On Multiplication of 2 × 2 Matrices

S. WINOGRAD
IBM Thomas J. Watson Research Center
Yorktown Heights, New York

Communicated by Alan J. Hoffman

ABSTRACT

The two main results of this note are:

(i) The minimum number of multiplications required to multiply two 2 × 2 matrices is seven.

(ii) The minimum number of multiplications/divisions required to multiply two complex numbers is three.

1. INTRODUCTION

In [2] Strassen showed that two 2 × 2 matrices can be multiplied using only seven multiplications without using the commutativity law. Hopcroft and Kerr showed in [1] that without using the commutativity law this number of multiplications is minimal. The purpose of this note is to show that the product of two 2 × 2 matrices requires at least seven multiplications, even when the commutativity law is used.

We will use the notation of [3, 4] as well as some of the results reported in these papers. In particular, we will need the following result: Let F be a field, let G be a subfield of F, and let $x_1, \ldots, x_n$ be indeterminates. Every algorithm which computes $\boldsymbol{\psi} = \Phi\mathbf{x}$ requires at least m multiplications if there are m columns of Φ such that no nontrivial linear combination of these columns with coefficients in G yields a vector all of whose components lie in G; where $\boldsymbol{\psi}$ is the (column) vector $(\psi_1, \psi_2, \ldots, \psi_t)$ of the expressions to be computed, Φ is a $t \times n$ matrix with entries in F, and $\mathbf{x}$ is the (column) vector $(x_1, x_2, \ldots, x_u)$. Moreover, the result holds even if multiplications by an element $g \in G$ are not counted.

2. NOTATION AND PRELIMINARY LEMMAS

Let

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

Linear Algebra and Its Applications 4(1971), 381–388

be the two 2×2 matrices to be multiplied. The evaluation of an N-step algorithm α is a function e_α: $\{1, 2, \ldots, N\} \to Q[a_{11}, \ldots, a_{22}, b_{11}, \ldots, b_{22}]$ (where Q is the field of rational numbers) such that either $e_\alpha(j) \in Q \cup \{a_{11}, \ldots, a_{22}, b_{11}, \ldots, b_{22}\}$ or else $e_\alpha(j)$ can be obtained by adding or subtracting or multiplying $e_\alpha(j_1)$ and $e_\alpha(j_2)$, where $j_1, j_2 < j$. The algorithm can compute the product of the two matrices if there exist j_1, j_2, j_3, j_4 such that

$$e_\alpha(j_1) = a_{11}b_{11} + a_{12}b_{21},$$

$$e_\alpha(j_2) = a_{11}b_{12} + a_{12}b_{22},$$

$$e_\alpha(j_3) = a_{21}b_{11} + a_{22}b_{21},$$

$$e_\alpha(j_4) = a_{21}b_{12} + a_{22}b_{22}.$$

If the jth step of the algorithm is such that $e_\alpha(j) \in Q \cup \{a_{11}, \ldots, a_{22}, b_{11}, \ldots, b_{22}\}$, then it is called a *data step*; otherwise it is called a *computation step*. If $e_\alpha(j)$ is obtained by adding, subtracting, or multiplying $e_\alpha(j_1)$ and $e_\alpha(j_2)$, then $e_\alpha(j_1)$ and $e_\alpha(j_2)$ are called the arguments of this computation step.

LEMMA 2.1. *Let α be an algorithm for multiplying two 2×2 matrices and let $p_1, p_2, \ldots, p_m$ be the results of the computation steps which are multiplications. Then for every step k there exist rational numbers r_i, $r_{i,j}$, and $r'_{i,j}$ (depending on k) such that*

$$e_\alpha(j) = r_0 + \sum_{i=1}^{m} r_i p_i + \sum_{i=1}^{2}\sum_{j=1}^{2} r_{i,j}a_{i,j} + \sum_{i=1}^{2}\sum_{j=1}^{2} r'_{i,j}b_{i,j}.$$

