On the Complexity of Matrix Product

Ran Raz*
Weizmann Institute
ranraz@wisdom.weizmann.ac.il

Abstract

Our main result is a lower bound of $\Omega(m^2 \log m)$ for the size of any arithmetic circuit for the product of two matrices, over the real or complex numbers, as long as the circuit doesn't use products with field elements of absolute value larger than 1 (where $m \times m$ is the size of each matrix). That is, our lower bound is super-linear in the number of inputs and is applied for circuits that use addition gates, product gates and products with field elements of absolute value up to 1.

We also prove size-depth tradeoffs for such circuits: We show that if a circuit, as above, is of depth d then its size is $\Omega(m^{2+1/O(d)})$.

1 Introduction

Matrix product is among the most extensively studied computational problems. Surprising upper bounds of $O(m^{2+\alpha})$ (where $\alpha < 1$, and $m \times m$ is the size of each matrix) were obtained by Strassen in [Str] and improved in many other works. The best current upper bound (obtained by Coppersmith and Winograd) achieves $\alpha \approx 0.376$ [CW] (see [Gat] for a survey). The best lower bounds, however, are linear lower bounds of between $2.5 \cdot m^2$ and $3 \cdot m^2$ (depending on the field) for the number of products needed [Bsh, Bla, Shp].

In particular, the following seminal problem is still open: Can matrix product be computed by circuits of size $O(m^2)$, that is, circuits of size linear in the number of inputs ? Super-linear lower bounds for matrix product are only known for bounded depth circuits [RS]. Note, however, that Strassen's method, as well as many other methods for matrix product, use circuits of larger depth.

The standard computational model for matrix product is by arithmetic circuits over some field F. The inputs for the circuit are the entries of the two matrices, and the allowed gates are product and addition over F. Products with field elements are also allowed. In this work, we take F to be the field of real numbers (all of our results hold for the complex numbers as well), and we restrict our arithmetic circuit in the following way: The circuit cannot use

*Research supported by US-Israel BSF grant 98-00349.

products with field elements of absolute value larger than 1. We call such a circuit *a bounded coefficients arithmetic circuit.*

We prove that any such circuit for matrix product is of size $\Omega(m^2 \log m)$.

More generally, if we require that the circuit doesn't use products with field elements of absolute value larger than $c = c(m)$ (for any function $c(m) \geq 1$), we obtain that any such circuit for matrix product is of size $\Omega(m^2 \log_{2c} m)$. This follows from the case $c = c(m) = 1$ by a simple reduction (just by replacing each product with a field element of absolute value smaller or equal to c by up to $\log_2 c$ additions and one product with a field element of absolute value smaller or equal to 1). Hence, in the rest of the paper, we concentrate on the case $c = 1$.

Besides our main result, we also prove size-depth tradeoffs for bounded coefficients arithmetic circuits for matrix product. We show that any such circuit of depth d is of size $\Omega(m^{2+1/O(d)})$. Note that for general arithmetic circuits only much weaker size-depth tradeoffs are known [RS].

Since a preliminary version of our paper has appeared, Burgisser and Lotz used our methods (together with some new methods) to obtain a similar lower bound of $\Omega(n \log n)$ for any bounded coefficients arithmetic circuit for the product of two polynomials of degree n (as well as for the division of two polynomials with reminder) [BL].

1.1 Previous work

Bounded coefficients arithmetic circuits were suggested and motivated as a natural model for arithmetic computations by Morgenstern [Mor] (1973) and by Chazelle [Cha] (1994). Morgenstern and Chazelle observed that many algorithms for arithmetic problems (e.g., the Fast Fourier Transform algorithm) do not use field elements at all (or use only small field elements). Morgenstern and Chazelle were mainly interested in the case of linear functions and proved lower bounds of $\Omega(n \log n)$ for several such functions (e.g., for the Fourier transform). Note that for general arithmetic circuits no super-linear lower bound is known for any linear function (or any constant degree polynomial).

Several works proved size-depth tradeoffs of $\Omega(n^{1+1/O(d)})$ for bounded coefficients arithmetic circuits [NW, Lok, Pud]. As in [Mor, Cha], the focus of these works was linear functions. As far as we know, no previous result was obtained for the complexity of matrix product (or similar functions) in the bounded coefficients model.

1.2 Organization of the paper

The paper is organized as follows. In Section 2, we give some basic definitions. In Section 3, we give lower bounds for linear functions. These lower bounds are then used in Section 4 to prove our lower bound for matrix product. The proof of the main lemma is deferred to Section 5. In Section 6, we prove our size-depth tradeoff for matrix product.

2 Preliminaries

As mentioned above, we consider arithmetic circuits over the field of real numbers. All of our results hold for the complex numbers as well.

An arithmetic circuit is a directed acyclic graph as follows: Nodes of in-degree 0 are called inputs and are labelled with input variables. Nodes of out-degree 0 are called outputs. Each edge is labelled with a field element (we think of this element as multiplying the outcome of the edge). Each node other than an input is labelled with either $+$ or $\times$ (in the first case the node is a plus gate and in the second case a product gate).

The computation is done in the following way. An input just computes the value of the variable that labels it. For every non-input node v, if $v_1, ..., v_k$ are the nodes that fan into v then we multiply the result of each v_i with the field element that labels the edge that connects it to v. If v is a plus gate we sum all the results, otherwise v is a product gate and we multiply all the results. Obviously, each node in the circuit computes a polynomial in the input variables.

The *size* of a circuit C is defined to be the number of edges in it and is denoted by $\text{Size}(C)$. The *depth* of a circuit C is defined to be the length of the longest directed path between an input and an output in C and is denoted by $\text{Depth}(C)$.

We say that an arithmetic circuit (over the real numbers) is a *bounded coefficients arithmetic circuit* if all field elements labelling the edges of the circuit are of absolute value smaller or equal to 1.

