Strassen’s $2 \times 2$ matrix multiplication algorithm: 
A conceptual perspective

Christian Ikenmeyer$^1$ and Vladimir Lysikov$^2$

$^1$Max Planck Institute for Software Systems, Saarland Informatics Campus
$^2$Department of Computer Science, Saarland University
cikenmey@mpi-sws.org vlysikov@cs.uni-saarland.de

February 14, 2019

Abstract

The main purpose of this paper is pedagogical.

Despite its importance, all proofs of the correctness of Strassen’s famous 1969 algorithm to multiply two $2 \times 2$ matrices with only seven multiplications involve some basis-dependent calculations such as explicitly multiplying specific $2 \times 2$ matrices, expanding expressions to cancel terms with opposing signs, or expanding tensors over the standard basis. This makes the proof nontrivial to memorize and many presentations of the proof avoid showing all the details and leave a significant amount of verifications to the reader.

In this note we give a short, self-contained, basis-independent proof of the existence of Strassen’s algorithm that avoids these types of calculations. We achieve this by focusing on symmetries and algebraic properties.

Our proof can be seen as a coordinate-free version of the construction of Clausen from 1988, combined with recent work on the geometry of Strassen’s algorithm by Chiantini, Ikenmeyer, Landsberg, and Ottaviani from 2016.

1 Introduction

The discovery of Strassen’s matrix multiplication algorithm [28] was a breakthrough result in computational linear algebra. The study of fast (subcubic) matrix multiplication algorithms initiated by this discovery has become an important area of research (see [3] for a survey and [21] for the currently best upper bound on the complexity of matrix multiplication). Fast matrix multiplication has countless applications as a subroutine in algorithms for a wide variety of problems, see e.g. [7, §16] for numerous applications in computational linear algebra. In practice, algorithms more sophisticated than Strassen’s are almost never implemented, but Strassen’s algorithm is used for multiplication of large matrices (see [12, 25, 19] on practical fast matrix multiplication).

The core of Strassen’s result is an algorithm for multiplying $2 \times 2$ matrices with only 7 multiplications instead of 8. It is a bilinear algorithm, which means that it arises from a decomposition of the form

$$XY = \sum_{k=1}^{7} u_k(X)v_k(Y)W_k,$$

where $u_k$ and $v_k$ are cleverly chosen linear forms on the space of $2 \times 2$ matrices and $W_k$ are seven explicit $2 \times 2$ matrices. Because of this structure it can be applied to block matrices, and its recursive application results in an algorithm for the multiplication of two $n \times n$ matrices using $O(n^{\log_2 7})$ arithmetic operations (see [7, §15.2] or [3] for details).

Because of the great importance of Strassen’s algorithm, our goal is to understand it on a deep level. In Strassen’s original paper, the linear forms $u_k, v_k$, and the matrices $W_k$ are given, but the verification of the correctness of the algorithm is left to the reader. Unfortunately, such a description does not yield many further immediate insights.
Shortly after Strassen’s paper, Gastinel [14] published a proof of the existence of decomposition (⋆) using simple algebraic transformations that is much easier to follow and verify. Many other papers provide alternative descriptions of Strassen’s algorithm or proofs of its existence. Brent [4] and Paterson [26] present the algorithm in a graphical form using $4 \times 4$ diagrams indicating which elements of the two matrices are used. A more formal version of these diagrams are matrices of linear forms, which are used, for example, by Fiduccia [13] (essentially the same proof appears in [29]), Brockett and Dobkin [5] and Lafon [20]. Makarov [22] gives a proof that uses ideas of Karatsuba’s algorithm for the efficient multiplication of polynomials. Büchi and Clausen [6] connect the existence of Strassen’s algorithm to the existence of special bases of the space of $2 \times 2$ matrices in which the multiplication table has a specific structure (their results are more general and apply not only to matrix multiplication). Alexeyev [1] describes several algorithms for matrix multiplication as embeddings of the matrix algebra into a 7-dimensional nonassociative algebra with a special properties.

