DEPTH-4 LOWER BOUNDS, 
DETERMINANTAL COMPLEXITY: 
A UNIFIED APPROACH

SURYAJITH CHILLARA 
AND PARtha MUKHOPADHYAY

Abstract. Tavenas (Proceedings of mathematical foundations of computer science (MFCS), 2013) has recently proved that any $n^{O(1)}$-variate and degree $n$ polynomial in VP can be computed by a depth-4 $\Sigma\Pi\Sigma\Pi$ circuit of size $2^{O(\sqrt{n \log n})}$. So, to prove $\text{VP} \neq \text{VNP}$ it is sufficient to show that an explicit polynomial in VNP of degree $n$ requires $2^{\omega(\sqrt{n \log n})}$ size depth-4 circuits. Soon after Tavenas’ result, for two different explicit polynomials, depth-4 circuit-size lower bounds of $2^{\Omega(\sqrt{n \log n})}$ have been proved (see Kayal et al. in Proceedings of symposium on theory of computing, ACM, 2014b. http://doi.acm.org/10.1145/2591796.2591847; Fournier et al. in Proceedings of symposium on theory of computing, ACM, 2014). In particular, using a combinatorial design Kayal et al. (2014b) construct an explicit polynomial in VNP that requires depth-4 circuits of size $2^{\Omega(\sqrt{n \log n})}$ and Fournier et al. (Proceedings of symposium on theory of computing, ACM, 2014) show that the iterated matrix multiplication polynomial (which is in VP) also requires $2^{\Omega(\sqrt{n \log n})}$ size depth-4 circuits.

In this paper, we identify a simple combinatorial property such that any polynomial $f$ that satisfies this property would achieve a similar depth-4 circuit-size lower bound. In particular, it does not matter whether $f$ is in VP or in VNP. As a result, we get a simple unified lower-bound analysis for the above-mentioned polynomials.

Another goal of this paper is to compare our current knowledge of the depth-4 circuit-size lower bounds and the determinantal complexity lower bounds. Currently, the best known determinantal complexity lower bound is $\Omega(n^2)$ for permanent of a $n \times n$ matrix (which is a $n^2$-variate and degree $n$ polynomial) due to Cai et al. (Proceedings of
symposium on theory of computing, ACM, 2008). We prove that the determinantal complexity of the iterated matrix multiplication polynomial is \( \Omega(dn) \) where \( d \) is the number of matrices and \( n \) is the dimension of the matrices. In particular, our result settles the determinantal complexity of the iterated matrix multiplication polynomial to \( \Theta(dn) \).

To the best of our knowledge, a \( \Theta(n) \) bound for the determinantal complexity for the iterated matrix multiplication polynomial was known only for any constant \( d > 1 \), due to Jansen (Theory Comput Syst 49(2):343–354, 2011).

**Keywords.** Lower bounds, Determinantal complexity, Constant depth, Arithmetic circuits

**Subject classification.** 68Q17

### 1. Introduction

In a surprising result, Agrawal & Vinay (2008) showed that proving exponential size lower bounds against depth-4 circuits imply exponential size lower bounds for general arithmetic circuits. In particular, given a polynomial-sized (or subexponential) general arithmetic circuit, it can be transformed into a depth-4 circuit of subexponential size. Koiran (2012) and Tavenas (2013) carefully analysed the chasm shown in Agrawal & Vinay (2008) and came up with an improved depth reduction.\(^1\)

**Theorem 1.1.** (Agrawal & Vinay (2008); Koiran (2012), Tavenas (2013)) Let \( f \) be a polynomial of degree \( d \) over \( n \) variables and is computed by an arithmetic circuit \( C \) of size \( s \). Then, for any \( 0 < t \leq d \), \( f \) can also be computed by a homogeneous \( \Sigma \Pi[O(d/t)]\Sigma \Pi[t] \) circuit of top fan-in \( s^{O(d/t)} \) and size \( s^{O(t+d/t)} \).

The above theorem tells us that proving a size lower bound of \( n^{w(d/t)} \) against the \( \Sigma \Pi[O(d/t)]\Sigma \Pi[t] \) would imply superpolynomial circuit size lower bounds against general circuits. Towards proving such lower bounds, Gupta et al. (2013) proved a lower bound of \( 2^{\Omega(n/t)} \) against \( \Sigma \Pi[O(d/t)]\Sigma \Pi[t] \) circuits computing the determinant

\(^{1}\) A newer proof of the chasm which yields a more structured depth-4 circuit was presented in Chillara et al. (2016).
or the permanent polynomial over a $n \times n$ matrix. They used the dimension of the shifted partial derivative space as the complexity measure.\footnote{Kayal (2012) introduced this measure to prove exponential circuit size lower bounds against the depth-four circuits of the form of sums of powers of constant degree homogeneous polynomials, which compute the monomial $x_1 x_2 \ldots x_n$.} Kayal et al. (2014b) pushed the bound of Gupta et al. (2013) to $N^{\Omega(d/t)}$ for $\Sigma^\ell \Pi[O(d/t)]^\ell \Pi[d]$ circuits computing an explicit polynomial in VNP of degree $d$ over $N$ variables. This candidate polynomial is based on the combinatorial designs of Nisan & Wigderson (1994) (see Definition 2.2).

In another surprising result, Fournier et al. (2014) proved that a matching bound could also be obtained for a polynomial in VP, the iterated matrix multiplication polynomial (see Definition 2.3). This tells us that the bound of Tavenas is tight (up to a constant in the exponent) and thus rules out any improvement of the depth reduction. However, it is important to note that the constant in the exponent of the bound proved by Kayal et al. (2014b) or the one by Fournier et al. (2014) is weaker than the constant in the bound of Tavenas (2013).

One of the main motivations of our study comes from this tantalizing fact that two seemingly different polynomials $NW_{n,r}$ (which is in VNP) and $IMM_{n,n}$ (which is in VP) behave very similarly as far as the $2^{\Omega(\sqrt{n \log n})}$-size lower bound against $\Sigma^\ell \Pi[O(\sqrt{n})] \Sigma^\ell \Pi[\sqrt{n}]$ circuits are concerned. In this paper, we seek a conceptual reason for this behaviour. We identify a simple combinatorial property such that any $n^{O(1)}$-variate polynomial of degree $d$ that satisfies it would require $2^{\Omega(\sqrt{d \log n})}$ size $\Sigma^\ell \Pi[O(\sqrt{d})] \Sigma^\ell \Pi[\sqrt{d}]$ circuits. We call this the Leading Monomial Distance Property. In particular, it does not matter whether the polynomial is easy (i.e. in VP) or hard (i.e. the polynomial is in VNP but not known to be in VP). As a result of this abstraction, we present a simple unified analysis of the bounded fan-in depth-4 circuit size lower bounds for the $NW_{n,r}$ and $IMM_{n,d}$ polynomials. Formally, we prove the following.

**Theorem 1.2.** Let $\delta, \varepsilon, c$ be some constants such that $\delta > 0$, $c > 1$ and $0 < \varepsilon \leq \frac{1}{40c}$. Let $t$ be a parameter such that $\omega(\ln d) \leq t \leq o\left(\frac{d}{\ln d}\right)$. Let $f$ be a polynomial over $N = d^{O(1)}(\geq d^2)$ variables, of
degree $d$. Let there be at least $d^{\delta k}$ different polynomials in $\langle \partial^k(f) \rangle$ for $k = \varepsilon_d^l$ such that any two of their leading monomials have a distance of at least $\Delta \geq \frac{d}{c}$. Then any depth-4 $\Sigma \Pi \Pi[O(d/t)] \Sigma \Pi[t]$ circuit that computes $f$ must be of size $e^{\Omega_{\delta,c} \left( \frac{d}{c} \ln N \right)}$.

By $\Omega_{\delta,c}(\cdot)$, we mean that the constant hidden under $\Omega$ is dependent on $\delta$ and $c$. It is now well understood that the currently known techniques cannot help prove better size lower bounds of the order of $n^{\omega(d/t)}$ against $\Sigma \Pi[O(d/t)] \Sigma \Pi[t]$ circuits computing explicit polynomials in $\text{VNP}$ of degree $d$ over $n^{O(1)}$-variables. However, some very interesting lower bounds were proved in the recent past over some related models: Kayal et al. (2014a) and Kumar & Saraf (2014a,b, 2016a,b) (see Saptharishi (2015) for an exposition of each of these results).

On the other hand, showing quasi-polynomial determinantal complexity lower bounds for an explicit polynomial in $\text{VNP}$ would imply Valiant’s hypothesis as well. Let us recall the following definitions before we introduce the notion of Determinantal Complexity. A multivariate polynomial family $\{f_n(X) \in \mathbb{F}[x_1, x_2, \ldots, x_{n^{O(1)}}] : n \geq 1\}$ is in the class $\text{VP}$ if $f_n$ has degree of at most $\text{poly}(n)$ and can be computed by an arithmetic circuit of size $\text{poly}(n)$. A multivariate polynomial $\{f_n(X) \in \mathbb{F}[x_1, x_2, \ldots, x_{n^{O(1)}}] : n \geq 1\}$ is in the class $\text{VNP}$ if it can be expressed as

$$f_n(X) = \sum_{Y \in \{0,1\}^m} g_{n+m}(X,Y)$$

where $m = |Y| = \text{poly}(n)$ and $g_{n+m}$ is a polynomial family in $\text{VP}$. The Permanent polynomial characterizes the class $\text{VNP}$ over the fields of all characteristics except 2 and the determinant polynomial characterizes the class $\text{VP}$ with respect to the quasi-polynomial projections.