We say that an arithmetic circuit is *linear* if all gates in it are plus gates (i.e., the circuit contains no product gates). Obviously, the outputs of a linear circuit are linear functions in the input variables. Let $L_1, ..., L_k$ be k linear functions (in the variables $z_1, ..., z_n$). It is well known (and easy to prove) that (over any field with characteristic 0) any arithmetic circuit for $L_1, ..., L_k$ can be translated into a linear circuit for $L_1, ..., L_k$, with only a constant-factor increase in the size and depth of the circuit. In the same way, any bounded coefficients arithmetic circuit for $L_1, ..., L_k$ can be translated into a bounded coefficients linear circuit for $L_1, ..., L_k$, with only a constant-factor increase in the size and depth of the circuit. We can hence assume w.l.o.g. that linear forms $L_1, ..., L_k$ are computed by linear circuits. Given an $n \times n$ matrix H, we say that a linear circuit computes H if it computes the linear functions that correspond to the rows of H, that is, the circuit computes the functions $\sum_{j=1}^{n} H_{i,j} \cdot z_j$, where $z_1, ..., z_n$ are the input variables for the circuit.

In this paper, we prove lower bounds on the size of circuits for the product of two $m \times m$ matrices. The input for such a circuit is of size $2m^2$, and it consists of two $m \times m$ matrices X, Y. The output is the matrix $X \cdot Y$. That is, there are m^2 outputs, and the $(i, j)^{th}$ output is: $\sum_{k=1}^{m} X_{i,k} \cdot Y_{k,j}$. Each output is a bilinear form in X and Y.

Since the product of two matrices is a bilinear form, it is natural to consider bilinear arithmetic circuits for it. We say that an arithmetic circuit is *bilinear* if each product gate in it computes the product of two linear functions, one in the variables $\{X_{i,j}\}$ and the other in the variables $\{Y_{i,j}\}$. Thus, a bilinear circuit has the following structure. First, there are many plus gates, computing linear forms in X and linear forms in Y. Then, there is one level of product gates that compute bilinear forms. Finally, there are many plus gates that eventually compute the outputs.

Obviously, the outputs of a bilinear circuit are bilinear functions in the input variables of X and Y. Let $f_1, ..., f_k$ be k bilinear functions (in the variables of X and Y). It is well known (and easy to prove) that (over any field with characteristic 0) any arithmetic circuit for $f_1, ..., f_k$ can be translated into a bilinear circuit for $f_1, ..., f_k$, with only a constant-factor increase in the size and depth of the circuit. In the same way, any bounded coefficients arithmetic circuit for $f_1, ..., f_k$ can be translated into a bounded coefficients bilinear circuit for $f_1, ..., f_k$, with only a constant-factor increase in the size and depth of the circuit. We can hence assume w.l.o.g. that bilinear forms $f_1, ..., f_k$ are computed by bilinear circuits.

3 Lower Bounds for Linear Functions

In this section, we prove lower bounds for the size of bounded coefficients linear circuits. In all that comes below, we assume w.l.o.g.[1] that all gates in the circuit are of fan-in 2.

Lower bounds for the size of bounded coefficients linear circuits were first proved by Morgenstern [Mor]. Morgenstern observed that for any matrix H, the size of any bounded coefficients linear circuit for H is bounded from below by $\log_2 |\mathrm{Det}[H]|$. For our purpose, we will need the following simple generalization of this result (Lemma 3.1).

Let $L_1, ..., L_k$ be k linear functions in the variables $z_1, ..., z_n$. We think of each L_i as a vector in the vector space R^n. For every $1 \le r \le n$, denote by $\mathrm{Vol}_r[L_1, ..., L_k]$ the maximal volume spanned by r vectors from $\{L_1, ..., L_k\}$ and $n - r$ arbitrary unit vectors (i.e., vectors with L^2-norm equal to 1). That is,

$\mathbf{Vol_r[L_1, ..., L_k]} = \mathrm{MAX}_{i_1,...,i_r,e_{r+1},...,e_n} |\mathrm{Det}[L_{i_1}, ..., L_{i_r}, e_{r+1}, ..., e_n]|$,
where $e_{r+1}, ..., e_n$ are arbitrary unit vectors in R^n. Equivalently,
$\mathbf{Vol_r[L_1, ..., L_k]} = \mathrm{MAX}_{i_1,...,i_r} (\mathrm{Det}[AA^T])^{1/2}$,
where A is the $n \times r$ matrix $[L_{i_1}, ..., L_{i_r}]$, and A^T is the transpose of A.

In the same way, for a matrix H of size $n \times n$, we define for every $1 \le r \le n$,

$\mathbf{Vol_r[H]} = \mathrm{Vol}_r[L_1, ..., L_n]$,
where $L_1, ..., L_n$ are the linear forms corresponding to the rows of H.

Lemma 3.1 *Let C be a bounded coefficients linear circuit for $L_1, ..., L_k$. Then, for every $1 \le r \le n$,*

$$Size(C) \ge \log_2(Vol_r[L_1, ..., L_k]).$$

Proof:
Let $s = \mathrm{Size}(C)$. Note that since C is a directed acyclic graph, it induces a partial order on

[1]Recall that the size of a circuit is defined as the number of edges in it, and not the number of nodes.

its nodes (a node v is larger than a node u if there exists a directed path from u to v). Let $f_1, ..., f_s$ be the linear functions corresponding to all nodes in C, and such that the order of $f_1, ..., f_s$ agrees with the partial order induced by the circuit C and $f_1 = z_1, ..., f_n = z_n$ (where $z_1, ..., z_n$ are the input variables for the circuit).

