Strassen's Algorithm Made (Somewhat) More Natural: A Pedagogical Remark

Ann Q. Gates and Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
{agates,vladik}@cs.utep.edu

Abstract

Strassen's 1969 algorithm for fast matrix multiplication [2] is based on the possibility to multiply two 2×2 matrices A and B by using 7 multiplications instead of the usual 8. The corresponding formulas are an important part of any algorithms course, but, unfortunately, even in the best textbook expositions (see, e.g., [1]), they look very ad hoc. In this paper, we show that the use of natural symmetries can make these formulas more natural.

Outline. The goal of this paper is to show that the use of symmetries can make Strassen's formulas for multiplying two 2×2 matrices A and B in 7 multiplications more natural. To achieve this goal, we will first describe two relevant symmetries: the first one is more straightforward, the second one is slightly more implicit. Then, we use these symmetries to select 7 combinations of matrix elements. Finally, we use the same symmetries to pair the combinations corresponding to A and B with each other and thus, to come up with Strassen's formulas.

First symmetry: renaming axes. From the mathematical viewpoint, a 2×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

describes a linear transformation of a 2-dimensional space into itself; the product of two matrices corresponds to the composition of two linear transformations. In this interpretation, the elements of the matrix describe the coordinates of the results $A\mathbf{e}_1$ and $A\mathbf{e}_2$ of applying this transformation to the unit vectors $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$ corresponding to the natural axes: $A\mathbf{e}_1 = a_{11} \cdot \mathbf{e}_1 + a_{21} \cdot \mathbf{e}_2$, and $A\mathbf{e}_2 = a_{12} \cdot \mathbf{e}_1 + a_{22} \cdot \mathbf{e}_2$. From this geometric viewpoint what matters is the transformation itself, while the order in which we describe the axes is irrelevant. If we change this order, we get new unit vectors $\mathbf{e}'_i = \mathbf{e}_{\pi(i)}$, where $\pi(1) = 2$ and $\pi(2) = 1$. In the new axes, the same transformation is represented by a new matrix $A' = \pi(A)$ with $a'_{ij} = a_{\pi(i)\pi(j)}$. Similarly, the transformation corresponding to the matrix B is represented, in the new axes, by the similarly permuted matrix $B' = \pi(B)$ with $b'_{ij} = b_{\pi(i)\pi(j)}$, and the product matrix $C = A \cdot B$ (corresponding to the composition of the two transformations) takes the form $C' = \pi(C)$ with $c'_{ij} = c_{\pi(i)\pi(j)}$.

From the algebraic viewpoint, transformation to the new axes means an index permutation $1 \leftrightarrow 2$. Since the matrix product does not change under the index permutation $1 \leftrightarrow 2$, it is natural to look for formulas for matrix multiplication which do not change under this transformation either.

Second symmetry: inverse transformations. The second natural symmetry comes from considering the inverse transformations A^{-1}, B^{-1}, and C^{-1}, namely, from the fact that if $A \cdot B = C$, then $C^{-1} = B^{-1} \cdot A^{-1}$. For a 2×2 matrix, the inverse is known to be equal to $A^{-1} = \frac{1}{|A|} \cdot A_m$, where $|A|$ denotes the determinant of the matrix A, and

$$A_m \stackrel{\text{def}}{=} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

is a transposition of a matrix formed by minors. The determinants are easy to handle, since they satisfy the equality $|C| = |A| \cdot |B|$. So, the above inverse-transformation property can be reformulated as follows: if $A \cdot B = C$, then $B_m \cdot A_m = C_m$.

The transformation from A to A_m consists of swapping diagonal elements a_{11} and a_{22} and changing the signs of non-diagonal elements a_{12} and a_{21}. Thus, it is reasonable to require that the formulas for matrix multiplication be invariant under this transformation.

Selecting seven linear combinations: general idea. We want to select seven linear combinations of the matrix elements in such a way that the multiplication of A and B can be reduced to multiplying each of seven combinations of elements of A by one of the seven similar combinations of the elements of B. Let us show how the above symmetries can help in this selection.

Selecting seven linear combinations: first preliminary consideration. A general linear combination $k_{11} \cdot a_{11} + \ldots + k_{22} \cdot a_{22}$ requires 4 multiplications to compute. Since our goal is to reduce the overall number of multiplications, we only want linear combinations which do not require multiplications at all, i.e., linear combinations which can be computed by using only addition and subtraction.