**Definition 1.3.** The determinantal complexity of a polynomial $f$, over $n$ variables is the minimum $m$ such that there are $m^2$ many affine linear polynomials $A_{k,\ell}$, $1 \leq k, \ell \leq m$ defined over the same set of variables and $f = \det((A_{k,\ell})_{1 \leq k, \ell \leq m})$. It is denoted by $\text{DetComp}(f)$. 
To resolve Valiant’s hypothesis, proving $\text{DetComp}(\text{Perm}_n) = n^{\omega(\log n)}$ is sufficient. von zur Gathen (1986) proved that $\text{DetComp}(\text{Perm}_n) \geq \sqrt{\frac{8}{7}} n$. Later, Cai (1990), Babai and Seress (cited in von zur Gathen (1987)), and Meshulam (1989) independently improved the lower bound to $\sqrt{2n}$. Mignon & Ressayre (2004) proved that $\text{DetComp}(\text{Perm}_n) \geq \frac{n^2}{2}$ over the fields of characteristic zero, using algebraic geometry. Subsequently, Cai et al. (2008) extended the result of Mignon & Ressayre (2004) to all fields of characteristic $\neq 2$. They also provided a simpler analysis.

For any polynomial $f$, Valiant (1979) proved that $\text{DetComp}(f) \leq 2(F(f) + 1)$ where $F(f)$ is the arithmetic formula complexity of $f$. Toda (1992) showed that $\text{DetComp}(f) = B(f)$ where $B(f)$ is the arithmetic branching program complexity of $f$.

Our main result in this context is a lower bound on the determinantal complexity of the iterated matrix multiplication polynomial.

**Theorem 1.4.** For all integers $n$ and $d > 1$, the determinantal complexity of the iterated matrix multiplication polynomial $\text{IMM}_{n,d}$ is at least $d(n - 1)/2$.

Since $\text{IMM}_{n,d}(X)$ has an algebraic branching program of size $O(dn)$, from the above theorem it follows that $\text{DetComp}(\text{IMM}_{n,d}(X)) = \Theta(dn)$. This improves upon the earlier bound of $\Theta(n)$ for the determinantal complexity of the iterated matrix multiplication polynomial for any constant $d > 1$, due to Jansen (2011). We follow the approach of Cai et al. (2008) and Mignon & Ressayre (2004) and use the rank of the Hessian matrix as our main technical tool.

As mentioned before, the current best known determinantal complexity lower bound for an explicit polynomial in VNP is only quadratic, for the permanent polynomial Mignon & Ressayre (2004). Before the result of Mignon & Ressayre (2004), the best known determinantal complexity lower bound for the $n \times n$ permanent polynomial was $\sqrt{2n}$ (cf. Cai 1990; von zur Gathen 1987). These results were proved using nontrivial algebraic-geometric concepts. One possible approach to prove that $\text{VP} \neq \text{VNP}$ could be by proving a super-quasi-polynomial determinantal complexity lower
bound for any other explicit polynomial in VNP. One such polynomial that we consider is the Nisan–Wigderson polynomial.

Here we first show that DetComp(NW_{n,εn}) \geq Ω(n^{1.5}) using elementary ideas. Similarly, we prove a weaker (compared to the one in Theorem 1.4) lower bound on the determinantal complexity of IMM_{n,d} using the method of partial derivatives.

In general, a lower bound on the determinantal complexity of a polynomial also implies a lower bound on the formula complexity. But here, in the case of the iterated matrix multiplication polynomial, the best bound on the determinantal complexity that we can get is Ω(dn) for it has an algebraic branching program of that size. A lower bound of Ω(dn) is trivial for the formula complexity as the polynomial is n^2d-variate. For the sake of completeness, we show a super-linear lower bound on the size of any arithmetic formula computing the iterated matrix multiplication polynomial.

**Theorem 1.5.** For all integers n, d > 0, any arithmetic formula computing the IMM_{n,d}(X) polynomial must be of size Ω(dn^3).

We prove this by adapting the argument of Kalorkoti (1985) to the iterated matrix multiplication polynomial.

## 2. Preliminaries

### Arithmetic circuits.

An arithmetic circuit over a field \( \mathbb{F} \) over the set of variables \( \{x_1, x_2, \ldots, x_n\} \) is a directed acyclic graph such that the internal nodes are labelled by addition or multiplication gates and the leaf nodes are labelled by variables or field elements. Any node with fan-out zero is an output gate. An arithmetic circuit computes a polynomial in the polynomial ring \( \mathbb{F}[x_1, x_2, \ldots, x_n] \).

The size of an arithmetic circuit is the number of nodes, and the depth is the length of a longest path from the root to a leaf node.

### Depth-4 circuits.

Usually, a depth-4 circuit over a field \( \mathbb{F} \) is denoted by \( \Sigma \Pi \Sigma \Pi \). A \( \Sigma \Pi^{[D]} \Sigma \Pi^{[l]} \) circuit computes the polynomials of the form

\[
f(x_1, \ldots, x_n) = \sum_{i=1}^{s} \prod_{j=1}^{d_i} Q_{i,j}(x_1, \ldots, x_n)
\]
where $d_i < D$ for all $i$ and each $Q_{i,j}$ is a polynomial of degree at most $t$ over $\mathbb{F}[x_1, \ldots, x_n]$.

The following beautiful lemma (from Gupta et al. 2013) is key to the asymptotic estimates required for the lower bound analyses.

**Lemma 2.1** (Lemma 6, Gupta et al. 2013). Let $a(n), f(n), g(n) : \mathbb{Z}_{\geq 0} \rightarrow \mathbb{Z}_{\geq 0}$ be integer-valued functions such that $(f + g) = o(a)$. Then,

$$\ln \frac{(a + f)!}{(a - g)!} = (f + g) \ln a \pm O\left(\frac{(f + g)^2}{a}\right)$$

**Polynomial families.**

**Definition 2.2** (Nisan–Wigderson polynomial). For integers $n > 0$ ranging over prime powers and an integer $r$, we define a polynomial family $\{\text{NW}_{n,r}\}$ as follows.

$$\text{NW}_{n,r}(X) = \sum_{a(z) \in \mathbb{F}_n[z]} x_{1,a(1)} x_{2,a(2)} \ldots x_{n,a(n)}$$

where $a(z)$ runs over all univariate polynomials of degree $< r$ and $X = \{x_{i,j} : (i, j) \in [n] \times [n]\}$.

Proposition 4 in Valiant (1979) tells us that, given a polynomial $P(X) \in \mathbb{F}[X]$ with $\{0, 1\}$ coefficients, if there is a polynomial time algorithm to test whether the coefficient of a given monomial is 1, then $P(X)$ is in VNP over $\mathbb{F}$. Given any monomial $m$ over $X$, we can decide in polynomial time if it indeed is a monomial in the polynomial $\{\text{NW}_{n,r}\}_{n>0}$ by checking whether it conforms to a univariate polynomial of degree at most $r$. Thus, $\{\text{NW}_{n,r}\}_{n>0}$ is in VNP.

**Definition 2.3** (Iterated matrix multiplication polynomial). The iterated matrix multiplication polynomial over $d$ generic $n \times n$ matrices $X_1, X_2, \ldots, X_d$ is the $(1, 1)$th entry in the product of these matrices. More formally, let $X_1, X_2, \ldots, X_d$ be $d$ generic $n \times n$ matrices over disjoint sets of variables. For any $k \in [d]$, let
Let $x_{ij}^{(k)}$ be the variable in $X_k$ indexed by $(i, j) \in [n] \times [n]$. Then the iterated matrix multiplication polynomial, denoted by the family $\{\text{IMM}_{n,d}\}$, is defined as follows.

$$\text{IMM}_{n,d}(X) = \sum_{i_1,i_2,\ldots,i_{d-1} \in [n]} x_{i_1,i_2,\ldots,x_{i_{d-2},i_{d-1}}}^{(d)}.$$

This is a canonical polynomial for the algebraic branching programs and thus is in VP.

**Shifted partial derivatives.** For a monomial $x^i = x_1^{i_1}x_2^{i_2}\ldots x_n^{i_n}$, let $\partial^i f$ be the partial derivative of $f$ with respect to the monomial $x^i$. The degree of the monomial is denoted by $|i|$ where $|i| = (i_1 + i_2 + \cdots + i_n)$. We also use the word length to refer to the degree of the monomial. We recall the following definition of shifted partial derivatives from Gupta et al. (2013).

**Definition 2.4.** Let $f(X) \in \mathbb{F}[X]$ be a multivariate polynomial. The span of the $\ell$-shifted $k$th-order derivatives of $f$, denoted by $\langle \partial^{=k}f \rangle_{\leq \ell}$, is defined as

$$\langle \partial^{=k}f \rangle_{\leq \ell} = \mathbb{F}\text{-span}\{x^i \cdot (\partial^j f) : i,j \in \mathbb{Z}^n_{\geq 0} \text{ with } |i| \leq \ell \text{ and } |j| = k\}$$

We denote by $\dim(\langle \partial^{=k}f \rangle_{\leq \ell})$ the dimension of the vector space $\langle \partial^{=k}f \rangle_{\leq \ell}$.