Verification of these proofs usually requires simple, but lengthy computations: expansion of explicit decompositions in some basis, multiplication of several matrices or following chains of algebraic transformations in which careful attention to details is required. To obtain a more conceptual proof of the existence of Strassen’s algorithm, we do not focus on the explicit algorithm, but on the algebraic properties of the $2 \times 2$ matrices, their transformations and symmetries of Strassen’s algorithm. It is well-known that the decomposition (⋆) is not unique. Given one decomposition, we can obtain another one by applying the identity

$$XY = A^{-1} \left[ (AXB^{-1})(BYC^{-1}) \right] C$$

and using the original decomposition for the product in the square brackets. Alternatively, we can talk about $2 \times 2$ matrices as linear maps between 2-dimensional vector spaces. Any choice of bases in these vector spaces gives a new bilinear algorithm. De Groote [18] proved that the algorithm with seven multiplications is unique up to these transformations (this result is also announced without a proof in [23], see also [24]). Thus, Strassen’s algorithm is unique in this sense and there should be a coordinate-free description of this algorithm which does not use explicit matrices. One such description is given in [10] and the proof of its correctness uses the fact that matrix multiplication is the unique (up to scale) bilinear map invariant under the transformations described above. This is a nontrivial fact which requires representation theory to prove. Moreover, the verification of the correctness in [10] is left to the reader.

Symmetries of Strassen’s algorithm are also useful for its understanding. Clausen [11] gives a description of Strassen’s algorithm in terms of special bases, as in [6], and notices that the elements of these bases form orbits under the action of the symmetric group $S_3$ on the space of $2 \times 2$ matrices defined via conjugation with specific matrices, i.e., Strassen’s algorithm is invariant under this action. Clausen’s construction is also described in [7, Ch.1]. Grochow and Moore [16, 17] generalize Clausen’s construction to $n \times n$ matrices using other finite group orbits. Another symmetry is only apparent in the trilinear representation of the algorithm: the decompositions (⋆) are in one-to-one correspondence with decompositions of the trilinear form $\text{tr}(XYZ)$ of the form

$$\text{tr}(XYZ) = \sum_{k=1}^{7} u_k(X)v_k(Y)w_k(Z)$$

where $u_k$, $v_k$ and $w_k$ are linear forms. The decomposition corresponding to Strassen’s algorithm is then invariant under the cyclic permutation of matrices $X, Y, Z$. This symmetry is exploited in the proof of Chatelin [9], which uses properties of polynomials invariant under this symmetry. He also notices the importance of a matrix which is related to the $S_3$ symmetry discussed above. The symmetries of Strassen’s algorithm are explored in detail in [8, 10]. Several earlier publications note their importance [15, 27]. The paper [2] explores symmetries of algorithms for $3 \times 3$ matrix multiplication.

In this paper we provide a proof of Strassen’s result which is

- **coordinate-free** — we do not use explicit matrices, which allows us to focus on the algebraic properties required to prove the correctness of the algorithm. We avoid all tedious explicit calculations, in particular any expansions of expressions and any verification of explicit sign cancellations. Our proof can be seen as a coordinate-free version of Clausen’s construction.

- **elementary** — our proof uses only simple facts from basic linear algebra and does not require knowledge of representation theory. This is also why we do not use tensor language. Proofs from [10] and [17] are based on more complicated mathematics and may offer other insights.

Formally, the result that we prove is the following.
Theorem 1 (Strassen [28]). Fix any field $F$. There exist fourteen linear forms $u_1, \ldots, u_7, v_1, \ldots, v_7 : \mathbb{F}^{2 \times 2} \to \mathbb{F}$ and seven matrices $W_1, \ldots, W_7 \in \mathbb{F}^{2 \times 2}$ such that for all pairs of $2 \times 2$ matrices $X$ and $Y$ the product satisfies

$$XY = \sum_{k=1}^{7} u_k(X)v_k(Y)W_k.$$  

(*)

Acknowledgements. The authors thank Alin Bostan, Joshua Grochow and anonymous referees for comments and pointers to the literature.