Since the order of $f_1, ..., f_s$ agrees with the order of the circuit, for every $i > n$ there exist $i_1, i_2 < i$ and c_1, c_2 of absolute value ≤ 1, such that, $f_i = c_1 \cdot f_{i_1} + c_2 \cdot f_{i_2}$. Hence, by the linear property of the determinant, it is easy to verify that

$$\mathrm{Vol}_r[f_1, ..., f_i] \leq 2 \cdot \mathrm{Vol}_r[f_1, ..., f_{i-1}],$$

and since $\mathrm{Vol}_r[f_1, ..., f_n] = 1$, we have

$$\mathrm{Vol}_r[f_1, ..., f_s] \leq 2^{s-n} < 2^s.$$

Since $\{f_1, ..., f_s\}$ include the functions $L_1, ..., L_k$, we have

$$\mathrm{Vol}_r[L_1, ..., L_k] \leq \mathrm{Vol}_r[f_1, ..., f_s] < 2^s.$$

$\square$

For a linear function L in n variables and for a vector space $V \subset \mathrm{R}^n$, denote by $\mathbf{Dist[L, V]}$ the L^2-distance between L and V (as before, we think of L as a vector in R^n). For r linear functions, $L_1, ..., L_r$, denote by $\mathbf{Span[L_1, ..., L_r]}$ the vector space in R^n spanned by $L_1, ..., L_r$. Let $L_1, ..., L_k$ be k linear functions in n variables. For every $1 \leq r \leq n$, denote

$\mathbf{Rig_r[L_1, ..., L_k]} = \mathrm{MIN}_V \mathrm{MAX}_i(\mathrm{Dist}[L_i, V])$,
where $V \subset \mathrm{R}^n$ is a vector space of dimension r.

In the same way, for a matrix H of size $n \times n$, we define for every $1 \leq r \leq n$,

$\mathbf{Rig_r[H]} = \mathrm{Rig}_r[L_1, ..., L_n]$,
where $L_1, ..., L_n$ are the linear forms corresponding to the rows of H.

A notion similar (but not identical) to $\mathrm{Rig}_r[H]$ was defined in [Lok] and was used there to prove size-depth tradeoffs for bounded coefficients arithmetic circuits. Here, we connect $\mathrm{Rig}_r[L_1, ..., L_k]$ to $\mathrm{Vol}_r[L_1, ..., L_k]$, and hence to the size of the smallest bounded coefficients arithmetic circuit for $L_1, ..., L_k$.

Lemma 3.2 *For every k linear functions $L_1, ..., L_k$, and every $1 \leq r \leq n$,*

$$\log_2(\mathit{Vol}_r[L_1, ..., L_k]) \geq r \cdot \log_2(\mathit{Rig}_r[L_1, ..., L_k]).$$

Proof:
Assume w.l.o.g. that $r < k$ (otherwise, $\mathrm{Rig}_r[L_1, ..., L_k] = 0$). Assume w.l.o.g. that the order of $L_1, ..., L_k$ is as follows: L_1 is a function $L \in \{L_1, ..., L_k\}$ such that $\mathrm{Vol}_1[L]$ is maximal. L_2 is a

function $L \in \{L_2, ..., L_k\}$ such that $\text{Vol}_2[L_1, L]$ is maximal, and so on (i.e., for every $1 \leq i \leq k$, we have that L_i is a function $L \in \{L_i, ..., L_k\}$ such that $\text{Vol}_i[L_1, ..., L_{i-1}, L]$ is maximal).

Denote $\upsilon_1 = \text{Vol}_1[L_1]$, and for every $1 < i \leq k$ denote $\upsilon_i = \text{Vol}_i[L_1, ..., L_i]/\text{Vol}_{i-1}[L_1, ..., L_{i-1}]$. Then, by our assumption on the order of $L_1, ..., L_k$, it is easy to verify that

$$\upsilon_1 \geq \upsilon_2 \geq \cdots \geq \upsilon_k.$$

Therefore,

$$\text{Vol}_r[L_1, ..., L_r] = \prod_{i=1}^{r} \upsilon_i \geq (\upsilon_{r+1})^r.$$

On the other hand (again by our assumption on the order of $L_1, ..., L_k$),

$$\upsilon_{r+1} = \text{MAX}_i(\text{Dist}[L_i, V]),$$

where $V = \text{Span}[L_1, ..., L_r]$, and hence,

$$\text{Rig}_r[L_1, ..., L_k] \leq \upsilon_{r+1}.$$

Thus,

$$\text{Vol}_r[L_1, ..., L_k] \geq \text{Vol}_r[L_1, ..., L_r] \geq (\upsilon_{r+1})^r \geq (\text{Rig}_r[L_1, ..., L_k])^r.$$

$\square$

Given a matrix H of size $m \times m$, we can use Lemma 3.1 and Lemma 3.2 to prove lower bounds for bounded coefficients arithmetic circuits for H. For our purpose, we will also need to prove lower bounds for bounded coefficients arithmetic circuits for the tensor product $I \otimes H$ (where I is the identity matrix of size $m \times m$). Recall that $I \otimes H$ is a matrix of size $m^2 \times m^2$ that consists of $m \times m$ blocks of size $m \times m$ each, such that the m blocks on the diagonal contain copies of the matrix H and all other blocks contain the zero matrix (of size $m \times m$). We will use the following proposition.

Proposition 3.3 *Let H be an arbitrary matrix of size $m \times m$ and let I be the identity matrix of size $m \times m$. Then, for every $1 \leq r \leq m$,*

$$\log_2(\mathit{Vol}_{r \cdot m}[I \otimes H]) \geq m \cdot \log_2(\mathit{Vol}_r[H]).$$

Proof:
By the properties of the determinant, for every matrix A (of size $m \times m$),

$$\text{Det}[I \otimes A] = (\text{Det}[A])^m.$$

Hence, by the definition of Vol,

$$\text{Vol}_{r \cdot m}[I \otimes H] \geq (\text{Vol}_r[H])^m.$$

$\square$

Corollary 3.4 *Let H be an arbitrary matrix of size $m \times m$ and let I be the identity matrix of size $m \times m$. Let C be a bounded coefficients linear circuit for $I \otimes H$. Then, for every $1 \le r \le m$,*

$$\mathit{Size}(C) \ge r \cdot m \cdot \log_2(\mathit{Rig}_r[H]).$$

Proof:
By Lemma 3.1, Proposition 3.3, and Lemma 3.2,

$$\text{Size}(C) \ge \log_2(\text{Vol}_{r \cdot m}[I \otimes H]) \ge m \cdot \log_2(\text{Vol}_r[H]) \ge m \cdot r \cdot \log_2(\text{Rig}_r[H]).$$

$\square$

4 Lower Bounds for Matrix Product

In this section, we prove a lower bound for the size of bounded coefficients arithmetic circuits for matrix product. Our bound is based on the lower bounds given in the previous section and on the following Lemma 4.1. The proof of Lemma 4.1 is given in the next section. In the following lemma, we assume for simplicity that $m/10$ is integer.