Let $\succ$ be any admissible monomial ordering Cox et al. (2007). The leading monomial of a polynomial $f(X) \in \mathbb{F}[X]$, denoted by $\text{LM}(f)$, is the largest monomial $x^i \in f(X)$ under the order $\succ$. The next lemma follows directly from Proposition 11 and Corollary 12 of Gupta et al. (2013).

**Lemma 2.5.** For any multivariate polynomial $f(X) \in \mathbb{F}[X]$, 

$$\dim(\langle \partial^{=k}f \rangle_{\leq \ell}) \geq \#\{x^i \cdot \text{LM}(g) : i,j \in \mathbb{Z}^n_{\geq 0} \text{ with } |i| \leq \ell, |j| = k, \text{ and } g \in \mathbb{F}\text{-span}\{\partial^j f\}\}$$

The following upper bound on the dimension of the shifted partial derivative space of the polynomials computed by $\Sigma \Pi^{[D]} \Sigma \Pi^{[d]}$ circuits was shown in Kayal et al. (2014b).
Lemma 2.6 (Lemma 4, Kayal et al. 2014b). If $C = \sum_{i=1}^{s'} Q_{i1} Q_{i2} \ldots Q_{iD}$ where each $Q_{ij} \in \mathbb{F}[X]$ is a polynomial of degree bounded by $t$. Then for any $k \leq D$,

$$\dim(\langle \partial^k(C) \rangle_{\leq \ell}) \leq s' \binom{D}{k} \left( N + \ell + k(t-1) \right)$$

**Leading monomial distance property.** To define the Leading Monomial Distance Property, we first define the notion of distance between two monomials.

**Definition 2.7.** Let $m_1, m_2$ be two monomials over a set of variables. Let $S_1$ and $S_2$ be the multisets of variables corresponding to the monomials $m_1$ and $m_2$, respectively. The distance $\text{dist}(m_1, m_2)$ between the monomials $m_1$ and $m_2$ is the

$$\min\{|S_1| - |S_1 \cap S_2|, |S_2| - |S_1 \cap S_2|\}$$

where the cardinalities are the order of the multisets.

For example, let $m_1 = x_1^2 x_2^2 x_3 x_4$ and $m_2 = x_1 x_2^2 x_3 x_5 x_6$. Then $S_1 = \{x_1, x_1, x_2, x_3, x_3, x_4\}$, $S_2 = \{x_1, x_2, x_2, x_3, x_5, x_6\}$, $|S_1| = 6$, $|S_2| = 6$ and $\text{dist}(m_1, m_2) = 3$. It is important to note that two distinct monomials could have distance 0 between them if one of them is a multiple of the other and hence the triangle inequality does not hold.

We say that a $n^{O(1)}$-variate and $n$-degree polynomial has the Leading Monomial Distance Property, if the leading monomials of a large subset ($\approx n^{\delta k}$) of its span of the derivatives (of order $\approx k$) have good pairwise distance for a suitable parameter $k$.

### 3. Unified analysis

In this section, we first prove a simple combinatorial lemma which we believe is the crux of the best known bounded fan-in depth-4 circuit size lower bound results. In fact, the lower bounds on the size of $\Sigma \Pi^{O(\sqrt{n})} \Sigma \Pi^{O(\sqrt{n})}$ circuits computing the polynomials $\text{NW}_{n,r}$ and $\text{IMM}_{n,n}$ follow easily from this lemma by suitably setting the parameters.
Lemma 3.1. Let \( m_1, m_2, \ldots, m_s \) be some monomials over \( N \) variables such that \( \text{dist}(m_i, m_j) \geq \Delta \) for all \( i \neq j \). Let \( M \) be the set of monomials of the form \( m_i m' \) where \( 1 \leq i \leq s \) and \( m' \) is a monomial of length at most \( \ell \) over the same set of \( N \) variables. Then, the cardinality of \( M \) is at least \( \left( sB - s^2 \left( \binom{N+\ell-\Delta}{N} \right) \right) \) where \( B = \binom{N+\ell}{N} \).

Proof. Let \( B_i \) be the set of all monomials \( m_i m' \) where \( m' \) is a monomial of length at most \( \ell \). It is easy to see that \( |B_i| = \binom{N+\ell}{N} \).

We would like to estimate \( |\bigcup_{i=1}^s B_i| \). Using the principle of inclusion and exclusion, we get \( |\bigcup_{i=1}^s B_i| \geq \sum_{i \in [s]} |B_i| - \sum_{i,j \in [s], i \neq j} |B_i \cap B_j| \).

Now we estimate the upper bound for \( |B_i \cap B_j| \) such that \( i \neq j \). Consider the monomials \( \tilde{m}_i \) and \( \tilde{m}_j \) in \( B_i \) and \( B_j \), respectively. For \( \tilde{m}_i \) and \( \tilde{m}_j \) to match, \( \tilde{m}_i \) should contain at least \( \Delta \) variables from \( \tilde{m}_j \) and similarly \( \tilde{m}_j \) should contain at least \( \Delta \) variables from \( \tilde{m}_i \). The rest of the at most \( \ell - \Delta \) degree monomials should be identical in \( \tilde{m}_i \) and \( \tilde{m}_j \). The number of such monomials over \( N \) variables is at most \( \binom{N+\ell-\Delta}{N} \). Thus, \( |B_i \cap B_j| \leq \binom{N+\ell-\Delta}{N} \).

Then the total number of monomials of the form \( m_i m' \) for all \( i \in [s] \) where \( m' \) is a monomial of length at most \( \ell \) is lower bounded as follows.

\[
|\bigcup_{i=1}^s B_i| \geq sB - s^2 \binom{N+\ell-\Delta}{N} = sB \left( 1 - \frac{s}{B} \binom{N+\ell-\Delta}{N} \right)
\]

□

We use the above lemma to prove our main theorem.

Proof of the Main Theorem (Theorem 1.2)

Proof. Consider a set of \( s \) many polynomials \( f_1, f_2, \ldots, f_s \in \langle \partial^{=k}(f) \rangle \) such that their leading monomials have a pairwise distance of at least \( \frac{d}{c} \).

By invoking Lemma 3.1 with the parameters \( s = d^{\delta k}, \Delta = d/c \), we get that \( |\bigcup_{i=1}^s B_i| \geq sB \left( 1 - \frac{s}{B} \binom{N+\ell-\Delta}{N} \right) \). To get an optimal lower bound on \( |\bigcup_{i=1}^s B_i| \), we will try to ensure that \( \frac{s \binom{N+\ell-\Delta}{N}}{\binom{N+\ell}{N}} \) is at most \( \frac{1}{p(d)} \) where \( p(d) \) is some function of \( d \) that we will soon fix.
Putting it all together, we get the following. Using Lemma 2.1, we shall tightly estimate the subsequent computations. Towards that end, we shall choose the parameter $\ell$ such that $\Delta^2 = o(N + \ell)$. Thus, by using Lemma 2.1 we can derive that $s \left( \frac{\ell}{N+\ell} \right)^\Delta \leq \frac{1}{p(d)}$. That is, $s \left( \frac{1}{1 + \frac{\ell}{N}} \right)^\Delta \leq \frac{1}{p(d)}$. We use the inequality $1 + x > e^{x/2}$ for $0 < x < 1$ to lower bound $\left( 1 + \frac{N}{\ell} \right)^\Delta$ by $e^{\frac{N\Delta}{2\ell}}$. Thus, it is enough to choose $\ell$ in a way that $s \cdot p(d)$ is at most $e^{\frac{N\Delta}{2\ell}}$. That is, we need $\ell$ to be at most $\frac{N\Delta}{2\ln(s \cdot p(d))}$. Let us fix $p(d)$ to $d^\delta k$ and substitute for the parameters $k$ and $\Delta$ where necessary. We now want $\ell$ to be at most $\frac{N\ell}{4c\delta d \ln d}$.

Let us fix $\ell$ to $\frac{10N\ell}{d\ln d}$. Since $\varepsilon \leq \frac{1}{4d\ell}$, we can now see that $\ell$ is indeed at most $\frac{N\ell}{4c\delta d \ln d}$, as required. Further, we can also verify that $\Delta^2$ is $o(N + \ell)$ using the facts that $t = \omega(d \ln d), N \geq d^2$ and $\Delta^2 = \frac{d^2}{c^2}$ (which implies that $\ell = \omega(d^2)$).

