Strassen's Matrix Multiplication Algorithm

Andreas Klappenecker

Suppose you have two $n \times n$ matrices $A = (a_{ij})_{1 \leq i,j \leq n}$ and $B = (b_{jk})_{1 \leq j,k \leq n}$ over a ring such as the real numbers, the complex numbers or the integers. Recall that the matrix product $C = AB$ is defined by the following rule

$$c_{ik} = \sum_{i=1}^{n} a_{ij} b_{jk} \tag{1}$$

for the entries of the matrix $C = (c_{ik})_{1 \leq i,k \leq n}$. For example

$$\begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}.$$

An inspection of this example shows that at most 8 multiplications and 4 additions are necessary to multiply two 2×2-matrices. More generally, we see that at most n multiplications and $n-1$ additions are necessary to calculate the coefficient c_{ik} with the help of equation (1). Thus, the multiplication of two $n \times n$-matrices needs at most n^3 multiplications and $n^3 - n^2$ additions.

We would like to reduce the number of multiplications at the expense of a higher number of additions. This makes sense when the cost of multiplication is higher than the cost of addition. We will make a short digression and have a look at some basic properties of matrices before discussing the core of Strassen's method.

Preliminaries. Recall that the addition $A + B$ of matrices A and B is defined by adding the corresponding entries

$$A + B = (a_{ij} + b_{ij})_{1 \leq i,j \leq n} = (a_{ij})_{1 \leq i,j \leq n} + (b_{ij})_{1 \leq i,j \leq n}.$$

The multiplication of a matrix A with a scalar α is defined by multiplying all entries with this scalar

$$\alpha A = (\alpha a_{ij})_{1 \leq i,j \leq n}$$

This allows us to write matrices in the form of linear combinations of other matrices. A basis for the $n \times n$ matrices is a set of matrices such that any

$n \times n$ matrix can be written as a linear combination in a unique way. For instance, any 2×2-matrix A can be written as a linear combination of

$$E_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \; E_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \; E_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \; E_3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Indeed,

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + a_{12} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + a_{21} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + a_{22} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

and it is easily verified that the coefficients are uniquely determined.

Multiplication Revisited. The multiplication of 2×2 matrices A and B can also be understood in the following way

$$AB = (a_{11}E_0 + a_{12}E_1 + a_{21}E_2 + a_{22}E_3)\,(b_{11}E_0 + b_{12}E_1 + b_{21}E_2 + b_{22}E_3) \qquad (2)$$

To evaluate this expression, we need to know the products E_iE_j. The results are summarized in the following multiplication table:

	E_0	E_1	E_2	E_3
E_0	E_0	E_1	0	0
E_1	0	0	E_0	E_1
E_2	E_2	E_3	0	0
E_3	0	0	E_2	E_3

Expressing C in the form $C = c_{11}E_0 + c_{12}E_1 + c_{21}E_2 + c_{22}E_3$, we see from the multiplication table that two product terms contribute to the term $c_{11}E_0$, namely

$$c_{11}E_0 = a_{11}E_0b_{11}E_0 + a_{12}E_1b_{21}E_2 = (a_{11}b_{11} + a_{12}b_{21})E_0$$

Inspecting the other terms $c_{12}E_1$, $c_{21}E_2$ and $c_{22}E_3$ allows us to recover the standard multiplication rule for matrices.

Bases. Note that the choice of the basis $E_0, \dots, E_3$ in equation (2) was rather arbitrary. In Strassen's method, equation (2) is replaced by a product with respect to two different bases:

$$AB = (\alpha_0A_0 + \alpha_1A_1 + \alpha_2A_2 + \alpha_3A_3)(\beta_0B_0 + \beta_1B_1 + \beta_2B_2 + \beta_3B_3) \qquad (3)$$

where

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},\ A_2 = \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix},\ A_3 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

and

$$B_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ B_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},\ B_2 = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix},\ B_3 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

Let us prove that $A_0, \ldots, A_3$ is indeed a basis. A direct calculation shows that

$$\begin{aligned} E_0 &= A_1, & E_2 &= -A_2 + A_0 - A_1, \\ E_1 &= A_3 - A_0 + A_1, & E_3 &= A_0 - A_1. \end{aligned}$$

Therefore, any 2×2 matrix can be written as a linear combination of the A_i's. To see this, just substitute the above expressions for E_i in an arbitrary linear combination $\sum_{i=0}^{3} \alpha_i E_i$. This shows that $A_0, \ldots, A_3$ is a generating set for the vector space of 2×2-matrices, and hence a basis, since it is minimal. A similar argument shows that the matrices $B_0, \ldots, B_3$ constitute a basis.