Lemma 4.1 (main lemma) *Let $L_1, ..., L_k$ be k linear functions (over R) in the m^2 variables $y_{1,1}, ..., y_{m,m}$ (we think of $y_{1,1}, ..., y_{m,m}$ as the entries of a matrix of size $m \times m$). Denote, $r = m/10$, and assume (for simplicity) that m is large enough (i.e., $m > m_0$, for some global constant m_0). Then, there exists a matrix Y of size $m \times m$ (over R), such that:*

1. *For every $1 \le i \le k$,*

$$|L_i(Y_{1,1}, ..., Y_{m,m})| \le \mathit{Rig}_{r \cdot m}[L_1, ..., L_k] \cdot (2 \ln k + 10)^{1/2}.$$

2.

$$\mathit{Rig}_r[Y] \ge \sqrt{m/9}.$$

We will now state and prove our main result.

Theorem 1 *Let C be a bounded coefficients arithmetic circuit (over the real or complex numbers) for the product of two matrices of size $m \times m$. Then,*

$$\mathit{Size}(C) = \Omega(m^2 \log m).$$

Proof:
First note that w.l.o.g. we can assume that the circuit is over the real numbers. This is true because any circuit over the complex numbers can be translated into a circuit over the real numbers (and vice-versa) with a constant-factor increase in its size. As before, we assume w.l.o.g. that all gates in the circuit are of fan-in 2. Recall also that we can assume w.l.o.g.

that the circuit is bilinear. We assume w.l.o.g. that m is large enough (and in particular, $m > m_0$, where m_0 is the global constant from Lemma 4.1), and we assume for simplicity that $m/10$ is integer. Define,

$$r = m/10.$$

Assume, for a contradiction to the statement of the theorem, that

$$\text{Size}(C) < 0.001 \cdot m^2 \log_2 m.$$

Denote by $v_1, ..., v_k$ the product gates of the circuit C. Since the circuit is bilinear, each product gate v_i computes the product of two linear functions, one in the variables $\{x_{i,j}\}$ (of the first matrix) and the other in the variables $\{y_{i,j}\}$ (of the second matrix). Denote the first linear function by R_i and the second linear function by L_i. Thus, v_i computes the product of $R_i(x_{1,1}, ..., x_{m,m})$ and $L_i(y_{1,1}, ..., y_{m,m})$.

Consider the linear functions $L_1, ..., L_k$. These functions are computed by a linear circuit of size smaller than $0.001 \cdot m^2 \log_2 m$ (in the input variables $y_{1,1}, ..., y_{m,m}$). Hence, by Lemma 3.1 and Lemma 3.2,

$$r \cdot m \cdot \log_2(\text{Rig}_{r \cdot m}[L_1, ..., L_k]) < 0.001 \cdot m^2 \log_2 m.$$

That is,

$$\text{Rig}_{r \cdot m}[L_1, ..., L_k] < m^{1/100}.$$

Hence, by Lemma 4.1, there exists a matrix Y of size $m \times m$ (over R), such that:

1. For every $1 \leq i \leq k$,

   $$|L_i(Y_{1,1}, ..., Y_{m,m})| \leq m^{1/100} \cdot (2 \ln k + 10)^{1/2} < m^{1/99}$$

   (for large enough m).

2. $$\text{Rig}_r[Y] \geq \sqrt{m/9}.$$

We fix the input variables $y_{1,1}, ..., y_{m,m}$ to be the entries $Y_{1,1}, ..., Y_{m,m}$. Denote the obtained circuit by C'. Since we fixed $y_{1,1}, ..., y_{m,m}$, each product gate v_i in C turned into a product with the field element $L_i(Y)$. The circuit C' is hence a linear arithmetic circuit, and it is not hard to see that it computes the matrix $I \otimes Y$ in the input variables $x_{1,1}, ..., x_{m,m}$.

The circuit C' is not a bounded coefficients arithmetic circuit, because the absolute value of each field element $L_i(Y)$ is not bounded by 1. We would like to convert C' into a bounded coefficients arithmetic circuit C''. This could be done by replacing the product with each field element $L_i(Y)$ by up to $\log_2 |L_i(Y)|$ additions plus one product with a field element of absolute value ≤ 1. Note, however, that k may be almost as large as the size of C' and hence this method may increase the size of C' by more than a constant factor. Instead, we will convert C' into C'' by the following two steps: First, replace the product with each field element $L_i(Y)$ by a product with the field element $L_i(Y)/m^{1/99}$ (which is of absolute value

≤ 1) and multiply each output of the circuit by the field element $m^{1/99}$. (Since the original circuit C was bilinear, it is not hard to see that this step doesn't change the outputs of the circuit and the circuit still computes $I \otimes Y$). Then, replace each product (of an output) with the field element $m^{1/99}$ by up to $\log_2(m^{1/99})$ additions plus one product with a field element of absolute value ≤ 1. Since the number of outputs is m^2, this increases the size of the circuit by at most $(1/99) \cdot m^2 \log_2 m$.

Thus, the obtained circuit C'' is a bounded coefficients arithmetic circuit that computes the matrix $I \otimes Y$, and such that

$$\mathrm{Size}(C'') < (1/90) \cdot m^2 \log_2 m.$$

However, by Corollary 3.4,

$$\mathrm{Size}(C'') \geq r \cdot m \cdot \log_2(\mathrm{Rig}_r[Y]) \geq (1/20) \cdot m^2 \log_2(m/9),$$

which is a contradiction (for large enough m). □

5 Proof of the Main Lemma

In this section, we give a proof of Lemma 4.1.