2 Preliminaries from linear algebra

The trace $\text{tr}(A)$ of a square matrix $A$ is the sum of its diagonal entries. If $\text{tr}(A)$ is zero, then the matrix $A$ is called traceless. Taking the trace of a product of (rectangular) matrices is invariant under cyclic shifts:

$$\text{tr}(A_1A_2 \cdots A_n) = \text{tr}(A_2 \cdots A_nA_1).$$

As a consequence, the trace of a matrix is invariant under conjugations:

$$\text{tr}(B^{-1}AB) = \text{tr}(ABB^{-1}) = \text{tr}(A).$$

Another implication is that if $u$ is a column vector and $v^T$ is a row vector, then $v^T u = \text{tr}(uv^T)$.

The characteristic polynomial of a $2 \times 2$ matrix $A$ is $\lambda^2 - \text{tr}(A) \lambda + \det(A)$. The Cayley–Hamilton theorem says that substituting $A$ for $\lambda$ yields the zero matrix.

3 Rotational symmetry

In this section we collect some standard facts about rotation matrices. We think of the $2 \times 2$ matrix $D$ as a rotation of the plane by $120^\circ$, but to make our approach work over every field we use a more algebraic definition for $D$.

Let $D$ have determinant 1 and trace $-1$, that is, $D$ has characteristic polynomial $\lambda^2 + \lambda + 1$. We assume that $D$ is not a multiple of the identity id (this is implicitly satisfied if the characteristic is not 3). For example, we could choose $D = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}$, the matrix that cyclically permutes the three vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Claim 2. The matrix $D$ has the following properties: $D^2 = \text{id}$ and $\text{id} + D + D^{-1} = 0$ and $\text{tr}(D^{-1}) = -1$.

Proof. The characteristic polynomial of $D$ is $\lambda^2 + \lambda + 1$. By the Cayley–Hamilton theorem $D^2 + D + \text{id} = 0$. Multiplying by $D$ we obtain $D + D^2 + D^3 = 0 = \text{id} + D + D^2$ and hence $D^3 = \text{id}$. Therefore $D^{-1} = D^2$ and thus $\text{id} + D + D^{-1} = 0$. This implies $\text{tr}(D^{-1}) = -\text{tr}(\text{id}) - \text{tr}(D) = -1$. □

For every column vector $u$ define $u^\perp$ as the row vector satisfying conditions $u^\perp u = 0$ and $u^\perp Du = 1$. If $u$ is not an eigenvector of $D$, then $u$ and $Du$ are linearly independent, so $u^\perp$ is uniquely defined. If, on the other hand, $u$ is an eigenvector of $D$, the two conditions are inconsistent and $u^\perp$ is undefined.

We fix a vector $u$ that is not an eigenvector of $D$ and define $u^\perp$ as above. In our example we could choose $u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, which is not an eigenvector of $D$.

A first simple observation relates $u^\perp$ and $(Du)^\perp$:

Claim 3. $u^\perp D^{-1} = (Du)^\perp$.

Proof. We need to verify the two defining properties for $(Du)^\perp$. We have $(u^\perp D^{-1})(Du) = u^\perp u = 0$ and $(u^\perp D^{-1})D(Du) = u^\perp Du = 1$ as required. □

The following observation complements the fact that $u^\perp Du = 1$.

Claim 4. $u^\perp D^{-1} u = -1$.

Proof. Using Claim 2 we have $\text{id} + D + D^{-1} = 0$ and thus $u^\perp u + u^\perp Du + u^\perp D^{-1} u = 0$.

Since $u^\perp u = 0$ and $u^\perp Du = 1$, the claim follows. □
4 Seven multiplications suffice

In this section we apply our structural insights from Section 3 to prove Theorem 1. We set $M := uu^\perp$. Clearly $\text{tr}(M) = u^\perp u = 0$ and we obtain the following identities that can be used to simplify products of $M$, $D$, and $D^{-1}$:

Claim 5. $M^2 = 0$ and $MDM = M$ and $MD^{-1}M = -M$.