Using Lemma 2.5, we infer that the dimension of $\langle \partial^{-k} f \rangle \leq k$ is at least $\left( 1 - \frac{1}{d\delta k} \right) s \left( \frac{N + \ell}{N} \right)$. Combining this with Lemma 2.6, we get that

$$s' \geq \frac{1 - \frac{1}{d\delta k}}{(D/k) \left( \frac{N + \ell + k(t - 1)}{N} \right)}.$$

For this value of $\ell$ and the range of $t$, it is easy to see that $(kt - k)^2 = o(\ell)$. By invoking Lemma 2.1 with $(kt - k)^2 = o(\ell)$, we can easily show the following.

$$s' \geq \frac{s \left( 1 - \frac{1}{d\delta k} \right)}{(D/k) \left( 1 + \frac{N}{\ell} \right)} \geq d^\delta k \left( 1 - \frac{1}{d\delta k} \right) \left( \frac{D}{k} \right) e^{\frac{N\Delta}{\ell}\kappa}.$$

Since $D = O(d/t)$ and $k = \varepsilon d/t$, we can estimate $(D/k)$ to be $e^{O_{\varepsilon} \left( \frac{d}{\ell} \right)}$ by Shannon’s entropy estimate for binomial coefficients. Putting it all together, we get the following.

$$s' \geq e^{\delta k \ln d - O_{\varepsilon} \left( \frac{d}{t} \right) \ln d - \frac{1}{d\delta k}}$$

$$= e^{\delta k \ln d - \frac{\delta k}{10} \ln d - O_{\varepsilon} \left( \frac{d}{\ell} \right) - \frac{1}{d\delta k}}$$

$$= e^{0.9\varepsilon \frac{d}{\ell} \ln d - O_{\varepsilon} \left( \frac{d}{\ell} \right) - \frac{1}{d\delta k}}$$

$$= e^{O_{\delta, \varepsilon} \left( \frac{d}{\ell} \ln N \right)}.$$
This completes the proof of the main theorem. □

4. Lower bounds for explicit polynomials

In this section, we shall apply Theorem 1.2 to two explicit polynomials, \( NW_{n,r} \) which is a polynomial in VNP and \( IMM_{n,d} \) which is a polynomial in VP, to derive exponential lower bounds against the depth-4 \( \Sigma \Pi^{[O(\sqrt{n})]} \Sigma \Pi^{[\sqrt{n}]} \) circuits computing them.

4.1. Nisan–Wigderson polynomial. Now we derive the depth-4 circuit size lower bound for \( NW_{n,r} \) polynomial by a simple application of Theorem 1.2 where \( d = n \) and \( t = \sqrt{n} \).

**Corollary 4.1.** For \( 0 < \varepsilon < 1/80 \), any \( \Sigma \Pi^{[O(\sqrt{n})]} \Sigma \Pi^{[\sqrt{n}]} \) circuit computing the polynomial \( NW_{n,r}(X) \) must be of size \( 2^{\Omega(\sqrt{n} \log n)} \) where \( r = \varepsilon \sqrt{n} \).

**Proof.** Recall that \( NW_{n,r}(X) = \sum_{a(z) \in \mathbb{F}[z]} x_{a(1)} \cdots x_{a(n)} \) where \( \mathbb{F} \) is a finite field of size \( n \) and \( a(z) \) is a univariate polynomial of degree \( \leq r - 1 \). Notice that any two monomials can intersect in at most \( r - 1 \) variables.

Let us fix \( k = r = \varepsilon \sqrt{n} \). We differentiate the polynomial \( NW_{n,r}(X) \) with respect to the first \( k \) variables of each monomial. After differentiation, we get \( n^k \) monomials of length \( (n - k) \) each. Since they are constructed from the image of univariate polynomials of degree at most \( (k - 1) \), the distance \( \Delta \) between any two monomials is at least \( n - 2k \) and therefore greater than \( n/2 \). So to get the required lower bound, we invoke Theorem 1.2 with \( \delta = 1 \) and \( c = 2 \). □

4.2. Iterated matrix multiplication polynomial. Next we derive the lower bound on the size of the depth-4 circuit computing \( IMM_{n,n} \).

**Corollary 4.2.** Any depth-4 \( \Sigma \Pi^{[O(\sqrt{n})]} \Sigma \Pi^{[\sqrt{n}]} \) circuit computing the \( IMM_{n,n}(X) \) polynomial must be of size \( 2^{\Omega(\sqrt{n} \log n)} \).

**Proof.** Recall that \( IMM_{n,n}(X) \) is a polynomial over \( (n - 2)n^2 + 2n \) variables \( \{X_1, X_2, \ldots, X_n\} \). We fix the following lexicographic
ordering on the variables of the set of matrices \( \{X_1, X_2, \ldots, X_n\} \) as follows: \( X_1 > X_2 > X_3 > \cdots > X_n \) and in any \( X_i \), the ordering is \( x_{i,1}^{(i)} > x_{i,2}^{(i)} > \cdots > x_{i,n}^{(i)} \). Choose a prime \( p \) such that \( \frac{n}{2} \leq p \leq n \). Consider a set of \( n/4k \) distance apart. Clearly \( 2k+1+\frac{(2k-1)n}{4k} < n \).

For each univariate polynomial \( a \) of degree at most \( (k-1) \), define a set \( S_a \) as follows. For all \( i \neq j \in [n] \) and \( 2 \leq r \leq 2k \), set

\[
x_{i,j}^{(q)} = 0 \text{ iff } r + \frac{(r-2)n}{4k} < q < r + \frac{(r-1)n}{4k}.
\]

The rest of the variables are left untouched. Next we differentiate the polynomial \( f(X) \) with respect to the sets of variables \( S_a \) indexed by the polynomials \( a(z) \in \mathbb{F}_p[z] \). Consider the leading monomial of the derivatives with respect to the sets \( S_a \) for all \( a(z) \in \mathbb{F}_p[z] \). Since \( |S_a \cap S_b| < k \) for \( a \neq b \), it is straightforward to observe that the distance between any two leading monomials is at least \( k \cdot \frac{n}{4k} = \frac{n}{4} \). The intuitive justification is that whenever there is a difference in \( S_a \) and \( S_b \), that difference can be stretched to a distance \( \frac{n}{4k} \) because of the restriction that eliminates the non-diagonal entries.

Now we prove the lower bound for the polynomial \( f(X) \) by applying Theorem 1.2. Notice that \( f(X) \) is a \( n^{O(1)} \)-variate polynomial of degree \( n \) such that there are at least \( (n/2)^k > n^{\frac{1}{2}(2k)} \) different polynomials in \( \langle \partial = 2k(f) \rangle \) such that any two of their leading monomials have distance \( \Delta \geq n/4 \). So we set the parameters \( \delta = 1/4 \) and \( c = 4 \) in Theorem 1.2. A simple calculation shows that the parameter \( \varepsilon \) can be fixed to some value smaller than \( 1/320 \).
Since $f(X)$ is a restriction of $\text{IMM}_{n,n}(X)$, any lower bound for $f(X)$ is a lower bound for $\text{IMM}_{n,n}(X)$ too. Otherwise, if $\text{IMM}_{n,n}(X)$ has a $2^{o(\sqrt{n}\log n)}$-sized $\Sigma \Pi [O(\sqrt{n})] \Sigma \Pi [\sqrt{n}]$ circuit, then we get a $2^{o(\sqrt{n}\log n)}$-sized $\Sigma \Pi [O(\sqrt{n})] \Sigma \Pi [\sqrt{n}]$ circuit for $f(X)$ by substituting for the variables according to the restriction. □

5. Determinantal complexity lower bounds via the partial derivatives

Let us recall the following definition for the sake of completeness.

**Definition 5.1.** The dimension of the space of partial derivatives of a polynomial $f$ with respect to a parameter $k$ is defined as $\Gamma_k(f) := \text{dim}(\partial^k f)$.

If a polynomial $f = \text{Det}_m(A(X))$, then we need $\Gamma_k(\text{Det}(A(X)))$ must be equal to $\Gamma_k(f)$. We will now lower bound dimension of the derivative space of a polynomial $\text{Det}_m(A(X))$.

**Derivative space of $\text{Det}_m$ polynomial.** We will now lower bound the derivative space of $\text{Det}_m(A(X))$ polynomial where $A(X)$ is a $m \times m$ matrix whose entries are linear polynomials over $\mathbb{F}[X]$. Now consider the polynomial $\text{Det}_m(Y)$ over $\mathbb{F}[Y]$ where $Y = \{y_{11}, \ldots, y_{mm}\}$. By the chain rule of derivatives,

$$\frac{\partial \text{Det}_m(A(X))}{\partial x_{i,j}} = \sum_{s,t \in [m]} \frac{\partial \text{Det}_m(Y)}{\partial y_{s,t}} \bigg|_{Y \leftarrow A(X)} \cdot \frac{\partial (A(X))_{s,t}}{\partial x_{i,j}}.$$

Generalizing this, we get that

$$\frac{\partial^{=k} \text{Det}_m(A(X))}{\partial x_{i_1,j_1} \cdots \partial x_{i_k,j_k}} = \sum_{s_1, t_1 \in [m]; \ldots; s_k, t_k \in [k]} \frac{\partial^{=k} \text{Det}_m(Y)}{\partial y_{s_1,t_1} \cdots \partial y_{s_k,t_k}} \bigg|_{Y \leftarrow A(X)} \cdot \frac{\partial (A(X))_{s_1,t_1}}{\partial x_{i_1,j_1}} \cdots \frac{\partial (A(X))_{s_k,t_k}}{\partial x_{i_k,j_k}}.$$

Since the entries of $A(X)$ are linear polynomials, $\frac{\partial (A(X))_{s,t}}{\partial x_{i,j}}$ is a constant for all $i, j, s, t$. We get that $\frac{\partial^{=k} \text{Det}_m(A(X))}{\partial x_{i_1,j_1} \cdots \partial x_{i_k,j_k}}$ is in
\[
\mathbb{F}\text{-span} \left\{ \frac{\partial^{=k} \operatorname{Det}_m(Y)}{\partial y_{s_1,t_1} \cdots \partial y_{s_k,t_k}} \bigg| Y \leftarrow A(X), s_p, t_p \in [m] \text{ and } p \in [k] \right\}
\]
and \( \partial^{=k} \operatorname{Det}_m(A(X)) \) is a subset of
\[
\mathbb{F}\text{-span} \left\{ \frac{\partial^{=k} \operatorname{Det}_m(Y)}{\partial y_{s_1,t_1} \cdots \partial y_{s_k,t_k}} \bigg| Y \leftarrow A(X), s_p, t_p \in [m] \text{ and } p \in [k] \right\}.
\]
and thus \( \dim(\partial^{=k} \operatorname{Det}_m(A(X))) \) is at most \( \dim((\partial^{=k} \operatorname{Det}_m(Y))) \).