We think of each linear function L in the variables $y_{1,1}, ..., y_{m,m}$ also as a vector in $\mathrm{R}^{m\times m}$, and we think of each vector in $\mathrm{R}^{m\times m}$ also as a linear function in the variables $y_{1,1}, ..., y_{m,m}$. Assignments to the variables $y_{1,1}, ..., y_{m,m}$ are matrices Z of size $m \times m$. We think of each such matrix also as a vector in $\mathrm{R}^{m\times m}$, and we think of each vector in $\mathrm{R}^{m\times m}$ also as a matrix of size $m \times m$. Given a linear function L in the variables $y_{1,1}, ..., y_{m,m}$ and an assignment Z to $y_{1,1}, ..., y_{m,m}$, the value $L(Z)$ is the value of the function L on the assignment Z. The norm that we use in $\mathrm{R}^{m\times m}$ is the L^2-norm. This norm is used to measure distances between vectors and lengths of vectors in $\mathrm{R}^{m\times m}$. Hence, it is also used to measure distances between matrices and norms of matrices, and distances between linear functions and norms of linear functions. We denote the L^2-norm of a linear function L by $\|L\|$, and we denote the L^2-norm of a matrix Z by $\|Z\|$ (this is known as the Frobenius norm of the matrix Z).

Denote,

$$R = \mathrm{Rig}_{r\cdot m}[L_1, ..., L_k].$$

By the definition of Rig, there exists a vector space $V \subset \mathrm{R}^{m\times m}$ of dimension $r \cdot m$, such that for every $1 \leq i \leq k$,

$$\mathrm{Dist}[L_i, V] \leq R.$$

Denote by $V^\perp \subset \mathrm{R}^{m\times m}$ the vector space orthogonal to V. Note that $V^\perp$ is a vector space of dimension $(m - r) \cdot m$. For every $1 \leq i \leq k$, we can write L_i as

$$L_i = L''_i + L'_i,$$

where $L''_i \in V$ and $L'_i \in V^{\perp}$. Since $\|L'_i\| = \mathrm{Dist}[L_i, V]$, we have for every $1 \leq i \leq k$,

$$\|L'_i\| \leq R.$$

Recall that we can think of V and $V^{\perp}$ also as subspaces of matrices Z of size $m \times m$. Obviously, $V^{\perp}$ is the vector space of all matrices $Z \in \mathrm{R}^{m \times m}$, such that every $L \in V$ satisfies $L(Z) = 0$. In the same way, V is the vector space of all matrices $Z \in \mathrm{R}^{m \times m}$, such that every $L \in V^{\perp}$ satisfies $L(Z) = 0$.

Our construction for the matrix Y will be probabilistic. We will define a random matrix Y that will satisfy the requirements of the lemma with high probability. The definition of Y will be in two stages. First, define the matrix W in the following way. Each entry $W_{i,j}$ is defined to be an independently chosen Gaussian random variable with expectation 0 and variance 1 (i.e., $W_{i,j}$ is chosen independently according to the distribution $N(0,1)$). Thus, the entries of the matrix W form a multi-normal distribution. We can write the matrix W as

$$W = W'' + W',$$

where $W'' \in V$ and $W' \in V^{\perp}$. We define,

$$Y = W'$$

(i.e., Y is the projection of W on $V^{\perp}$). We will show that with high probability Y satisfies the requirements of the lemma.

Claim 5.1 *With high probability (say, with probability of at least* 0.98*), for every* $1 \leq i \leq k$*,*

$$|L_i(Y)| \leq R \cdot (2 \ln k + 10)^{1/2}.$$

Proof:
Note that $L''_i(Y) = L''_i(W') = 0$ and that $L'_i(W'') = 0$. Hence, for every $1 \leq i \leq k$,

$$L_i(Y) = L'_i(Y) = L'_i(W).$$

Each $L'_i(W)$ is a weighted sum of independently chosen Gaussian random variables with expectation 0 and variance 1, and hence $L'_i(W)$ is a Gaussian random variable with expectation 0 and variance $\|L'_i\|^2$.

Since $\|L'_i\| \leq R$, the probability of the event $|L'_i(W)| > R \cdot (2 \ln k + 10)^{1/2}$ can be bounded by $2/(e^5 \cdot k)$ (see for example [ASE], Appendix A). Hence, by the union bound, with probability of at least 0.98, for every $1 \leq i \leq k$,

$$|L_i(Y)| = |L'_i(W)| \leq R \cdot (2 \ln k + 10)^{1/2}.$$

$\square$

Thus, with high probability, the matrix Y satisfies the first requirement of the lemma. To prove the second requirement, we will need the following claim.

Claim 5.2 *Assume that m is large enough (i.e., $m > m_0$, for some global constant m_0). With high probability (say, with probability of at least 0.97), for any matrix D of size $m \times m$ and rank r,*

$$\|Y - D\| \geq m/3.$$

Proof:
For the proof of the claim, we will use the spectral method developed by Lokam in [Lok]. Lokam proves a similar lemma for the Hadamard matrix (and for a generalized Hadamard matrix). We will use a similar method, plus some additional facts and observations, to prove our claim for the matrix Y.

Let A be a matrix of size $m \times m$ (of real or complex numbers). The i^{th} *singular value*, $\sigma_i(A)$, is defined by

$$\sigma_i(A) = \sqrt{\lambda_i(AA^*)},$$

where A^* is the conjugate transpose of A, and $\lambda_i(AA^*)$ is the i^{th} largest eigenvalue of AA^* (for $1 \leq i \leq m$).

It is well known that for every matrix A (of size $m \times m$), there exist unitary matrices U, V (of size $m \times m$), such that U^*AV is a diagonal matrix with values $\sigma_1(A), ..., \sigma_m(A)$ on the diagonal (see, e.g., [GV], Sec. 2.3.).

For the proof of the claim we will need the following six facts. The facts are true for any constant $\epsilon > 0$. The global constant m_0 (from the statement of the claim) depends on the actual ϵ chosen (i.e., we assume that $m > m_0(\epsilon)$).

1. With high probability (say, with probability of at least 0.99),

$$\|W\| \geq (1 - \epsilon) \cdot m.$$

   Proof:
   Note that $\|W\|^2$ is the sum of the squares of m^2 standard Gaussian random variables. Hence, $\|W\|^2$ is a random variable with expectation m^2 and variance $2m^2$ and (by the central limit theorem) with very high probability its value is very close to its expectation. In particular, for large enough m, the probability for $\|W\| < (1 - \epsilon) \cdot m$ is smaller than 0.01 (this follows, e.g., by Chernoff bounds, see e.g., [ASE] Appendix A).