Proof.

\begin{align*}
M^2 &= (uu^\perp)(uu^\perp) = u(u^\perp u)u^\perp = 0. \\
MDM &= (uu^\perp)D(uu^\perp) = u(u^\perp Du)u^\perp = uu^\perp = M. \\
MD^{-1}M &= (uu^\perp)D^{-1}(uu^\perp) = u(u^\perp D^{-1}u)u^\perp = -uu^\perp = -M,
\end{align*}

where in the last line we used Claim 4.

By Claim 2, conjugation with $D$ is a map of order 3 on the vector space of all $2 \times 2$ matrices, i.e., for any matrix $A$ there is a triple of conjugates $A \mapsto D^{-1}AD \mapsto DAD^{-1} \mapsto A$. Moreover, if $A$ is traceless, then so are its conjugates.

Claim 6. The matrices $M$, $D^{-1}MD$, and $DMD^{-1}$ form a basis of the vector space of traceless matrices.

Proof. Since $M$ is traceless, its conjugates are also traceless. Hence it is enough to prove that $M$, $D^{-1}MD$ and $DMD^{-1}$ are linearly independent.

Since $u$ is not an eigenvector of $D$, the vectors $u$ and $Du$ are linearly independent and thus form a basis of the space of column vectors. The row vectors $u^\perp$ and $u^\perp D^{-1} = (Du)^\perp$ (Claim 3) are orthogonal to $u$ and $Du$, respectively. Therefore they form a basis of the space of row vectors. Thus, the four matrices

$$u \cdot u^\perp = M, \quad u \cdot u^\perp D^{-1} = MD^{-1}, \quad Du \cdot u^\perp = DM, \quad Du \cdot u^\perp D^{-1} = DMD^{-1}$$

form a basis of the space of $2 \times 2$ matrices. The matrices $M$ and $DMD^{-1}$ are contained in this basis. Adding up all four matrices, we get $(id + D)M(id + D^{-1})$, which can be simplified to $(-D^{-1})M(-D) = D^{-1}MD$ using Claim 2. Therefore the matrices $M, DMD^{-1}, D^{-1}MD$ are linearly independent.

Since $D$ and $D^{-1}$ have trace $-1 \neq 0$ (Claim 2), adding $D$ or $D^{-1}$ to the basis in Claim 6 yields two bases for the full space of $2 \times 2$ matrices: \{\(D, M, D^{-1}MD, DMD^{-1}\)\} and \{\(D^{-1}, M, D^{-1}MD, DMD^{-1}\)\}.

Using the properties $D^2 = D^{-1}$, $D^{-2} = D$ and $M^2 = 0$ from Claim 2 and Claim 5, we can write down the multiplication table with respect to these two bases. We further simplify it using the identities $MDM = M$ and $MD^{-1}M = -M$ from Claim 5.

|     | $D^{-1}$ | $M$ | $D^{-1}MD$ | $DMD^{-1}$ |
|-----|---------|-----|------------|-------------|
| $D$ | id     | $DMD$ | $MD$       | $D^{-1}MD^{-1}$ |
| $M$ | $MD^{-1}$ | 0   | $MDMD^{-1}$ | $MD^{-1}$ |
| $D^{-1}MD$ | $D^{-1}M$ | $D^{-1}MDM$ | $0$ | $DMD^{-1}MD^{-1}$ |
| $DMD^{-1}$ | $DMD$ | $DMD^{-1}M$ | $DMDMD^{-1}$ | $0$ |