We note the following simple property of the derivative space of the determinant polynomial. This follows from the fact that a \( k \)th-order derivative corresponds to a minor of the order \( (n-k) \) and any two distinct minors do not share a monomial.\(^3\)

**Proposition 5.2.** For any \( k \), \( \Gamma_k(\operatorname{Det}_m(Y)) = \binom{m}{k}^2 \).

Invoking Proposition 5.2 and from the discussion above, we get that
\[
\Gamma_k(\operatorname{Det}_m(A(X))) \leq \Gamma_k(\operatorname{Det}_m(Y)) = \binom{m}{k}^2.
\]

**Derivative space of Nisan–Wigderson polynomial.** Let us recall that \( \text{NW}_{n,\varepsilon n}(X) = \sum_{a(z) \in \mathbb{F}[z]} x_{1,a(1)} x_{2,a(2)} \ldots x_{n,a(n)} \) where \( \mathbb{F} \) is a finite field of size \( n \) and \( a(z) \) is a univariate polynomial of degree \( < \varepsilon n \) where \( \varepsilon \in (0, 0.5) \). Notice that any two of its monomials can intersect in at most \( \varepsilon n - 1 \) variables. We now differentiate the polynomial \( \text{NW}_{n,\varepsilon n}(X) \) with respect to the first \( k = \varepsilon n \) variables of every monomial. After the differentiation, we get \( n^{\varepsilon n} \) distinct monomials each of which is of length \((1-\varepsilon)n\). Thus, \( \Gamma_{\varepsilon n}(\text{NW}_{n,\varepsilon n}) \geq n^{\varepsilon n} \).

**Theorem 5.3.** For any \( \varepsilon \in (0, 0.5) \), it is true that \( \operatorname{DetComp}(\text{NW}_{n,\varepsilon n}) \geq \Omega(n^{1.5}) \). This holds over any field.

\(^3\) A much stronger statement about the determinantal ideal can be found in (Theorem 22) Gupta et al. (2013) and the references therein.
Proof. If \( \text{NW}_{n,\varepsilon n}(X) = \text{Det}_m(A(X)) \) for some \( m \times m \) matrix \( A(X) \), then the dimension of the partial derivative space of the \( \text{NW}_{n,\varepsilon n}(X) \) polynomial of order \( k \) must be equal to the dimension of the partial derivative space of the \( \text{Det}_m(A(X)) \) polynomial of the same order. Thus for \( k = \varepsilon n \),

\[
\Gamma_k(\text{Det}_m(A(X))) = \Gamma_k(\text{NW}_{n,\varepsilon n}(X))
\]

\[
\left(\frac{m}{k}\right)^2 \geq n^{\varepsilon n}
\]

\[
\left(\frac{e \cdot m}{k}\right)^{2k} \geq n^{\varepsilon n}
\]

\[
m \geq \frac{\varepsilon n \cdot \sqrt{n}}{e} = \Omega(n^{1.5})
\]

Thus, \( m \) has to be at least \( \Omega(n^{1.5}) \) for the \( \text{NW}_{n,\varepsilon n}(X) \) polynomial to be written as an affine projection of the \( \text{Det}_m \) polynomial, that is as \( \text{Det}_m(A_{k,\ell}) \) where \( A_{k,\ell}, 1 \leq k, \ell \leq m \) are linear polynomials in \( \mathbb{F}[X] \). \( \square \)

Derivative space of the iterated matrix multiplication polynomial. The iterated matrix multiplication polynomial is defined over the disjoint sets of variables \( X_1, X_2, \ldots, X_d \).

\[
\text{IMM}_{n,d}(X) = \sum_{i_1, i_2, \ldots, i_{n-1} \in [n]} x_1^{(1)} x_{i_1, i_2}^{(2)} \cdots x_{i(d-2), i(d-1)}^{(d)} x_{i(d-1), i(1)}^{(d-1)}.
\]

We will lower bound \( \Gamma_k(\text{IMM}_{n,d}(X)) \) by the dimension of a specific subspace of the derivative space of \( \text{IMM}_{n,d}(X) \). That is, dimension of the entire derivative space is lower bounded by the dimension of the subspace that we will now consider. Let a subset \( J = \{j_1, j_2, \ldots, j_k\} \subseteq [d] \), be called a good set of size \( k \) if it is an ordered set with \( k \) elements \( j_1 < j_2 < \cdots < j_k \) where \( j_1 \geq 1, j_k \leq d \) and \( |j_{s+1} - j_s| \geq 2 \) for all \( s \in [k-1] \). For any good set \( J \) of size \( k \), consider the corresponding sets of variables \( X_{j_1}, X_{j_2}, \ldots, X_{j_k} \). Let \( M_J \) be a set of set-multilinear\(^4\) monomials of degree exactly equal

\(^4\) A monomial \( m \) of degree \( d \) is said to be set-multilinear with respect to the disjoint variable sets \( X_1, X_2, \ldots, X_d \) if \( |m \cap X_i| = 1 \) for all \( i \in [d] \).
to $k$ defined over $X_{j_1}, X_{j_2}, \ldots, X_{j_k}$. Let $M$ be a set of monomials of degree $k$, which is defined as follows.

$$M = \sum_{J \text{ is a good set of size } k} M_J$$

It is easy to see that the partial derivatives of $\text{IMM}_{n,d}$ with respect to the monomials in $M$ are pairwise distinct. The number of good sets of size $k$ is equal to $\binom{d-k+1}{k}$, and the number of monomials in $M_J$ is $n^{2k}$. Thus,

$$\Gamma_k(\text{IMM}_{n,d}(X)) \geq \binom{d-k+1}{k} \cdot n^{2k}.$$

Since $\Gamma_k(\text{IMM}_{n,d}(X)) = \Gamma_k(\text{Det}_m(A(X)))$, for $k = \delta d$ for a suitable $\delta \in (0, 1)$, we get the following.

$$\left(\frac{m}{k}\right)^2 \geq \binom{d-k+1}{k} \cdot n^{2k}$$

$$\left(\frac{e \cdot m}{k}\right)^{2k} \geq \frac{d-k+1}{k} \cdot n^{2k}$$

$$\frac{e \cdot m}{k} \geq \sqrt{\frac{d-k+1}{k}} \cdot n$$

$$m \geq e^{-1} n \cdot \sqrt{k(d-k+1)} = \Omega(dn).$$

We will improve on this result by a constant factor in Section 6.

6. Determinantal complexity lower bounds via the Hessian

Determinantal complexity of $\text{IMM}_{n,d}$ via the approach of Mignon and Ressayre. We start by recalling a few facts from Cai et al. (2008). Let $\text{IMM}_{n,d}$ be the target polynomial over $N = n^2 d$ variables. Let $A_{k,\ell}(X)$, $1 \leq k, \ell \leq m$ be the affine linear polynomials over $\mathbb{F}[X]$ such that $\text{IMM}_{n,d}(X) = \det((A_{k,\ell}(X))_{1 \leq k, \ell \leq m})$. Consider a point $X_0 \in \mathbb{F}^N$ such that $\text{IMM}_{n,d}(X_0) = 0$. The affine linear functions $A_{k,\ell}(X)$ can be expressed as $L_{k,\ell}(X - X_0) + y_{k,\ell}$ where $L_{k,\ell}$ is a linear form and $y_{k,\ell}$ is a constant from the
field. Thus, \((A_{k,\ell}(X))_{1\leq k,\ell\leq m} = (L_{k,\ell}(X - X_0))_{1\leq k,\ell\leq m} + Y_0\). If \(\text{IMM}_{n,d}(X_0) = 0\), then \(\det(Y_0) = 0\). Let \(C\) and \(D\) be two non-singular matrices such that \(CY_0D\) is a diagonal matrix.

\[
CY_0D = \begin{pmatrix}
0 & 0 \\
0 & I_s
\end{pmatrix}
\]

Since \(\det(Y_0) = 0\), \(s < m\). As in Cai et al. (2008) we too consider the case where \(s = m - 1\). Since the first row and the first column of \(CY_0D\) are zero, we may multiply \(CY_0D\) by \(\text{diag}(\det(C)^{-1}, 1, \ldots, 1)\) and \(\text{diag}(\det(D)^{-1}, 1, \ldots, 1)\) on the left and the right sides. Without loss of generality, we may assume that \(\det(C) = \det(D) = 1\). By multiplying with \(C\) and \(D\) on the left and the right and by suitably renaming \((L_{k,\ell}(X - X_0))_{1\leq k,\ell\leq m}\) and \(Y_0\), we get that

\[
\text{IMM}_{n,d}(X) = \det((L_{k,\ell}(X - X_0))_{1\leq k,\ell\leq m} + Y_0))
\]

where \(Y_0 = \text{diag}(0, 1, \ldots, 1)\).