Since the ultimate goal is to reduce the computation time, we should also minimize the number of additions and subtractions. It is therefore reasonable to consider only the simplest possible linear combinations, each of which is either one of the matrix's elements, or a sum of two elements, or a difference between two elements.

Selecting seven linear combinations: second preliminary consideration. In linear algebra, we often need to fictitiously enlarge a matrix to a larger one by adding zeros. This is how, e.g., Strassen's algorithm is used for matrices of size $n \times n$ with $n \neq 2^k$: we add 0's so that the size becomes 2^k, and then apply Strassen's method.

In particular, an arbitrary real number a – i.e., a 1×1 matrix – can be represented as a 2×2 matrix if we just "pad" it with zeros:

$$\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}.$$

The product of matrices thus obtained from the numbers a and b is exactly the padded number $a \cdot b$. Thus, real numbers and their multiplication can be viewed as a degenerate case of matrices. In this degenerate case, the number is the element a_{11}.

It is therefore reasonable to require that, first, the value a_{11} itself be one of the linear combinations, and second, that all other basic linear combinations turn into a_{11} when the situation becomes degenerate.

In view of the first preliminary consideration, this second condition means that these combinations are of the type $a_{11} + a_{ij}$ or $a_{11} - a_{ij}$, for some i and j.

Selecting seven linear combinations: actual selection. We have already required that a_{11} should be one of these combinations. Since we want the set of all combinations to be permutation-invariant, the result a_{22} of applying the $1 \leftrightarrow 2$ permutation to a_{11} should also be one of the combinations. So, we already have two of them:

$$a_{11} \qquad\qquad a_{22}. \tag{1}$$

One of the combinations should include an element a_{12}. Due to the second preliminary consideration, this combination should be either of the type $a_{11}+a_{12}$ or $a_{11}-a_{12}$. Without losing generality, we select $a_{11}+a_{12}$ (the combination $a_{11}-a_{12}$ leads to similar formulas).

Since we require that the set of selected combinations be invariant with respect to both symmetries, we conclude that together with this combination, we must have a combination $a_{22}+a_{21}$ which is obtained from this one by permutation:

$$a_{11}+a_{12} \qquad a_{22}+a_{21},$$

as well as the combinations which are obtained from these ones by using the second symmetry (swapping diagonal elements a_{11} and a_{22} and changing the signs of non-diagonal elements a_{12} and a_{21}):

$$\frac{a_{11}+a_{12}}{a_{22}-a_{12}} \qquad \frac{a_{22}+a_{21}}{a_{11}-a_{21}}. \tag{2}$$

We already have $2+4=6$ combinations described by the formulas (1) and (2). To complete our list of seven combination, we must therefore pick one more. Since the list of combinations must be invariant, this seventh combination must not change under both transformations. Since the second transformation changes the sign of non-diagonal elements a_{12} and a_{21}, this seventh combination cannot contain these elements; so the desired combination must contain only diagonal elements a_{11} and a_{22}. Thus, this combination is equal either to the sum $a_{11}+a_{22}$ or to the difference $a_{11}-a_{22}$ between the diagonal elements. The difference does change under swap, so the only invariant combination is the sum:

$$a_{11}+a_{22}. \tag{3}$$

Pairing the combinations: symmetry requirements. This combination should be invariant under the permutation π, so if we pair a combination a with a combination b, then the result $\pi(a)$ of applying the $1 \leftrightarrow 2$ to a should be paired with $\pi(b)$.

Similarly, this combination should be invariant under the second symmetry $A \to A_m$, so if we pair a combination a with a combination b, then the A-combination A which is similar to b_m must be paired with the B-combination of which is similar to a_m. (The second symmetry is slightly more difficult to describe than permutation invariance because for permutation, we have $\pi(A)\cdot\pi(B)=\pi(A\cdot B)$, while for the second symmetry, we have $(A\cdot B)^{-1}=B^{-1}\cdot A^{-1}$ hence $B_m\cdot A_m=(A\cdot B)_m$.)

Pairing the combinations: plan. Our seven combinations form three groups: a group of one (described by the formula (3)), a group of two (described by the formula (1)), and a group of four (described by the formula (2)). It is natural to start with the smallest group (3), then pair combinations from the next smallest group (1), and conclude with the pairing for the largest group (2).