2. With high probability (say, with probability of at least 0.99),

$$\|W''\| \leq (1 + \epsilon) \cdot \sqrt{r \cdot m}.$$

   Proof:
   Recall that the entries of W form a multi-normal distribution. Since a multi-normal distribution doesn't change under unitary transformations and since W'' is the projection of W on V, we can present $\|W''\|^2$ as the sum of the squares of $r \cdot m$ standard Gaussian random variables. Hence, $\|W''\|^2$ is a random variable with expectation rm and variance $2rm$ and (by the central limit theorem) with very high probability its value is very

close to its expectation. In particular, for large enough m, the probability for $\|W''\| > (1+\epsilon)\cdot\sqrt{r\cdot m}$ is smaller than 0.01 (this follows, e.g., by Chernoff bounds, see e.g., [ASE] Appendix A).

3. With high probability (say, with probability of at least 0.99),

$$\sigma_1(W) < (2+\epsilon)\cdot\sqrt{m}.$$

   Proof:
   The proof was given in [Gem] (see also [Sil]).

4. For any matrix D of size $m \times m$ and rank r,

$$\sigma_{r+1}(D), ..., \sigma_m(D) = 0.$$

   Proof:
   As mentioned above, there exist unitary matrices U, V (of size $m \times m$), such that U^*DV is a diagonal matrix with values $\sigma_1(D), ..., \sigma_m(D)$ on the diagonal. Since unitary transformations do not change the rank of a matrix, we conclude that $\sigma_{r+1}(D), ..., \sigma_m(D) = 0$.

5. For any matrix A of size $m \times m$,

$$\|A\|^2 = \sigma_1^2(A) + \cdots + \sigma_m^2(A).$$

   Proof:
   As mentioned above, there exist unitary matrices U, V (of size $m \times m$), such that U^*AV is a diagonal matrix with values $\sigma_1(A), ..., \sigma_m(A)$ on the diagonal. Since unitary transformations do not change the norm of a matrix, we conclude that $\|A\|^2 = \sigma_1^2(A) + \cdots + \sigma_m^2(A)$.

6. For any two matrices A, B of size $m \times m$,

$$\sum_{i=1}^{m}[\sigma_i(A) - \sigma_i(B)]^2 \leq \|A - B\|^2.$$

   Proof:
   This inequality is know as Hoffman-Wielandt inequality [HW]. For a proof of this version of the inequality, see [GV] Sec. 8.3.

We are now ready to complete the proof of the claim. Assume that the above 6 facts are all true for $\epsilon = 0.01$ (for large enough m, this happens with probability of at least 0.97). Let D be any matrix of size $m \times m$ and rank r. By fact 6 and fact 4,

$$\|W - D\|^2 \geq \sum_{i=1}^{m}[\sigma_i(W) - \sigma_i(D)]^2 \geq \sum_{i=r+1}^{m}[\sigma_i(W)]^2.$$

By fact 5, fact 3 and fact 1 (and since $\epsilon = 0.01$ and $r = m/10$),

$$\sum_{i=r+1}^{m} [\sigma_i(W)]^2 = \|W\|^2 - \sum_{i=1}^{r} [\sigma_i(W)]^2 \geq \|W\|^2 - 4.0401 \cdot r \cdot m \geq 0.98 \cdot m^2 - 0.40401 \cdot m^2.$$

Hence,

$$\|W - D\|^2 \geq (0.75 \cdot m)^2,$$

and by the triangle inequality and fact 2,

$$\|Y - D\| \geq \|W - D\| - \|W - Y\| \geq 0.75 \cdot m - 0.32 \cdot m > m/3.$$

□

Let us now finish the proof of Lemma 4.1. Note that if $\mathrm{Rig}_r[Y] < \sqrt{m/9}$ then (by the definition of Rig) there exists a matrix D of rank r, such that, all rows of $Y - D$ are of L^2-norm $< \sqrt{m/9}$, and hence $\|Y - D\|^2 < m \cdot m/9$ (in contradiction to Claim 5.2).

Thus, by Claim 5.1 we know that with high probability Y satisfies the first requirement of the lemma, and by Claim 5.2 we know that with high probability Y satisfies the second requirement of the lemma. Altogether, with probability of at least 0.95, the matrix Y satisfies both requirements. □

6 Size-Depth Tradeoffs

The following lemma is implicit in [Lok].

Lemma 6.1 *Let C be a bounded coefficients linear circuit of size s and depth d for the linear functions $L_1, ..., L_k$ (in n variables). Then, for every $1 \leq r \leq n$,*

$$Rig_r[L_1, ..., L_k] \leq \left(\frac{s}{r}\right)^{2d}.$$

Proof:
Let $v_1, ..., v_l$ be all nodes in C of in-degree larger than s/r. Obviously, $l \leq r$. Denote by $f_1, ..., f_l$ the l linear functions outputted at the nodes $v_1, ..., v_l$. Denote by C' the circuit C after removing from it the l nodes $v_1, ..., v_l$ and all edges connected to them. Then, each L_i can be written as $L_i = L''_i + L'_i$, where L''_i is a linear combination of the functions $f_1, ..., f_l$, and L'_i is the i^{th} output of the circuit C'.

Since the maximal in-degree in C' is at most s/r and since C' is of depth d, the L^1-norm of each L'_i is bounded by $(s/r)^d$, and hence, its L^2-norm is bounded by $(s/r)^{2d}$. Hence, if we denote $V = \mathrm{Span}[f_1, ..., f_l]$ then for every $1 \leq i \leq k$, we have $\mathrm{Dist}[L_i, V] \leq (s/r)^{2d}$. Thus, $\mathrm{Rig}_r[L_1, ..., L_k] \leq \mathrm{Rig}_l[L_1, ..., L_k] \leq (s/r)^{2d}$. □

For our size-depth tradeoff for matrix product, we will also need the following version of Lemma 4.1. Note that for the proof of Theorem 1 we could have used Lemma 6.2 rather than Lemma 4.1. We preferred to use Lemma 4.1 because it makes the proof of Theorem 1 more intuitive.