We use \(\text{Hess}_{\text{IMM}_{n,d}}(X)\) to denote the Hessian matrix of the polynomial \(\text{IMM}_{n,d}\) which is defined as follows.

\[
\text{Hess}_{\text{IMM}_{n,d}}(X) = (H_{s;ij,t;kl}(X))_{1\leq i,j,k,\ell\leq n, 1\leq s,t\leq d}
\]

where

\[
H_{s;ij,t;kl}(X) = \frac{\partial^2 \text{IMM}_{n,d}(X)}{\partial x_{ij}^{(s)} \partial x_{k,\ell}^{(t)}}.
\]

where \(x_{ij}^{(s)}\) and \(x_{k,\ell}^{(t)}\) denote the \((i,j)\)th and \((k,\ell)\)th entries of the variable sets \(X_s\) and \(X_t\), respectively.

By taking second-order derivatives and evaluating the Hessian matrices of \(\text{IMM}_{n,d}(X)\) and \(\det((A_{k,\ell}(X))_{1\leq k,\ell\leq m})\) at \(X_0\), we obtain \(\text{Hess}_{\text{IMM}_{n,d}}(X_0) = L\text{Hess}_{\det}(Y_0)L^T\) where \(L\) is a \(N \times m^2\) matrix with entries from the field. It follows that \(\text{rank}(\text{Hess}_{\text{IMM}_{n,d}}(X_0)) \leq \text{rank}(\text{Hess}_{\det}(Y_0))\). It was observed in the earlier work of Mignon & Ressayre (2004) and Cai et al. (2008) that it is relatively easy to get an upper bound for \(\text{rank}(\text{Hess}_{\det}(Y_0))\). The main task is to construct a point \(X_0\) such that \(f(X_0) = 0\), yet the rank of \(\text{Hess}_f(X_0)\) is high.

Before we give an explicit construction of a point \(X_0 \in \mathbb{F}^{n^2d}\) such that \(\text{IMM}_{n,d}(X_0) = 0\) and \(\text{rank}(\text{Hess}_{\text{IMM}_{n,d}}(X_0)) \geq d(n-1)\), we briefly recall the upper bound argument for the rank of \(\text{Hess}_{\det}(Y_0)\) from Section 2.1 of Cai et al. (2008), for the sake of completeness.
Upper bound for the rank of $\text{Hess}_{\det}(Y_0)$. When we take a partial derivative of the determinant polynomial with respect to the variable $y_{i,j}$, the result is a minor that is obtained by striking out the row $i$ and column $j$. The second-order derivative of $\det(Y)$ with respect to the variables $y_{i,j}$ and $y_{k,\ell}$ eliminates the rows $\{i, k\}$ and the columns $\{j, \ell\}$. Considering the form of $Y_0$, the nonzero entries in $\text{Hess}_{\det}(Y_0)$ are obtained only if $1 \in \{i, k\}$ and $1 \in \{j, \ell\}$ and thus $(ij, k\ell)$ are of the form $(11, tt)$, $(tt, 11)$, $(t1, 1t)$ or $(1t, t1)$ for any $t > 1$. Thus, $\text{rank}(\text{Hess}_{\det}(Y_0))$ is at most $2m$.

Lower bound for the rank of $\text{Hess}_{\text{IMM}_{n,d}}(X_0)$. In this section, we shall prove Theorem 1.4. In particular, we give a polynomial time algorithm to construct a point $X_0$ explicitly such that $\text{IMM}_{n,d}(X_0) = 0$ and $\text{rank}(\text{Hess}_{\text{IMM}_{n,d}}(X_0)) \geq d(n - 1)$. Since $\text{rank}(\text{Hess}_{\det}(Y_0)) = 2m$ and $\text{rank}(\text{Hess}_{\text{IMM}_{n,d}}(X_0)) \leq \text{rank}(\text{Hess}_{\det}(Y_0))$, we get that $m \geq d(n - 1)/2$. As mentioned in Introduction, the determinantal complexity of $\text{IMM}_{n,d}(X)$ is $O(dn)$. Together, it implies that $m = \Theta(dn)$.

**Theorem 6.1.** For any integers $n, d > 1$, there is a point $X_0 \in \mathbb{F}^{n^2d}$ such that $\text{IMM}_{n,d}(X_0) = 0$ and $\text{rank}(\text{Hess}_{\text{IMM}_{n,d}}(X_0)) \geq d(n - 1)$. Moreover, the point $X_0$ can be constructed explicitly in polynomial time.

**Proof.** We prove the theorem by induction on $d$, the degree of the polynomial. For the purpose of induction, we ensure that the entries indexed by the indices $(1, 2), (1, 3), \ldots, (1, n)$ of the matrix obtained after multiplying the first $(d - 1)$ matrices are not all zero at $X_0$.

We first prove the base case for $d = 2$. The corresponding polynomial is $\text{IMM}_{n,2}(X) = \sum_{i=1}^{n} x_{1,i}^{(1)} x_{i,1}^{(2)}$. It is easy to observe that the rank of the corresponding Hessian matrix is $2n > 2(n - 1)$ at any point since each nonzero entry of the Hessian matrix is 1 and the structure of the Hessian matrix is the following:

$$\text{Hess}_{\text{IMM}_{n,2}}(X) = \begin{bmatrix} 0 & B_{1,2} \\ B_{2,1} & 0 \end{bmatrix}$$
where $B_{2,1} = B_{1,2}^T$ and the matrix $B_{1,2}$ is formally described as

$$(B_{1,2})_{x^{(1)}_{i,j}, x^{(2)}_{k,l}} = \begin{cases} 1 & \text{if } i = l = 1 \text{ and } j = k \\ 0 & \text{otherwise.} \end{cases}$$

We set the values of the variables as follows: $x^{(1)}_{1,1} = 0$, $x^{(2)}_{1,1} = 1$, $x^{(2)}_{2,1} = x^{(2)}_{3,1} = \cdots = x^{(2)}_{n,1} = 0$ and $x^{(1)}_{1,2}, x^{(1)}_{1,3}, \ldots, x^{(1)}_{1,n}$ to arbitrary values but not all to zero. The point thus obtained (say $X_0$) is clearly a zero of the polynomial $\text{IMM}_{n,2}(X)$.

For induction hypothesis, assume that the statement of the theorem is true for the case where the number of matrices being multiplied is $\leq d$. Consider the polynomial $\text{IMM}_{n,d+1}(X)$.

$$\text{IMM}_{n,(d+1)}(X) = \sum_{i_1, i_2, \ldots, i_{d-1}, i_d \in [n]} x^{(1)}_{i_1, i_2} x^{(2)}_{i_1, i_2} \cdots x^{(d-1)}_{i_1, i_2} x^{(d)}_{i_1, i_2} x^{(d+1)}_{i_1, i_2}$$

Let the matrix obtained after multiplying the first $d$ matrices be $(P_{k,l})_{(k,l)\in [n] \times [n]}$ where

$$P_{k,l}(X) = \sum_{i_1, i_2, \ldots, i_{d-1} \in [n]} x^{(1)}_{k, i_1} x^{(2)}_{i_1, i_2} \cdots x^{(d-1)}_{i_1, i_2} x^{(d)}_{i_1, i_2} x^{(d+1)}_{i_1, i_2}$$

for $1 \leq k, l \leq n$.

Thus, we have the following expression.

$$\text{IMM}_{n,(d+1)}(X) = P_{1,1}(X)x^{(d+1)}_{1,1} + P_{1,2}(X)x^{(d+1)}_{2,1} + \cdots + P_{1,n}(X)x^{(d+1)}_{n,1}$$

Now consider the Hessian matrix $\text{Hess}_{\text{IMM}_{n,d+1}}(X)$ which is a $(d+1)n^2 \times (d+1)n^2$-sized matrix.

$$\text{Hess}_{\text{IMM}_{n,d+1}}(X) = \begin{bmatrix} 0 & B_{1,2} & B_{1,3} & B_{1,4} & \cdots & B_{1,(d+1)} \\ B_{2,1} & 0 & B_{2,3} & B_{2,4} & \cdots & B_{2,(d+1)} \\ B_{3,1} & B_{3,2} & 0 & B_{3,4} & \cdots & B_{3,(d+1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ B_{(d+1),1} & B_{(d+1),2} & \cdots & \cdots & B_{(d+1),d} & 0 \end{bmatrix}$$

Each $B_{i,j}$ is a block of size $n^2 \times n^2$ and is indexed by the variable sets $X_i$ and $X_j$, respectively. Consider the block $B_{(d+1),d}$ which is
indexed by the variable sets $X_{d+1}$ and $X_d$. The only nonzero rows in $B_{(d+1),d}$ are indexed by the variables $x_{1,1}^{(d+1)}, x_{2,1}^{(d+1)}, \ldots, x_{n,1}^{(d+1)}$. The potential nonzero entries for the row $x_{1,1}^{(d+1)}$ are indexed by the columns $x_{1,1}^{(d)}, x_{2,1}^{(d)}, \ldots, x_{n,1}^{(d)}$. Similarly the potential nonzero entries for the row $x_{2,1}^{(d+1)}$ are indexed by the columns $x_{1,2}^{(d)}, x_{2,2}^{(d)}, \ldots, x_{n,2}^{(d)}$ and so on.