Pairing the combination (3). Let us first find out what is paired with the combination $a=a_{11}+a_{22}$. The pairing should be permutation-invariant, so if a is paired with b, then $\pi(a)$ should be paired with $\pi(b)$. The combination $a_{11}+a_{22}$ is permutation-invariant ($\pi(a)=a$), so if we pair a with b, we should also pair a with $\pi(b)$. Since we want to pair each of the seven A-combinations with one of the seven similar B-combinations and vice versa, each A-combination gets paired with only one B-combination. Thus, we must have $\pi(b)=b$. In other words, the paired combination b must also be permutation-invariant. Among our seven combinations (1)–(3), only one is permutation-invariant: (3). Thus, the combination $a_{11}+a_{22}$ must be paired with the similar combination $b_{11}+b_{22}$. As a result, we get the first product from Strassen's algorithm:

$$(a_{11}+a_{22})\cdot(b_{11}+b_{22}). \tag{4}$$

Pairing the combinations (1). Now, we go to the next largest group (1). Let us start by deciding what to pair a_{11} with. We already used one B-combination, so we can, in principle, pair a_{11} with six remaining B-combinations. Which of them should we choose? We can dismiss a pairing of a_{11} with b_{22} since the product $a_{11} \cdot b_{22}$ does not occur in any of the four elements of the product matrix $c_{ij} = a_{i1} \cdot b_{1j} + a_{i2} \cdot b_{2j}$ and is, therefore, useless. So, we are left with five pairing possibilities.

Up to now, the selection of combinations and the pairing of combination was uniquely determined by symmetries. Now is the first time when symmetry is not sufficient. We have to explicitly mention that only one of the five pairings leads to fast matrix multiplication: the pairing of a_{11} with $b_{22} - b_{12}$:

$$a_{11} \cdot (b_{22} - b_{12}). \tag{5}$$

From the viewpoint of simplicity, the fact that we have to use a non-symmetry argument can be perceived as a negative. The positive side is that this is the only use of non-symmetries in the selection of pairings; for all other combinations, pairing follows from symmetries. Indeed, due to permutation-invariance, the result a_{22} of permuting a_{11} should be paired with $b_{11} - b_{21}$:

$$a_{22} \cdot (b_{11} - b_{21}). \tag{6}$$

Similarly, if we apply the second symmetry to the combination (5), we conclude that the combination b_{22} (which is the B-analogue of $(a_{11})_m = a_{22}$) should be paired with the combination $a_{11} + a_{12}$ – which is the B-analogue of $(b_{22} - b_{12})_m = b_{11} + b_{12}$. As a result, we get the following pairing:

$$(a_{11} + a_{12}) \cdot b_{22}. \tag{7}$$

If we apply the second symmetry to the combination (6), we get the pairing

$$(a_{22} + a_{21}) \cdot b_{11}. \tag{8}$$

Pairing the remaining combinations (2). After the pairing (4)–(8), we have two un-paired A-combinations $a_{22} - a_{12}$ and $a_{11} - a_{21}$, and two un-paired B-combinations $b_{11} + b_{12}$ and $b_{22} + b_{21}$. Which should we pair with which? Let us start with finding a pair for $a_{11} - a_{21}$. Pairing this combination with $b_{22} - b_{21}$ would be useless, because neither of the four products $a_{11} \cdot b_{22}$, $a_{11} \cdot b_{21}$, $a_{21} \cdot b_{22}$, and $a_{21} \cdot b_{21}$ occurs in the definition of any of the four elements of the product matrix. Thus, we must pair $a_{11} - a_{21}$ with $b_{11} + b_{21}$:

$$(a_{11} - a_{21}) \cdot (b_{11} + b_{21}). \tag{9}$$

Now, we only have one A-combination and one B-combination left, so we must pair them with each other:

$$(a_{22} - a_{12}) \cdot (b_{22} + b_{12}). \tag{10}$$

Conclusion. We got the formulas (4)–(10) which are exactly Strassen's formulas.

Acknowledgments. This work was partially supported by NASA under cooperative agreement NCC5-209 and by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-00-1-0365.

References

[1] Th. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to algorithms*, MIT Press, Cambridge, MA, and Mc-Graw Hill Co., N.Y., 1994.

[2] V. Strassen, "Gaussian Elimination is Not Optimal", *Numerische Mathematick*, 1969, Vol. 14, No. 3, pp. 354–356.