Lemma 6.2 *Let $L_1, ..., L_k$ be k linear functions (over R) in the m^2 variables $y_{1,1}, ..., y_{m,m}$ (we think of $y_{1,1}, ..., y_{m,m}$ as the entries of a matrix of size $m \times m$). Denote, $r = m/10$, and assume (for simplicity) that m is large enough (i.e., $m > m_0$, for some global constant m_0). Then, there exists a matrix Y of size $m \times m$ (over R), such that:*

1. *For every $1 \leq i \leq k$,*

$$|L_i(Y_{1,1}, ..., Y_{m,m})| \leq Rig_{r \cdot m}[L_1, ..., L_k] \cdot (2 \ln k + 10)^{1/2}.$$

2. $$Rig_{r \cdot m}[I \otimes Y] \geq \sqrt{m/9}.$$

Proof:
The proof is similar to the proof of Lemma 4.1.

We define R, W, W'' and Y as in the proof of Lemma 4.1. Thus, by Claim 5.1, with high probability, the matrix Y satisfies the first requirement of the lemma. To prove the second requirement, we will need the following version of Claim 5.2.

Claim 6.1 *Assume that m is large enough (i.e., $m > m_0$, for some global constant m_0). With high probability (say, with probability of at least 0.97), for any matrix D of size $m^2 \times m^2$ and rank $r \cdot m$,*

$$\|(I \otimes Y) - D\| \geq m^{1.5}/3.$$

Proof:
The proof is similar to the proof of Claim 5.2. The first three facts (out of the six given in Claim 5.2) are replaced by the following three facts. As before, the facts are true for any constant $\epsilon > 0$. The global constant m_0 (from the statement of the claim) depends on the actual ϵ chosen (i.e., we assume that $m > m_0(\epsilon)$).

1. With high probability (say, with probability of at least 0.99),

$$\|I \otimes W\| \geq (1 - \epsilon) \cdot m^{1.5}.$$

   Proof:
   Obvious, since $\|I \otimes W\| = m^{0.5} \cdot \|W\|$.

2. With high probability (say, with probability of at least 0.99),

$$\|I \otimes W''\| \leq (1 + \epsilon) \cdot r^{0.5} \cdot m.$$

   Proof:
   Obvious, since $\|I \otimes W''\| = m^{0.5} \cdot \|W''\|$.

3. With high probability (say, with probability of at least 0.99),

$$\sigma_1(I \otimes W) < (2+\epsilon) \cdot m^{0.5}.$$

Proof:
Obvious, since $\sigma_1(I \otimes W) = \sigma_1(W)$.

The proof of the claim is now completed as before. Assume that the above 6 facts are all true for $\epsilon = 0.01$ (for large enough m, this happens with probability of at least 0.97). Let D be any matrix of size $m^2 \times m^2$ and rank $r \cdot m$. By fact 6 and fact 4,

$$\|(I \otimes W) - D\|^2 \geq \sum_{i=1}^{m^2} [\sigma_i(I \otimes W) - \sigma_i(D)]^2 \geq \sum_{i=r\cdot m+1}^{m^2} [\sigma_i(I \otimes W)]^2.$$

By fact 5, fact 3 and fact 1 (and since $\epsilon = 0.01$ and $r = m/10$),

$$\sum_{i=r\cdot m+1}^{m^2} [\sigma_i(I \otimes W)]^2 = \|I \otimes W\|^2 - \sum_{i=1}^{r\cdot m} [\sigma_i(I \otimes W)]^2 \geq 0.98 \cdot m^3 - 0.40401 \cdot m^3.$$

Hence,

$$\|(I \otimes W) - D\|^2 \geq (0.75 \cdot m^{1.5})^2,$$

and by the triangle inequality and fact 2,

$$\|(I \otimes Y) - D\| \geq \|(I \otimes W) - D\| - \|(I \otimes W) - (I \otimes Y)\| > m^{1.5}/3.$$

□

Let us now finish the proof of Lemma 6.2. Note that if $\mathrm{Rig}_{r\cdot m}[I \otimes Y] < \sqrt{m/9}$ then (by the definition of Rig) there exists a matrix D of rank $r \cdot m$, such that, all rows of $(I \otimes Y) - D$ are of L^2-norm $< \sqrt{m/9}$, and hence $\|(I \otimes Y) - D\|^2 < m^2 \cdot m/9$ (in contradiction to Claim 6.1).

Thus, by Claim 5.1 we know that with high probability Y satisfies the first requirement of the lemma, and by Claim 6.1 we know that with high probability Y satisfies the second requirement of the lemma. Altogether, with probability of at least 0.95, the matrix Y satisfies both requirements. □

We will now state and prove our size-depth tradeoff for matrix product. We didn't attempt here to optimize the constant ϵ.

Theorem 2 *Let C be a bounded coefficients arithmetic circuit of depth d (over the real or complex numbers) for the product of two matrices of size $m \times m$. Then, for some global constant $\epsilon > 0$ (say, $\epsilon = 1/20$),*

$$Size(C) = \Omega(m^{2+\epsilon/d}).$$

Proof:
The proof follows the lines of the proof of Theorem 1. As before, w.l.o.g. we assume that the circuit is over the reals and that the circuit is bilinear. We assume w.l.o.g. that m is large enough (and in particular, $m > m_0$, where m_0 is the global constant from Lemma 6.2), and we assume for simplicity that $m/10$ is integer. Define,

$$r = m/10.$$

Assume, for a contradiction to the statement of the lemma, that

$$\mathrm{Size}(C) < 0.001 \cdot m^{2+\epsilon/d}.$$

As before, denote by $v_1, ..., v_k$ the product gates of the circuit C. Since the circuit is bilinear, each product gate v_i computes the product of two linear functions, one in the variables $\{x_{i,j}\}$ (of the first matrix) and the other in the variables $\{y_{i,j}\}$ (of the second matrix). Denote the first linear function by R_i and the second linear function by L_i. Thus, v_i computes the product of $R_i(x_{1,1}, ..., x_{m,m})$ and $L_i(y_{1,1}, ..., y_{m,m})$.