Consider the entries corresponding to the indices $(x_{1,1}^{(d+1)}, x_{1,1}^{(d)})$, $(x_{1,1}^{(d+1)}, x_{2,1}^{(d)})$, $\ldots$, $(x_{1,1}^{(d+1)}, x_{n,1}^{(d)})$, say $Q_1, Q_2, \ldots, Q_n$, respectively, where

$$Q_j = \sum_{i_1,i_2,\ldots,i_{d-2} \in [n]} x_{1,i_1}^{(1)} x_{1,i_2}^{(2)} \cdots x_{i,(d-2),j}^{(d-1)} \text{ for } 1 \leq j \leq n.$$ 

For the other rows indexed by the variables $x_{2,1}^{(d+1)}, x_{3,1}^{(d+1)}, \ldots, x_{n,1}^{(d+1)}$, the sequence of potential nonzero entries is the same ($Q_1, Q_2, \ldots, Q_n$), but their positions are shifted by a column compared to the previous nonzero row. Formally,

$$(B_{(d+1),d})_{x_{i,j}^{(d+1)},x_{k,\ell}^{(d)}} = \begin{cases} Q_k & \text{if } j = 1, \ell = i, \text{ and } i, k \in [n] \\ 0 & \text{otherwise.} \end{cases}$$

$Q_1, Q_2, \ldots, Q_n$ are also the entries indexed by the indices $(1,1), (1,2), \ldots, (1,n)$ of the matrix obtained after multiplying the first $(d - 1)$ matrices. By induction hypothesis, we know that the entries indexed by the indices $(1,2), \ldots, (1,n)$ are not all zero at the point $X_0$, the zero of the polynomial \text{IMM}_{n,d}(X).$ Consider the rows indexed by the variables $x_{1,1}^{(d+1)}, x_{2,1}^{(d+1)}, \ldots, x_{n,1}^{(d+1)}$. The row supports for rows indexed by these variables in $B_{(d+1),d}$ are disjoint, and at least one of the entries per row is nonzero. Thus, the rows indexed by the variables $x_{1,1}^{(d+1)}, x_{2,1}^{(d+1)}, \ldots, x_{n,1}^{(d+1)}$ are linearly independent. It is important to note that $P_{1,1}(X)$ is an instance of \text{IMM}_{n,d}(X) over $X_1, X_2, \ldots, X_d$.

Let $X_0$ be the zero of the polynomial \text{IMM}_{n,d}(X).$ Now, we extend $X_0$ and construct a new point inductively. We first fix the variables appearing in $P_{1,1}(X)$ by the values assigned by $X_0$. We set $x_{1,1}^{(d+1)} = 1$ and $x_{2,1}^{(d+1)} = x_{3,1}^{(d+1)} = \cdots = x_{n,1}^{(d+1)} = 0$. We will fix
the rest of the variables later. We call the new point which is a zero of the polynomial $\text{IMM}_{n,(d+1)}(X)$, as $X_0$ as well.

Now, consider the first $d \times d$ blocks of the Hessian matrix $\text{Hess}_{\text{IMM}_{n,(d+1)}}(X_0)$. By induction hypothesis, the rank of this minor of $\text{Hess}_{\text{IMM}_{n,(d+1)}}(X_0)$ is at least $d(n - 1)$. These $d \times d$ blocks precisely represent the Hessian matrix of $P_{1,1}(X)$, which is also the Hessian matrix of the polynomial $\text{IMM}_{n,d}(X)$ at the point $X_0$. The only nonzero entries in the columns indexed by the variable set $X^{(d)}$ are indexed by the variables $x^{(d)}_{1,1}, x^{(d)}_{2,1}, \ldots, x^{(d)}_{n,1}$. This is because the other variables of $X_d$ do not appear in $\text{IMM}_{n,d}(X)$ and the setting of the variables $x^{(d+1)}_{1,1} = 1$ and $x^{(d+1)}_{2,1} = x^{(d+1)}_{3,1} = \cdots = x^{(d+1)}_{n,1} = 0$.

The row in $B_{(d+1),d}$ indexed by $x^{(d+1)}_{1,1}$ is the only row that interferes with any of the rows of $B_{1,d}, B_{2,d}, \ldots, B_{d,d}$. The rows indexed by the variables $x^{(d+1)}_{2,1}, x^{(d+1)}_{3,1}, \ldots, x^{(d+1)}_{n,1}$ in $B_{(d+1),d}$ are linearly independent of the rows of $B_{1,d}, B_{2,d}, \ldots, B_{d,d}$ due to aforementioned reasons. These rows contribute $n - 1$ to the rank and hence extend the rank of $\text{Hess}_{\text{IMM}_{n,(d+1)}}$ at the point described to $\geq (d+1)(n - 1)$.

For the purpose of induction, we must verify that the entries indexed by the indices $(1, 2), (1, 3), \ldots, (1, n)$ of the matrix obtained after multiplying the first $d$ matrices are not all zero at $X_0$. These entries are the polynomials $P_{1,2}, P_{1,3}, \ldots, P_{1,n}$. We shall express each of the polynomials in terms of $Q_1, Q_2, \ldots, Q_n$ as follows.

$$P_{1j} = Q_1 x^{(d)}_{1,j} + Q_2 x^{(d)}_{2,j} + \cdots + Q_n x^{(d)}_{n,j} \quad \text{for } 2 \leq j \leq n.$$

By induction hypothesis, we already know that $Q_2, Q_3, \ldots, Q_n$ are not all zero at $X_0$. Notice that the variables in $X^{(d)} \backslash \{x^{(d)}_{1,1}, x^{(d)}_{2,1}, \ldots, x^{(d)}_{n,1}\}$ were never set in the previous steps of induction. This is because of the fact that they do not appear in the polynomial $P_{1,1}$. Therefore, we can fix these variables suitably such that $P_{1,2}, P_{1,3}, \ldots, P_{1,n}$ are not all zero when evaluated at the point $X_0$. (In fact, we can make all of them nonzero.) It is clear that we construct the point $X_0$ in polynomial time. This completes the proof. □
7. Formula-size lower bound for $\text{IMM}_{n,d}$

In this section, we shall prove a super-linear but subquadratic lower bound on the size of any formula that computes the $\text{IMM}_{n,d}$ polynomial. The following proof is an adaptation of the proof strategy of Kalorkoti (1985). Let us first recall the notion of algebraic independence and transcendence degree.

**Definition 7.1.** A set of polynomials $f_1, f_2, \ldots, f_m \in \mathbb{F}[X]$ are said to be algebraically independent if the only polynomial $F \in \mathbb{F}[y_1, y_2, \ldots, y_m]$ satisfying $F(f_1, f_2, \ldots, f_m) \equiv 0$ is the zero polynomial.

The transcendental degree of the polynomials $f_1, f_2, \ldots, f_m \in \mathbb{F}[X]$, denoted by $\text{trdeg}(f_1, f_2, \ldots, f_m)$, is the maximal size of the subset $S$ of $[m]$ such that $\{f_i\}_{i \in S}$ are algebraically independent.

We shall now define the notion of transcendental degree of a polynomial with respect to a subset of its variables.

**Definition 7.2.** Let $f \in \mathbb{F}[X]$ be a polynomial and $X' \subset X$ a set of variables. Let $f$ be expressed as $\sum_{m \in M} f_m \cdot m$ where $M$ is set of all monomials over the variables in $X'$ and degree at most $\deg(f)$. The complexity measure $\text{trdeg}_{X'}(f)$ is defined as the transcendental degree of $\{f_m\}_{m \in M}$.

The following lemma is the key to the formula size lower bound in Kalorkoti (1985) (cf. Saptharishi 2015).

**Lemma 7.3.** Let $f \in \mathbb{F}[X]$ and $X_1, X_2, \ldots, X_t$ be a partition of $X$. Then, every arithmetic formula for $f$ must be of size $\Omega(\sum_{i \in [t]} \text{trdeg}_{X_i}(f))$.

**Theorem 7.4.** For all integers $n, d > 0$, any arithmetic formula computing the $\text{IMM}_{n,d}(X)$ polynomial must be of size $\Omega(dn^3)$.