Proof of Theorem 1. Notice that in the body of the table only (scalar multiples of) 7 matrices are used, and the entries are aligned in such a way that two occurrences of the same matrix are either in the same row or in the same column. At this point we are done proving Theorem 1, because the existence of such a pattern gives a simple way to construct a matrix multiplication algorithm as follows. To multiply matrices $A$ and $B$, represent them in the bases \{\(D, M, D^{-1}MD, DMD^{-1}\)\} and \{\(D^{-1}, M, D^{-1}MD, DMD^{-1}\)\}, respectively:

\[
X = x_1 D + x_2 M + x_3 D^{-1}MD + x_4 DMD^{-1}
\]

\[
Y = y_1 D^{-1} + y_2 M + y_3 D^{-1}MD + y_4 DMD^{-1}
\]

(4.1)
Note that the $x_i$ are linear forms in the entries of $X$ and the $y_j$ are linear forms in the entries of $Y$. We expand the product $XY$ and group together summands according to the table:

$$
XY = x_1 \times y_1 \times \text{id} \\
+ x_2 \times (y_1 + y_4) \times MD^{-1} \\
+ x_3 \times (y_1 + y_2) \times D^{-1}M \\
+ x_4 \times (y_1 + y_3) \times DMD \\
+ (x_1 - x_4) \times y_2 \times DM \\
+ (x_1 - x_2) \times y_3 \times MD \\
+ (x_1 - x_3) \times y_4 \times D^{-1}MD^{-1} \\
\uparrow \\
u_k(X) \quad v_k(Y) \quad W_k
$$

This finishes the proof.

Remark. Taking the trace in (4.1) and using the fact that $M$ and its conjugates are traceless, we see that $\text{tr}(X) = x_1$, $\text{tr}(D) = -x_1$, and $\text{tr}(Y) = -y_1$. Thus the first of the 7 summands is $\text{tr}(X) \text{tr}(Y) \text{id}$.

References

[1] Valery B. Alekseyev. Maximal extensions with simple multiplication for the algebra of matrices of the second order. Discrete Math. Appl., 7(1):89–102, 1996. doi:10.1515/dma.1997.7.1.89.

[2] Grey Ballard, Christian Ikenmeyer, Joseph M. Landsberg, and Nick Ryder. The geometry of rank decompositions of matrix multiplication II: $3 \times 3$ matrices. Preprint arXiv:1801.00843, arXiv, 2018. arXiv:1801.00843.

[3] Markus Bläser. Fast matrix multiplication. Theory Comput. Grad. Surv., 5:1–60, 2013. doi:10.4086/toc.gs.2013.005.

[4] Richard P. Brent. Algorithms for matrix multiplication. Tech. Report STAN-CS-70-157, Stanford University, Department of Computer Science, 1970. doi:10.21236/ad0705509.

[5] Roger W. Brockett and David Dobkin. On the optimal evaluation of a set of bilinear forms. In Proc. 5th ACM STOC, pages 88–95, 1973. doi:10.1145/800125.804039.

[6] Werner Büchi and Michael Clausen. On a class of primary algebras of minimal rank. Linear Algebra Appl., 69:249–268, 1985. doi:10.1016/0024-3795(85)90080-1.

[7] Peter Bürgisser, Michael Clausen, and M. Amin Shokrollahi. Algebraic Complexity Theory, volume 315 of Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 1997. doi:10.1007/978-3-662-03338-8.

[8] Vladimir P. Burichenko. On symmetries of the Strassen algorithm. Preprint arXiv:1408.6273, arXiv, 2014. arXiv:1408.6273.

[9] Philippe Chatelin. On transformations of algorithms to multiply $2 \times 2$ matrices. Inf. Process. Lett., 22(1):1–5, 1986. doi:10.1016/0020-0190(86)90033-5.

[10] Luca Chiantini, Christian Ikenmeyer, Joseph M. Landsberg, and Giorgio Ottaviani. The geometry of rank decompositions of matrix multiplication I: $2 \times 2$ matrices. Exp. Math., Advance online publication, 2017. doi:10.1080/10586458.2017.1403981.

[11] Michael Clausen. Beiträge zum Entwurf schneller Spektraltransformationen. Habilitationsschrift, Universität Karlsruhe, 1988.