Proof. By definition $e_\alpha(1) \in Q \cup \{a_{1,1}, \ldots, a_{2,2}, b_{1,1}, \ldots, b_{2,2}\}$, so it is of the desired form. Also, if the jth step is a multiplication computation step, then $e_\alpha(j) = p_k$ for some k. Assume that the result of the lemma holds for all steps smaller than l. Then either $e_\alpha(l) \in Q \cup \{a_{11}, \ldots, b_{22}\}$, in which case it is of the form defined by Lemma 2.1, or else $e_\alpha(l) = e_\alpha(j_1) \pm e_\alpha(j_2)$, and by assumption

$$e_\alpha(j_1) = r_0 + \sum_{i=1}^{m} r_i p_i + \sum_{j=1}^{2}\sum_{i=1}^{2} r_{i,j}a_{i,j} + \sum_{j=1}^{2}\sum_{i=1}^{2} r'_{i,j}b_{i,j}$$

and

$$e_\alpha(j_2) = s_0 + \sum_{i=1}^{m} s_i p_i + \sum_{j=1}^{2}\sum_{i=1}^{2} s_{i,j}a_{i,j} + \sum_{j=1}^{2}\sum_{i=1}^{2} s'_{i,j}b_{i,j},$$

so

$$e_\alpha(l) = (r_0 + s_0) + \sum_{i=1}^{m} (r_i + s_i)p_i + \sum_{j=1}^{2}\sum_{i=1}^{2} (r_{i,j} + s_{i,j})a_{i,j} + \sum_{j=1}^{2}\sum_{i=1}^{2} (r'_{i,j} + s'_{i,j})b_{i,j},$$

which proves the lemma.

LEMMA 2.2. *Let α be an algorithm for computing the product of two 2×2 matrices which has m multiplication steps. Then there exists an algorithm α' requiring only m steps such that the arguments for each multiplication are of the form $\sum_{j=1}^{2}\sum_{i=1}^{2} r_{i,j}a_{i,j} + \sum_{j=1}^{2}\sum_{i=1}^{2} r'_{i,j}b_{i,j}$.*

Proof. Let u be a polynomial in the $a_{i,j}$'s and $b_{i,j}$'s, i.e., $u \in Q[a_{11}, \ldots, b_{22}]$. Define the mappings L_i: $Q[a_{1,1}, \ldots, b_{2,2}] \to Q[a_{1,1}, \ldots, b_{2,2}]$. $i = 0, 1, 2, 3$ by $L_0(u)$ is the constant term, $L_1(u)$ is the linear lemma, $L_2(u)$ is the quadratic term, and $L_3(u) = u - L_0(u) - L_1(u) - L_2(u)$. Let $e_\alpha(j) = e_\alpha(j_1) \cdot e_\alpha(j_2) = u \cdot u'$. Then $e_\alpha(j) = L_0(u)L_0(u') + L_0(u)(u' - L_0(u')) + L_0(u')(u - L_0(u)) + (u - L_0(u))(u' - L_0(u'))$. $L_0(u)L_0(u')$ is a rational number, $L_0(u)(u' - L_0(u'))$ and $L_0(u')(u - L_0(u))$ are multiplications by a rational number (and, therefore, are not counted as multiplications), so we can obtain an algorithm α^* from α such that α^* has the same number of multiplications as α and, if $e_{\alpha^*}(j) = u \cdot u'$, then $L_0(u) = L_0(u') = 0$. (Note that, if u is computed, then $u - L_0(u)$ can be computed without extra multiplications).

If $L_0(u) = L_0(u') = 0$, then $L_2(u \cdot u') = L_1(u) \cdot L_1(u')$. By Lemma 2.1, if α^* is an algorithm which can compute the product of two 2×2 matrices, then, for each i,

$$j = \{1, 2\} \sum_{k=1}^{2} a_{ik}b_{kj} = r_0 + \sum_{i=1}^{m} r_i p_i + \sum_{j=1}^{2}\sum_{i=1}^{2} r_{i,j}a_{i,j} + \sum_{j=1}^{2}\sum_{i=1}^{2} r'_{i,j}b_{i,j}.$$