Consider the linear functions $L_1, ..., L_k$. These functions are computed by a linear circuit of depth at most d and size at most $0.001 \cdot m^{2+\epsilon/d}$ (in the input variables $y_{1,1}, ..., y_{m,m}$). Hence, by Lemma 6.1,

$$\mathrm{Rig}_{r \cdot m}[L_1, ..., L_k] < (0.01)^{2d} \cdot m^{2\epsilon}.$$

Hence, by Lemma 6.2, there exists a matrix Y of size $m \times m$ (over R), such that:

1. For every $1 \leq i \leq k$,

   $$|L_i(Y_{1,1}, ..., Y_{m,m})| \leq (0.01)^{2d} \cdot m^{2\epsilon} \cdot (2 \ln k + 10)^{1/2} < (0.01)^{2d} \cdot m^{2\epsilon} \cdot \ln m,$$

   (for large enough m).

2. $$\mathrm{Rig}_{r \cdot m}[I \otimes Y] \geq \sqrt{m/9}.$$

Denote,

$$c = (0.01)^{2d} \cdot m^{2\epsilon} \cdot \ln m.$$

We fix the input variables $y_{1,1}, ..., y_{m,m}$ to be the entries $Y_{1,1}, ..., Y_{m,m}$. Denote the obtained circuit by C'. Since we fixed $y_{1,1}, ..., y_{m,m}$, each product gate v_i in C turned into a product with the field element $L_i(Y)$. The circuit C' is hence a linear arithmetic circuit for the matrix $I \otimes Y$ in the input variables $x_{1,1}, ..., x_{m,m}$.

As before, the circuit C' is not a bounded coefficients arithmetic circuit. We will convert C' into a bounded coefficients arithmetic circuit C'' by the following two steps: First, replace the product with each field element $L_i(Y)$ by a product with the field element $L_i(Y)/c$ (which is of absolute value ≤ 1) and multiply each output of the circuit by the field element c. Then, replace each product (of an output) with the field element c by $2d$ consecutive additions of fan-in $c^{1/2d}$ each (plus one product with a field element of absolute value ≤ 1). Since the number of outputs is m^2, this increases the size of the circuit by at most $m^2 \cdot 2d \cdot c^{1/2d}$ and hence the size of C'' is at most $0.01 \cdot m^{2+\epsilon/d} \cdot \ln m$ (for large enough m).

Thus, the obtained circuit C'' is a bounded coefficients arithmetic circuit of depth $3d$ that computes the matrix $I \otimes Y$, and such that

$$\text{Size}(C'') < 0.01 \cdot m^{2+\epsilon/d} \cdot \ln m.$$

However, since C'' is of depth $3d$, by Lemma 6.1,

$$\text{Size}(C'') \geq r \cdot m \cdot (\text{Rig}_{r \cdot m}[I \otimes Y])^{1/6d} \geq 0.01 \cdot m^{2+1/12d},$$

which is a contradiction (for large enough m and, say, $\epsilon = 1/20$). □

Acknowledgment

I would like to thank Pavel Pudlak, Sasha Razborov, Amir Shpilka and Avi Wigderson for very helpful conversations.

References

[ASE] N. Alon, J.H. Spencer and P.Erdos. *The Probabiliatic Method.* John Wiley and Sons, Inc., 1992.

[Bla] M. Bläser. A $2.5n$ lower bound for the rank of $n \times n$ matrix multiplication over arbitrary fields. In *40th IEEE Symposium on Foundations of Computer Science*, pages 45–50, 1999.

[Bsh] N.H. Bshouty. A lower bound for matrix multiplication. *SIAM Journal on Computing*, 18:759–765, 1989.

[BL] P. Burgisser and M. Lotz. Lower Bounds on the Bounded Coefficient Complexity of Bilinear Maps. In *43th IEEE Symposium on Foundations of Computer Science*, 2002.

[Cha] B. Chazelle. A spectral approach to lower bounds with applications to geometric searching. *SIAM Journal on Computing*, 27:545–556, 1998.

[CW] D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progressions. *J. Symbolic Comput.*, 9:251–280, 1990.

[Gat] J.V.Z. Gathen. Algebraic complexity theory. *Ann. Rev. Computer Science*, 1988, pages 317–347.

[Gem] S. Geman. A limit theorem for the norm of random matrices. *Annals of Probability*, 8:252–261, 1980.

[GV] G.H. Golub and C.F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, 1983.

[HW] A.J. Hoffman and H.W. Wielandt. The variation of the spectrum of normal matrices. *Duke Mathematical Journal* 20:37–39, 1953.

[Lok] S.V. Lokam. Spectral methods for matrix rigidity with applications to size-depth trade-offs and communication complexity. In *36th IEEE Symposium on Foundations of Computer Science*, pages 6–15, 1995.

[Mor] J. Morgenstern. Note on a lower bound of the linear complexity of the fast fourier transform. *JACM*, 20(2): 305–306, 1973.

[NW] N. Nissan and A. Wigderson. On the complexity of bilinear forms. In *27th ACM Symposium on Theory of Computing*, pages 723–732, 1995.

[Pud] P. Pudlak. A note on using the detrminant for proving lower bounds on the size of linear circuits. *Electronic Colloquium on Computational Complexity (ECCC)*, Report No. 42, 1998.

[RS] R. Raz and A. Shpilka. Lower bounds for matrix product, in bounded depth circuits with arbitrary gates. In *33rd ACM Symposium on Theory of Computing*, pages 409–418, 2001.

[Shp] A. Shpilka. Lower bounds for matrix product. In *42nd IEEE Symposium on Foundations of Computer Science*, 2001.

[Sil] J.W. Silverstein. The smallest eigenvalue of a large dimensional Wishart matrix. *Annals of Probability*, 13:1364–1368, 1985.

[Str] V. Strassen. Gaussian elimination is not optimal. *Numer. Math*, 13:354–356, 1969.