Depth-4 Lower Bounds, Determinantal Complexity : A Unified Approach

Suryajith Chillara[*]
Chennai Mathematical Institute
suryajith@cse.iitb.ac.in

Partha Mukhopadhyay
Chennai Mathematical Institute.
partham@cmi.ac.in

November 5, 2018

Abstract

Tavenas 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 $2^{O(\sqrt{n}\log n)}$ [Tav13]. So, to prove VP $\neq$ 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 [KSS14] and [FLMS14]). In particular, using combinatorial design Kayal et al. [KSS14] construct an explicit polynomial in VNP that requires depth-4 circuits of size $2^{\Omega(\sqrt{n}\log n)}$ and Fournier et al. [FLMS14] 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) [CCL08]. 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. So for $d = n$, we get that the iterated matrix multiplication polynomial achieves the current best known lower bounds in both fronts: depth-4 circuit size and determinantal complexity. Our result also 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 $d > 1$ [Jan11].

1 Introduction

In a surprising result, Agrawal and Vinay [AV08] showed that proving exponential size lower bounds against depth four circuits imply exponential size lower bounds for general arithmetic circuits. In particular, given a polynomial (or sub-exponential) sized general arithmetic circuit, it can be transformed into a depth four circuit of sub-exponential size. Koiran [Koi12] and Tavenas [Tav13] carefully analyzed the chasm shown in [AV08] and came up with an *improved* depth reduction[1].

Theorem 1.1 ([AV08, Koi12, Tav13, CKSV16]). *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 $s^{O(t+d/t)}$.*

[*] Supported by TCS research fellowship.

[1] A newer proof of the chasm which yields a more structured depth four circuit was presented in [CKSV16].

The above theorem tells us that proving a size lower bound of $n^{\omega(d/t)}$ against the $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ would imply super polynomial circuit size lower bounds against general circuits. Towards proving such lower bounds Gupta et al. [GKKS13] proved a lower bound of $2^{\Omega(n/t)}$ against $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuits computing the determinant or the permanent polynomial over a $n \times n$ matrix. They used the dimension of the *shifted partial derivative* space as the complexity measure [2]. Kayal, Saha and Saptharishi [KSS14] pushed the bound of Gupta et al. [GKKS13] to $N^{\Omega(d/t)}$ for $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuits computing an explicit polynomial in VNP of degree d over N variables. This candidate polynomial is based on the combinatorial designs of Nisan and Wigderson [NW94] (see Definition 2.2).

In another surprising result Fournier et al. [FLMS14] 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, Saha and Saptharishi [KSS14] or the one by Fournier et al., is weaker than the constant in the bound of Tavenas [Tav13].

One of the main motivations of our study comes from this tantalizing fact that two seemingly different polynomials $\mathsf{NW}_{n,r}$ (which is inVNP) and $\mathsf{IMM}_{n,n}$ (which is in VP) behave very similarly as far as the $2^{\Omega(\sqrt{n}\log n)}$-size lower bound against $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ 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 $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ 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 four circuit size lower bounds for the $\mathsf{NW}_{n,r}$ and $\mathsf{IMM}_{n,d}$ polynomials. Formally, we prove the following.

Theorem 1.2. *Let f be a $n^{O(1)}$-variate polynomial of degree n. Let there be $s \geq n^{\delta k}$ (δ is any constant > 0) different polynomials in $\langle \partial^{=k}(f)\rangle$ for $k = \varepsilon\sqrt{n}$ such that any two of their leading monomials have pair-wise distance of at least $\Delta \geq \frac{n}{c}$ for any constant $c > 1$, and $0 < \varepsilon < \frac{1}{40c}$. Then any depth-4 $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ circuit that computes f must be of size $e^{\Omega_{\delta,c}(\sqrt{n}\ln n)}$.*

It is now well understood that the current known techniques can not 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 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 [KS14a, KLSS14, KS14b, KS16a, KS16b] (see [Sap15] for an exposition of each of these results).

Let us recall that 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 VP if f_n has degree of at most poly(n) and can be computed by an arithmetic circuit of size poly(n). It is in 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 VP. Permanent polynomial characterizes the class VNP over the fields of all characteristics except 2 and the determinant polynomial characterizes the class 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* DetComp(f). ◊

To resolve Valiant's hypothesis, proving DetComp(Perm_n) $= n^{\omega(\log n)}$ is sufficient. Von zur Gathen [vzG86] proved that DetComp(Perm_n) $\geq \sqrt{\frac{8}{7}}n$. Later Cai [Cai90], Babai and Seress [vzG87], and Meshulam [Mes89] independently improved the lower bound to $\sqrt{2}n$. Mignon and Ressayre [MR04] proved that DetComp(Perm_n) $\geq \frac{n^2}{2}$ over the fields of characteristic zero, using algebraic geometry. Subsequently, Cai et al. [CCL08] extended the result of Mignon and Ressayre [MR04] to all fields of characteristic $\neq 2$. They also provided a simpler analysis.

[2] Kayal [Kay12] 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 $x_1 x_2 \ldots x_n$.