Applying L_2 to both sides, we obtain that

$$\sum_{k=1}^{2} a_{ik}b_{kj} = \sum_{i=1}^{m} r_i L_2(p_i).$$

Let $p_i = u_i \cdot u_i'$; then we obtain that

$$\sum_{k=1}^{2} a_{ik}b_{kj} = \sum_{i=1}^{m} r_i L_1(u_i) \cdot L_1(u_i').$$

We can now construct the algorithm α' required by Lemma 2.2 by first computing, for each i, $L_1(u_i)$ and $L_1(u_i')$; note the no multiplication which is counted is required. We then use m multiplications to form $L_1(u_i) \cdot$

$L_1(u_i')$, $i = 1, 2, \ldots, m$, and finally by using only additions and multiplication by a fixed scalar we obtain

$$\sum_{k=1}^{2} a_{ik}b_{kj} = \sum_{i=1}^{m} r_i L_1(u_i) \cdot L_2(u_i').$$

LEMMA 2.3. *Let* $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ *and* $(\alpha_1', \alpha_2', \alpha_3', \alpha_4')$ *be two linearly independent vectors in* Q^4. *Then any algorithm which computes both*

$$f_1 = a_{11}(\alpha_1 b_{11} + \alpha_2 b_{12}) + a_{12}(\alpha_1 b_{21} + \alpha_2 b_{22}) + a_{21}(\alpha_3 b_{11} + \alpha_4 b_{12}) + a_{22}(\alpha_3 b_{21} + \alpha_4 b_{22})$$

and

$$f_2 = a_{11}(\alpha_1' b_{11} + \alpha_2' b_{12}) + a_{12}(\alpha_1' b_{21} + \alpha_2' b_{22}) + a_{21}(\alpha_3' b_{11} + \alpha_4' b_{12}) + a_{22}(\alpha_3' b_{21} + \alpha_1' b_{22})$$

requires at least four multiplications.

Proof. The expressions f_1 and f_2 can also be written as

$$f_1 = b_{11}(\alpha_1 a_{11} + \alpha_3 a_{21}) + b_{12}(\alpha_2 a_{11} + \alpha_4 a_{21}) + b_{21}(\alpha_1 a_{12} + \alpha_3 a_{22}) + b_{22}(\alpha_2 a_{12} + \alpha_4 a_{22}),$$

$$f_2 = b_{11}(\alpha_1' a_{11} + \alpha_3' a_{21}) + b_{12}(\alpha_2' a_{11} + \alpha_4' a_{21}) + b_{21}(\alpha_1' a_{12} + \alpha_3' a_{22}) + b_{22}(\alpha_2' a_{12} + \alpha_4' a_{22}).$$

Assume that only three multiplications are necessary to compute f_1 and f_2. Then it follows from the theorem of [3] and [4] that as vectors over the rationals the vectors

$$\begin{bmatrix} \alpha_1 b_{11} + \alpha_2 b_{12} \\ \alpha_1' b_{11} + \alpha_2' b_{12} \end{bmatrix}, \quad \begin{bmatrix} \alpha_1 b_{21} + \alpha_2 b_{22} \\ \alpha_1' b_{21} + \alpha_2' b_{22} \end{bmatrix}, \quad \begin{bmatrix} \alpha_3 b_{11} + \alpha_4 b_{12} \\ \alpha_3' b_{11} + \alpha_4' b_{12} \end{bmatrix}, \quad \begin{bmatrix} \alpha_3 b_{21} + \alpha_4 b_{22} \\ \alpha_3' b_{21} + \alpha_4' b_{22} \end{bmatrix}$$

are linearly dependent, and also that the vectors

$$\begin{bmatrix} \alpha_1 a_{11} + \alpha_3 a_{21} \\ \alpha_1' a_{11} + \alpha_3' a_{21} \end{bmatrix}, \quad \begin{bmatrix} \alpha_2 a_{11} + \alpha_4 a_{21} \\ \alpha_2' a_{11} + \alpha_4' a_{21} \end{bmatrix}, \quad \begin{bmatrix} \alpha_1 a_{12} + \alpha_3 a_{22} \\ \alpha_1' a_{12} + \alpha_3' a_{22} \end{bmatrix}, \quad \begin{bmatrix} \alpha_2 a_{12} + \alpha_4 a_{22} \\ \alpha_2' a_{12} + \alpha_4' a_{22} \end{bmatrix}$$