**Proof.** The main idea is to find a suitable partition of the input variables. For simplicity, we assume that $d$ is a multiple of four. Let $M_1, M_2, \ldots, M_d$ be the generic $n \times n$ matrices being multiplied in $\text{IMM}_{n,d}(X)$ polynomial. For all $i$ such that $i$ is of the form $4t + 1$
or $4t + 2$, $t \in [0, d/4 - 1]$, partition the variables in the matrices $M_t$ and $M_{i+2}$ by grouping the $j$th row of $M_t$ and $j$th column of $M_{i+2}$ together, for all $j \in [n]$. We shall denote such a set by $X_{i,j} = \{x^{(i)}_{j,1}, \ldots, x^{(i)}_{j,n}, x^{(i+2)}_{1,1}, \ldots, x^{(i+2)}_{n,1}\}$. The final partition of the variables is as follows.

$$X = \bigsqcup_{i \in \{4t+1, 4t+2:t \in [0,d/4-1]\}} \bigsqcup_{1 \leq j \leq n} X_{i,j}.$$  

Now we express the polynomial $\text{IMM}_{n,d}(X)$ w.r.t the set of variables $X_{i,j}$ as explained in definition Definition 7.2.

$$\text{IMM}_{n,d}(X) = \sum_{k,\ell \in [n]} (x^{(i+1)}_{k,\ell} P_1) \cdot x^{(i)}_{j,k} x^{(i+2)}_{\ell,j} + P_2.$$  

The first summand in the above expression is the summation of all monomials that contain the variables $x^{(i)}_{j,k}$ and $x^{(i+2)}_{\ell,j}$ for all $k, \ell \in [n]$ and $P_2$ is the summation of the rest of the monomials. Formally,

$$P_1(X) = \sum_{a_i \in [n]} x^{(1)}_{1,a_1} \cdots x^{(i-1)}_{a_{i-2},j} x^{(i+3)}_{j,a_i+3} \cdots x^{(d)}_{a_{d-1},1}.$$  

Now, $\text{trdeg}_{X_{i,j}}(\text{IMM}_{n,d}(X))$ is at least the transcendental degree of the set of polynomial $P = \{P_1 \cdot x^{(i+1)}_{k,\ell}\}_{k,\ell \in [n]}$. Notice that $|P| = n^2$. Let us introduce new variables $Y = \{y_1, \ldots, y_{n^2}\}$ to lexicographically correspond to polynomials in $P = \{P_1 \cdot x^{(i+1)}_{k,\ell}\}_{k,\ell \in [n]}$. To prove their algebraic independence, we need to prove that there is no nonzero polynomial over $F[Y]$ such that substitution for $y_i$ with the corresponding polynomials in $P$ makes it a zero polynomial over $F[X]$.

For the sake of contradiction, let us assume that there is a polynomial $g \in F[Y]$ that annihilates the polynomials in $P$. Consider two distinct monomials $m_1 = y^{\alpha_1}_1 y^{\alpha_2}_2 \cdots y^{\alpha_{n^2}}_{n^2}$ and $m_2 = y^{\beta_1}_1 y^{\beta_2}_2 \cdots y^{\beta_{n^2}}_{n^2}$ in $g$ such that $\alpha \neq \beta$. Consider $m'_1 = m_1 |_{y_i \leftarrow P_i \in P}$ and $m'_2 = m_2 |_{y_i \leftarrow P_i \in P}$. We can see that $m'_1 = P_1(\sum_{r \in [n^2]} \alpha_r) \left( x^{(i+1)}_{1,1} \right)^{\alpha_1} \left( x^{(i+1)}_{1,2} \right)^{\alpha_2} \cdots \left( x^{(i+1)}_{n,n} \right)^{\alpha_{n^2}}$ and $m'_2 = P_1(\sum_{r \in [n^2]} \beta_r) \left( x^{(i+1)}_{1,1} \right)^{\beta_1} \left( x^{(i+1)}_{1,2} \right)^{\beta_2} \cdots \left( x^{(i+1)}_{n,n} \right)^{\beta_{n^2}}$.
W.l.o.g, let us assume that $\alpha_1 > \beta_1$. The overall degree of $x_{1,1}^{(i)}$ in $m'_1$ is equal to $\alpha_1$, and similarly the overall degree of the variable $x_{1,1}^{(i)}$ in $m'_2$ is equal to $\beta_1$, and hence, the monomials in $m'_1$ and $m'_2$ are distinct. So, one can conclude that no two distinct monomials in $g$ can share a monomial after the substitution. Hence, the polynomial $g$ cannot annihilate the polynomials in $P$. From Lemma 7.3, we get that the size of any arithmetic formula computing $\text{IMM}_{n,d}(X)$ is of size at least $\sum_{i,j} \text{trdeg}_{X_{i,j}}(\text{IMM}_{n,d}(X)) = \Omega(dn^3)$.

\qed

Acknowledgements

This work was done when Suryajith Chillara was a graduate student at Chennai Mathematical Institute. Suryajith Chillara was supervised by Partha Mukhopadhyay and was supported by TCS research fellowship. We thank the anonymous reviewers for their invaluable feedback that helped improve the paper and take the current form.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

MANINDRA AGRAWAL & V VINAY (2008). Arithmetic circuits: A chasm at depth four. In Proceedings of Foundations of Computer Science (FOCS), 67–75. IEEE.

JIN-YI CAI (1990). A note on the determinant and permanent problem. Information and Computation 84(1), 119–127.

JIN-YI CAI, XI CHEN & DONG LI (2008). A quadratic lower bound for the permanent and determinant problem over any characteristic $\neq 2$. In Proceedings of Symposium on Theory of Computing, 491–498. ACM.

SURYAJITH CHILLARA, MRINAL KUMAR, RAMPRASAD SAPTHARISHI & V. VINAY (2016). The Chasm at Depth Four, and Tensor Rank: Old results, new insights. Electronic Colloquium on Computational Complex-
ity (ECCC) 23, 96. URL http://eccc.hpi-web.de/report/2016/096.

David Cox, John Little & Donal O’shea (2007). Ideals, varieties, and algorithms, volume 3. Springer.

Hervé Fournier, Nutan Limaye, Guillaume Malod & Srikanth Srinivasan (2014). Lower bounds for depth 4 formulas computing iterated matrix multiplication. In Proceedings of Symposium on Theory of Computing, 128–135. ACM. URL http://doi.acm.org/10.1145/2591796.2591824.

Joachim von zur Gathen (1986). Permanent and Determinant. In Proceedings of Foundations of Computer Science (FOCS), 398–401. IEEE Computer Society.

Joachim von zur Gathen (1987). Permanent and determinant. Linear Algebra and its Applications 96, 87–100.

Ankit Gupta, Pritish Kamath, Neeraj Kayal & Ramprasad Saptharishi (2013). Approaching the chasm at depth four. In Proceedings of the Conference on Computational Complexity (CCC).

Maurice Jansen (2011). Lower Bounds for the Determinantal Complexity of Explicit Low Degree Polynomials. Theory of Computing Systems 49(2), 343–354.

Kyriakos Kalorkoti (1985). A Lower Bound for the Formula Size of Rational Functions. SIAM Journal of Computing 14(3), 678–687.

Neeraj Kayal (2012). An exponential lower bound for the sum of powers of bounded degree polynomials. Electronic Colloquium on Computational Complexity (ECCC) 19, 81. URL http://eccc.hpi-web.de/report/2012/081.

Neeraj Kayal, Nutan Limaye, Chandan Saha & Srikanth Srinivasan (2014a). An Exponential Lower Bound for Homogeneous Depth Four Arithmetic Circuits. In Proceedings of Foundations of Computer Science (FOCS). IEEE.

Neeraj Kayal, Chandan Saha & Ramprasad Saptharishi (2014b). A super-polynomial lower bound for regular arithmetic formulas. In Proceedings of Symposium on Theory of Computing, 146–153. ACM. URL http://doi.acm.org/10.1145/2591796.2591847.
Pascal Koiran (2012). Arithmetic circuits: The chasm at depth four gets wider. *Theor. Comput. Sci.* **448**, 56–65.

Mrinal Kumar & Shubhangi Saraf (2014a). The limits of depth reduction for arithmetic formulas: it’s all about the top fan-in. In *Proceedings of Symposium on Theory of Computing*, 136–145. ACM. URL http://doi.acm.org/10.1145/2591796.2591827.

Mrinal Kumar & Shubhangi Saraf (2014b). On the power of homogeneous depth 4 arithmetic circuits. In *Proceedings of Foundations of Computer Science (FOCS)*. IEEE.

Mrinal Kumar & Shubhangi Saraf (2016a). Arithmetic circuits with locally low algebraic rank. In *Proceedings of Conference on Computational Complexity (CCC)*.

Mrinal Kumar & Shubhangi Saraf (2016b). Sums of Products of Polynomials in Few Variables: Lower Bounds and Polynomial Identity Testing. In *Proceedings of Conference on Computational Complexity (CCC)*.

Roy Meshulam (1989). On two extremal matrix problems. *Linear Algebra and its Applications* **114**, 261–271.

Thierry Mignon & Nicolas Ressayre (2004). A quadratic bound for the determinant and permanent problem. *International Mathematics Research Notices* **2004**(79), 4241–4253.

Noam Nisan & Avi Wigderson (1994). Hardness vs Randomness. *J. Comput. Syst. Sci.* **49**(2), 149–167.

Ramprasad Saptharishi (2015). A survey of lower bounds in arithmetic circuit complexity. URL https://github.com/dasarpmar/lowerbounds-survey/releases/. Github survey.

Sébastien Tavenas (2013). Improved Bounds for Reduction to Depth 4 and Depth 3. In *Proceedings of Mathematical Foundations of Computer Science (MFCS)*, 813–824.

Seinosuke Toda (1992). Classes of arithmetic circuits capturing the complexity of computing the determinant. *IEICE Transactions on Information and Systems* **75**(1), 116–124.
Leslie G Valiant (1979). Completeness classes in algebra. In Proceedings of Symposium on Theory of Computing (STOC), 249–261. ACM.

Manuscript received 3 December 2017

Suryajith Chillara
Department of CSE
IIT Bombay
Mumbai, India
suryajith@cse.iitb.ac.in

Partha Mukhopadhyay
Chennai Mathematical Institute
Siruseri, India
partham@cmi.ac.in