[12] Jean-Guillaume Dumas and Victor Y. Pan. Fast matrix multiplication and symbolic computation. Preprint arXiv:1612.05766, arXiv, 2016. arXiv:1612.05766.

[13] Charles M. Fiduccia. On obtaining upper bounds on the complexity of matrix multiplication. In Complexity of Computer Computations, pages 31–40, 1972. doi:10.1007/978-1-4684-2001-2_4.
[14] Noël Gastinel. Sur le calcul des produits de matrices. *Numer. Math.*, 17(3):222–229, 1971. doi:10.1007/BF01436378.

[15] Ann Q. Gates and Vladik Kreinovich. Strassen’s algorithm made (somewhat) more natural: A pedagogical remark. *Bull. EATCS*, 73:142–145, 2001. URL: https://digitalcommons.utep.edu/cs_techrep/502/.

[16] Joshua A. Grochow and Cristopher Moore. Matrix multiplication algorithms from group orbits. Preprint arXiv:1612.01527, arXiv, 2016. arXiv:1612.01527.

[17] Joshua A. Grochow and Cristopher Moore. Designing Strassen’s algorithm. Preprint arXiv:1708.09398, arXiv, 2017. arXiv:1708.09398.

[18] Hans F. de Groote. On varieties of optimal algorithms for the computation of bilinear mappings II: Optimal algorithms for $2 \times 2$-matrix multiplication. *Theor. Comp. Sci.*, 7(2):127–184, 1978. doi:10.1016/0304-3975(78)90045-2.

[19] Jianyu Huang, Leslie Rice, Devin A. Matthews, and Robert A. van de Geijn. Generating families of practical fast matrix multiplication algorithms. In *Proc. IPDPS 2017*, pages 656–667, 2017. doi:10.1109/IPDPS.2017.56.

[20] Jean-Claude Lafon. Optimum computation of $p$ bilinear forms. *Linear Algebra Appl.*, 10(3):225–240, 1975. doi:10.1016/0024-3795(75)90071-3.

[21] François Le Gall. Powers of tensors and fast matrix multiplication. In *Proc. ISSAC 2014*, pages 296–303, 2014. doi:10.1145/2608628.2608664.

[22] Oleg M. Makarov. The connection between two multiplication algorithms. *USSR Comput. Math. Math. Phys.*, 15(1):218–223, 1975. doi:10.1016/0041-5553(75)90149-4.

[23] Victor Y. Pan. О схемах вычисления произведений матриц и обратной матрицы [On algorithms for matrix multiplication and inversion]. *Усп. мат. наук*, 27(5(167)):249–250, 1972. Translation available in [24]. URL: http://mi.mathnet.ru/umn5125.

[24] Victor Y. Pan. Better late than never: Filling a void in the history of fast matrix multiplication and tensor decompositions. Preprint arXiv:1411.1972, arXiv, 2014. arXiv:1411.1972.

[25] Victor Y. Pan. Fast matrix multiplication and its algebraic neighbourhood. *Sb. Math.*, 208(11):1661–1704, 2017. doi:10.1070/SM8833.

[26] Mike Paterson. Complexity of product and closure algorithms for matrices. In *Proc. ICM 1974*, volume 2, pages 483–489, 1974. URL: https://www.mathunion.org/fileadmin/ICM/Proceedings/ICM1974.2/ICM1974.2.ocr.pdf#page=491 [cited 2018-02-03].

[27] Mike Paterson. Strassen symmetries. Presentation at Leslie Valiant’s 60th birthday celebration, 30.05.2009, Bethesda, Maryland, USA, 2009. URL: https://www.cis.upenn.edu/~mkearns/valiant/paterson.ppt [cited 2018-02-03].

[28] Volker Strassen. Gaussian elimination is not optimal. *Numer. Math.*, 13(4):354–356, 1969. doi:10.1007/BF02165411.

[29] Gideon Yuval. A simple proof of Strassen’s result. *Inf. Process. Lett.*, 7(6):285–286, 1978. doi:10.1016/0020-0190(78)90018-2.