are linearly dependent. Therefore there must exist four rational numbers u_1, u_2, v_1, v_2 such that

$$u_1 \begin{bmatrix} \alpha_1 \\ \alpha_1' \end{bmatrix} + u_2 \begin{bmatrix} \alpha_3 \\ \alpha_3' \end{bmatrix} = 0,$$

$$u_1 \begin{bmatrix} \alpha_2 \\ \alpha_2' \end{bmatrix} + u_2 \begin{bmatrix} \alpha_4 \\ \alpha_4' \end{bmatrix} = 0,$$

$$v_1\begin{bmatrix}\alpha_1\\ \alpha_1'\end{bmatrix}+v_2\begin{bmatrix}\alpha_2\\ \alpha_2'\end{bmatrix}=0,$$

$$v_1\begin{bmatrix}\alpha_3\\ \alpha_3'\end{bmatrix}+v_2\begin{bmatrix}\alpha_4\\ \alpha_4'\end{bmatrix}=0,$$

and therefore the space spanned by

$$\begin{bmatrix}\alpha_1\\ \alpha_1'\end{bmatrix},\qquad \begin{bmatrix}\alpha_2\\ \alpha_2'\end{bmatrix},\qquad \begin{bmatrix}\alpha_3\\ \alpha_3'\end{bmatrix},\qquad \begin{bmatrix}\alpha_4\\ \alpha_4'\end{bmatrix}$$

is one-dimensional, contradicting the assumption that $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ and $(\alpha_1', \alpha_2', \alpha_3', \alpha_4')$ are linearly independent.

3. THE RESULT

THEOREM 3.1. *Every algorithm for multiplying two* 2×2 *matrices requires at least seven multiplications.*

Proof. Assume there exists an algorithm using only six multiplications. Using Lemma 2.2, there exist six products $p_1, p_2, \ldots, p_6$ and twenty-four rational numbers $r_{i,j}^k$, $i, j \in \{1, 2\}$, $k \in \{1, 2, \ldots, 6\}$ such that

$$\sum_{k=1}^{2} a_{ik}b_{kj} = \sum_{k=1}^{6} r_{i,j}^k p_k, \qquad i, j \in \{1, 2\}.$$

Let R be the 4×6 matrix

$$R = \begin{bmatrix} r_{1,1}^1 & r_{1,1}^2 & \cdots & r_{1,1}^6 \\ r_{1,2}^1 & \cdots & \cdot & r_{1,2}^6 \\ r_{2,1}^1 & \cdot & \cdot & r_{2,1}^6 \\ r_{2,2}^1 & \cdots & \cdots & r_{2,2}^6 \end{bmatrix},$$

$\mathbf{p}$ the (column) vector $(p_1, p_2, \ldots, p_6)$, and $\mathbf{c}$ the (column) vector $(c_{11}, c_{12}, c_{21}, c_{22})$, where $c_{i,j} = \sum_{k=1}^{2} a_{ik}b_{kj}$; then $R\mathbf{p} = \mathbf{c}$. The rank of R has to be four since no nontrivial linear combination of the $c_{i,j}$'s vanishes; therefore there exists a 4×4 matrix D such that (possibly after a permutation of the columns of R and renaming the p_i's) the matrix $R' = D \cdot R$ is of the form

$$\begin{bmatrix} & r_1 & s_1 \\ I & r_2 & s_2 \\ & r_3 & s_3 \\ & r_4 & s_4 \end{bmatrix}.$$