Multiplying 2×2 Matrices. After all this preparation, we are now able to derive Strassen's multiplication algorithm for 2×2 matrices. First we need a multiplication table that shows us how to evaluate equation (3):

	B_0	B_1	B_2	B_3
A_0	A_0	B_1	B_2	B_3
A_1	A_1	0	B_2	A_1
A_2	A_2	B_1	A_2	0
A_3	A_3	A_3	0	B_3

As a consequence, we can reduce

$$C = AB = \left(\sum_{i=0}^{3} \alpha_i A_i\right)\left(\sum_{j=0}^{3} \beta_j B_j\right) = \sum_{i,j} \alpha_i \beta_j (A_i B_j) \tag{4}$$

to the following expression

$$C = AB = p_1 A_0 + p_2 A_1 + p_3 A_2 + p_4 A_3 + p_5 B_1 + p_6 B_2 + p_7 B_3 \tag{5}$$

where

$$p_1 = \alpha_0\beta_0; \quad p_2 = \alpha_1(\beta_0+\beta_3); \quad p_3 = \alpha_2(\beta_0+\beta_2); \quad p_4 = \alpha_3(\beta_0+\beta_1);$$
$$p_5 = (\alpha_0+\alpha_2)\beta_1; \quad p_6 = (\alpha_0+\alpha_1)\beta_2; \quad p_7 = (\alpha_0+\alpha_3)\beta_3. \qquad (6)$$

This expression is obtained from the product (4) by collecting terms with the same product. Please consult the multiplication table above to confirm this result. A comparison of coefficients in (5) yields the following simple rules for the entries of the product matrix C:

$$\begin{array}{rclrcl} c_{11} & = & p_1+p_2+p_6+p_7 & c_{12} & = & p_4-p_6 \\ c_{21} & = & -p_3+p_7 & c_{22} & = & p_1+p_3+p_4+p_5 \end{array} \qquad (7)$$

It remains to determine the coefficients α_i and β_i from the entries of the matrices A and B respectively. We expand $A = \sum_i \alpha_i A_i$ for this purpose:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \alpha_0 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \alpha_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

Comparing coefficients gives the equations

$$\begin{array}{rclrcl} a_{11} & = & \alpha_0+\alpha_1; & a_{21} & = & -\alpha_2; \\ a_{12} & = & \alpha_3; & a_{22} & = & \alpha_0+\alpha_2+\alpha_3; \end{array}$$

and solving for the α_i's yields

$$\begin{array}{rclrcl} \alpha_3 & = & a_{12}; & \alpha_0 & = & -a_{12}+a_{21}+a_{22}; \\ \alpha_2 & = & -a_{21}; & \alpha_1 & = & a_{11}+a_{12}-a_{21}-a_{22}. \end{array} \qquad (8)$$

A similar calculation determines the coefficients β_j to be

$$\begin{array}{rclrcl} \beta_0 & = & b_{11}+b_{12}-b_{21} & \beta_2 & = & -b_{12} \\ \beta_1 & = & -b_{11}-b_{12}+b_{21}+b_{22} & \beta_3 & = & b_{21} \end{array} \qquad (9)$$

So the product of the 2×2 matrices A and B can be obtained in the following way. Calculate the intermediate terms α_i and β_i from the coefficients of the matrices A and B. This steps merely needs additions and subtractions. Calculate then the 7 product terms p_i. Use (7) to find the matrix entries of the product $C = AB$.

Exercise. Use Strassen's method to multiply the matrices $A = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

Multiplication of $\mathbf{N \times N}$ Matrices. A remarkable property of Strassen's algorithm for 2×2 matrices is that it can be used to accelerate the multiplication of larger matrices as well. Note that only addition, subtraction and multiplication is used in the 2×2 case. Moreover, the method described above does not exploit the commutative law $ab = ba$. Indeed, all products p_i are of the form

$$\text{(terms from matrix } A\text{)(terms from matrix } B\text{)}$$

Thus the algorithm works for any ring, in particular for matrix rings. The practical consequence is the following: suppose we have two matrices A and B of size $2m \times 2m$. We partition the matrices into four different parts:

$$A = \left(\begin{array}{c|c} a_{11} & a_{12} \\ \hline a_{21} & a_{22} \end{array}\right) \qquad B = \left(\begin{array}{c|c} b_{11} & b_{12} \\ \hline b_{21} & b_{22} \end{array}\right)$$

where a_{ij} and b_{ij} are $m \times m$ matrices. We get $m \times m$ matrices α_i and β_j with the help of the equations (8) and (9). We recurse to form the matrix products (6). Once the products p_i are computed, it is possible to contruct the $m \times m$ matrices c_{ik} using (7). The result is stored in the $2m \times 2m$ matrix

$$C = \left(\begin{array}{c|c} c_{11} & c_{12} \\ \hline c_{21} & c_{22} \end{array}\right).$$

This recursive algorithm can be extended to matrices of arbitrary size by embedding them into larger matrices of size $2^n \times 2^n$.

Complexity. Strassen showed that a slight variation of the above method yields a matrix multiplication algorithm with 7 multiplications and 18 additions. Winograd improved this further to 7 multiplications and 15 additions. If $T(n)$ denotes the total number of arithmetic operations to compute the product of two $n \times n$ matrices, then the recursive method yields

$$T(n) \leq 7T(n/2) + 15n^2$$

arithmetic operations. Therefore, $T(n) \in \Theta(n^{\log_2 7})$ according to the Master theorem.