multiplications and less than $6 \cdot 7^k$ additions and subtractions.

Fact 2. The product of two matrices of order n may be computed with $< 4.7n^{\log 7}$ arithmetical operations.

Proof. Put $k = \lfloor \log n - 4 \rfloor$, $m = \lfloor n2^{-k} \rfloor + 1$; then $n \le m2^k$. Imbedding matrices of order n into matrices of order $m2^k$ reduces our task to that of estimating the number of operations of $\alpha_{m,k}$. By Fact 1 this number is

$$
\begin{aligned}
&(5 + 2m)m^2 7^k - 6(m2^k)^2 \\
&< (5 + 2(n2^{-k} + 1))(n2^{-k} + 1)^2 7^k \\
&< 2n^3(7/8)^k + 12.03n^2(7/4)^k \quad \text{(here we have used } 16 \cdot 2^k \le n\text{)} \\
&= (2(8/7)^{\log n - k} + 12.03(4/7)^{\log n - k})n^{\log 7} \\
&\le \max_{4 \le t \le 5} (2(8/7)^t + 12.03(4/7)^t)n^{\log 7} \\
&\le 4.7 \cdot n^{\log 7}
\end{aligned}
$$

by a convexity argument.

We now turn to matrix inversion. To apply the algorithms below it is necessary to assume not only that the matrix is invertible but that all occurring divisions make sense (a similar assumption is of course necessary for Gaussian elimination).

We define algorithms $\beta_{m,k}$ which invert matrices of order $m2^k$, by induction on k: $\beta_{m,0}$ is the usual Gaussian elimination algorithm. $\beta_{m,k}$ already being known, define $\beta_{m,k+1}$ as follows:

If A is a matrix of order $m2^{k+1}$ to be inverted, write

$$
A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix},
$$

where the A_{ik}, C_{ik} are matrices of order $m2^k$. Then compute

$$
\begin{aligned}
\mathrm{I} &= A_{11}^{-1},\\
\mathrm{II} &= A_{21} \cdot \mathrm{I},\\
\mathrm{III} &= \mathrm{I} \cdot A_{12},\\
\mathrm{IV} &= A_{21} \cdot \mathrm{III},\\
\mathrm{V} &= \mathrm{IV} - A_{22},\\
\mathrm{VI} &= \mathrm{V}^{-1},\\
C_{12} &= \mathrm{III} \cdot \mathrm{VI},\\
C_{21} &= \mathrm{VI} \cdot \mathrm{II},\\
\mathrm{VII} &= \mathrm{III} \cdot C_{21},\\
C_{11} &= \mathrm{I} - \mathrm{VII},\\
C_{22} &= -\mathrm{VI}
\end{aligned}
$$

using $\alpha_{m,\,k}$ for multiplication, $\beta_{m,\,k}$ for inversion and the usual algorithm for addition or subtraction of two matrices of order $m2^k$.

By induction on k one easily sees

Fact 3. $\beta_{m,\,k}$ computes the inverse of a matrix of order $m2^k$ with $m2^k$ divisions, $\leq \frac{6}{5}m^3 7^k - m2^k$ multiplications and $\leq \frac{6}{5}(5+m)m^2 7^k - 7(m2^k)^2$ additions and subtractions of numbers. The next Fact follows in the same way as Fact 2.

Fact 4. The inverse of a matrix of order n may be computed with $< 5.64 \cdot n^{\log 7}$ arithmetical operations.

Similar results hold for solving a system of linear equations or computing a determinant (use $\det A = (\det A_{11}) \det(A_{22} - A_{21}A_{11}^{-1}A_{12})$).

Reprinted from Strassen (1969), with permission from Springer.