Let $D\mathbf{c}$ be the vector $\mathbf{c}' = (c'_{11}, c'_{12}, c'_{13}, c'_{14})$, where then each $c'_{i,j}$ is of the form $d_{m,1}c_{11} + d_{m,2}c_{12} + d_{m,3}c_{2,1} + d_{m,4}c_{2,2}$ for some $m \in \{1, 2, 3, 4\}$. Each $c'_{i,j}$ is of the form

$$a_{11}(d_{m,1}b_{1,1} + d_{m,2}b_{1,2}) + a_{12}(d_{m,1}b_{2,1} + d_{m,2}b_{2,2}) + a_{21}(d_{m,3}b_{1,1} + d_{m,4}b_{1,2}) + a_{22}(d_{m,3}b_{2,1} + d_{m,4}b_{2,2}).$$

Since each $c'_{i,j}$ can be computed using only three multiplications, it follows that $d_{m,1}b_{11} + d_{m,2}b_{1,2}$, $d_{m,1}b_{2,1} + d_{m,2}b_{22}$, $d_{m,3}b_{1,1} + d_{m,4}b_{1,2}$, $d_{m,3}b_{2,1} + d_{m,4}b_{2,2}$ are linearly dependent and therefore that the determinant

$$\begin{vmatrix} d_{m,1} & d_{m,2} \\ d_{m,3} & d_{m,4} \end{vmatrix}$$

vanishes.

Using Lemma 2.3, we obtain that at most one of the r_i's is 0 and at most one of the s_i's is 0. We assume with no loss of generality that $r_1 \neq 0_2$, $r_2 \neq 0$, and $r_3 \neq 0$. Since D is nonsingular, the rank of either

$$\begin{bmatrix} d_{1,1} & d_{1,2} \\ d_{2,1} & d_{2,2} \\ d_{3,1} & d_{3,2} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} d_{1,3} & d_{1,4} \\ d_{2,3} & d_{2,4} \\ d_{3,3} & d_{3,4} \end{bmatrix}$$

is two, and we will assume that $(d_{1,1}, d_{1,2})$ and $(d_{2,1}, d_{2,2})$ are linearly independent. We may also assume that $(d_{3,1}, d_{3,2})$ is not 0, since if it is we replace D by

$$D' = \begin{bmatrix} \frac{1}{s_1} & 0 & 0 & 0 \\ -\frac{s_2}{s_1} & 1 & 0 & 0 \\ -\frac{s_3}{s_1} & 0 & 1 & 0 \\ -\frac{s_4}{s_1} & 0 & 0 & 1 \end{bmatrix} \times D$$

and then $D'R$ is of the same form as R' (after exchanging the first and fifth columns, and p_1 and p_5). Since $(d_{1,1}, d_{1,2})$ and $(d_{1,2}, d_{2,2})$ are linearly independent, and $(d_{3,1}, d_{3,2})$ is not 0, then either $(d_{1,1}, d_{1,2})$ and $(d_{3,1}, d_{3,2})$ are linearly independent or $(d_{2,1}, d_{2,2})$ and $(d_{3,1}, d_{3,2})$ are linearly independent. Assume with no loss of generality that $(d_{1,1}, d_{1,2})$ and $(d_{3,1}, d_{3,2})$ are linearly independent; so $(d_{m,3}, d_{m,4}) = e_m(d_{m,1}, d_{m,2})$ for $m = 1, 2, 3$.

Since each of the expressions $c_2' - (s_2/s_1)c_1'$ and $c_3' - (s_2/s_1)c_1'$ can be calculated using only three products, it follows that

$$\begin{vmatrix} d_{2,1} - \frac{s_2}{s_1} d_{1,1} & d_{2,2} - \frac{s_2}{s_1} d_{1,2} \\ d_{2,3} - \frac{s_2}{s_1} d_{1,3} & d_{2,4} - \frac{s_2}{s_1} d_{1,4} \end{vmatrix} = 0, \tag{1}$$

$$\begin{vmatrix} d_{3,1} - \frac{s_3}{s_1} d_{1,1} & d_{3,2} - \frac{s_3}{s_1} d_{1,2} \\ d_{3,3} - \frac{s_3}{s_1} d_{1,3} & d_{3,4} - \frac{s_3}{s_1} d_{1,4} \end{vmatrix} = 0. \tag{2}$$

Using the relations developed above, we obtain that (1) holds if and only if $e_1 = e_2$, and (2) holds if and only if $e_1 = e_3$. Let α, β be such that $(d_{3,1}, d_{3,2}) = \alpha(d_{1,1}, d_{1,2}) + \beta(d_{2,1}, d_{2,2})$; since $e_1 = e_2 = e_3$ we obtain that $(d_{3,3}, d_{3,4}) = \alpha(d_{1,3}, d_{1,4}) + \beta(d_{2,3}, d_{2,4})$, contradicting the assumption that D is nonsingular. We have thus obtained a contradiction from the assumption that two 2×2 matrices can be multiplied using only six multiplications, which proves the theorem.

4. PRODUCTS OF COMPLEX NUMBERS

The method of proving the main result of the paper can be also used to show that at least three multiplications or divisions of real numbers are necessary to multiply two complex numbers given in Cartesian form. That is, to form $(a + ib)(c + id) = (ac - bd) + i(ad + bc)$ requires at least three multiplications or divisions of real numbers.

Before proving this theorem, we note that it is possible to compute a complex product using only three multiplications. For example,

$$ac - bd = ac - bd,$$

$$ad + bc = (a + b)(c + d) - ac - bd.$$

So the three products which are formed are ac, bd, $(a + b)(c + d)$.

THEOREM 4.1. *Every algorithm for computing $ac - bd$ and $ad + bc$ requires at least three multiplications or divisions.*

Proof. Assume there exists an algorithm requiring only two multiplications or divisions. Let p_1, p_2 be the result of these multiplications or divisions. Assume that p_1 was formed before p_2, so, though p_2 may depend

on p_1, p_1 is just a product or quotient of linear terms. In a way similar to the proof of Lemma 2.1, it can be shown that

$$ac - bd = r_1p_1 + r_2p_2 + r_3a + r_4b + r_5c + r_6d + r_7,$$

$$ad + bc = s_1p_1 + s_2p_2 + s_3a + s_4b + s_5c + s_6d + s_7$$

for some rational numbers r_i, s_j, $i, j = 1, 2, \ldots, 7$. Since at least two multiplications or divisions are required to compute $ac - bd$, it follows that $r_2 \neq 0$. Similarly we obtain $s_2 \neq 0$. Therefore

$$\begin{aligned} a\left(c - \frac{r_2}{s_2}d\right) - b\left(d + \frac{r_2}{s_2}c\right) = &\left(r_1 - \frac{r_2}{s_2}s_1\right)p_1 + \left(r_3 - \frac{r_2}{s_2}s_3\right)a \\ &+ \left(r_4 - \frac{r_2}{s_2}s_4\right)b + \left(r_5 - \frac{r_2}{s_2}s_5\right)c \\ &+ \left(r_6 - \frac{r_2}{s_2}s_6\right)d + \left(r_7 - \frac{r_2}{s_2}s_7\right). \end{aligned}$$

Consequently the expression $a(c - (r_2/s_2)d) - b(d + (r_2/s_2)c)$ can be computed using only one multiplication or division. This implies that $c - (r_2/s_2)d$ and $d + (r_2/s_2)c$ are linearly dependent. Therefore

$$0 = \begin{vmatrix} 1 & -\dfrac{r_2}{s_2} \\ \dfrac{r_2}{s_2} & 1 \end{vmatrix} = 1 + \frac{r_2^2}{s_2^2},$$

which is a contradiction. Therefore at least three multiplications or divisions are necessary.

REFERENCES

1 J. E. Hopcroft and L. R. Kerr, Some techniques for proving certain simple programs optimal, *Proc. Tenth Ann. Symposium on Switching and Automata Theory*, 1969, pp. 36–45.

2 V. Strassen, Gaussian Elimination is not optimal, *Numer. Math.* (4) **13**(August 1969), 354–356.

3 S. Winograd, On the number of multiplications required to compute certain functions, *Proc. Nat. Acad. Sci. USA*, (5) **58**(November 1967), 1840–1842.

4 S. Winograd, On the number of multiplications necessary to compute certain functions, IBM RC 2285, November 1968.

Received February, 1970

Linear Algebra and Its Applications **4**(1971), 381–388