For any polynomial f, Valiant [Val79] proved that $\mathrm{DetComp}(f) \leq 2(F(f)+1)$ where $F(f)$ is the arithmetic formula complexity of f. Later, Nisan [Nis91] proved that $\mathrm{DetComp}(f) = O(B(f))$ 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 any integers n and $d > 1$, the determinantal complexity of the iterated matrix multiplication polynomial $\mathsf{IMM}_{n,d}$ is $0.5dn$.*

Since $\mathsf{IMM}_{n,d}(X)$ has an algebraic branching program of size $O(dn)$ [Nis91], from the above theorem it follows that $\mathrm{DetComp}(\mathsf{IMM}_{n,d}(X)) = \Theta(dn)$. This improves upon the earlier bound of $\Theta(n)$ for the determinantal complexity of the iterated matrix multiplication polynomial for any constant $d > 1$ [Jan11]. Similar to the approach of [CCL08] and [MR04], we also use the the rank of 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 [MR04]. Before the result of Mignon and Ressayre, the best known determinantal complexity lower bound for the $n \times n$ permanent polynomial was $\sqrt{2}n$ [Cai90, vzG87]. These results were proved using nontrivial algebraic-geometric concepts. One possible approach to prove that $\mathsf{VP} \neq \mathsf{VNP}$ 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 $\mathrm{DetComp}(\mathsf{NW}_{n,\varepsilon n}) \geq \Omega(n^{1.5})$ using elementary ideas. Similarly we prove a *weaker* (compared to the one in Theorem 1.4) lower bound on the determinantal complexity of $\mathsf{IMM}_{n,d}$ using the method of partial derivatives.

In general, a *strong enough* 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 $\Omega(dn)$ for it has an algebraic branching program of that size. This does not imply any non trivial bound on the formula complexity of the polynomial. 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 $\mathsf{IMM}_{n,d}(X)$ polynomial must be of size $\Omega(dn^3)$.*

We prove this by adapting the argument of Kalorkoti [Kal85] 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 $\mathbb{F}[x_1, x_2, \ldots, x_n]$. 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 $\Sigma\Pi^{[D]}\Sigma\Pi^{[t]}$ 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 [GKKS13]) is key to the asymptotic estimates required for the lower bound analyses.

Lemma 2.1 (Lemma 6, [GKKS13]). *Let $a(n), f(n), g(n) : \mathbb{Z}_{\geq 0} \to \mathbb{Z}_{\geq 0}$ be the 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 $\left\{\mathsf{NW}_{n,r}\right\}$ as follows.*

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

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

Proposition 4 in [Val79] tells us that if there is a polynomial time algorithm to test if the coefficient of a given monomial is 1 in the given polynomial $P(X) \in \mathbb{F}[X]$ with $\{0,1\}$ coefficients, 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 $\{\mathsf{NW}_{n,r}\}_{n>0}$ by checking if it *conforms* to a univariate polynomial of degree at most r. Thus, $\{\mathsf{NW}_{n,r}\}_{n>0}$ is in VNP.

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

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

$\Diamond$

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

Shifted partial derivatives

For a monomial $\mathbf{x^i} = x_1^{i_1} x_2^{i_2} \dots x_n^{i_n}$, let $\partial^{\mathbf{i}} f$ be the partial derivative of f with respect to the monomial $\mathbf{x^i}$. The degree of the monomial is denoted by $|\mathbf{i}|$ where $|\mathbf{i}| = (i_1 + i_2 + \cdots + i_n)$. We also use the word length to refer to the degree of the monomial. We recall the following definition of shifted partial derivatives from [GKKS13].

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

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

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

Let $\succ$ be any admissible monomial ordering. The *leading monomial* of a polynomial $f(X) \in \mathbb{F}[X]$, denoted by $\mathsf{LM}(f)$ is the largest monomial $\mathbf{x^i} \in f(X)$ under the order $\succ$. The next lemma follows directly from Proposition 11 and Corollary 12 of [GKKS13].

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

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

In [KSS14], the following upper bound on the dimension of the shifted partial derivative space of the polynomials computed by $\Sigma\Pi^{[D]}\Sigma\Pi^{[t]}$ circuits was shown.

Lemma 2.6 (Lemma 4, [KSS14]). *If $C = \sum_{i=1}^{s'} Q_{i1}Q_{i2}\dots Q_{iD}$ where each $Q_{ij} \in \mathbb{F}[X]$ is a polynomial of degree bounded by t. Then for any $k \le D$,*

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

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 (multi)-sets of variables corresponding to the monomials m_1 and m_2 respectively. The distance $\mathrm{dist}(m_1, m_2)$ 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 (multi)-sets.* ◊

For example, let $m_1 = x_1^2x_2x_3^2x_4$ and $m_2 = x_1x_2^2x_3x_5x_6$. 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 $\mathrm{dist}(m_1, m_2) = 3$.

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 pair-wise 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 four circuit size lower bound results. In fact, the lower bounds on the size of $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ circuits computing the polynomials $\mathsf{NW}_{n,r}$ and $\mathsf{IMM}_{n,n}$ follow easily from this lemma by suitably setting the parameters.

Lemma 3.1. *Let $m_1, m_2, \dots, m_s$ be the monomials over N variables such that $\mathrm{dist}(m_i, m_j) \ge \Delta$ for all $i \ne j$. Let M be the set of monomials of the form $m_i m'$ where $1 \le i \le s$ and m' is a monomial of length at most ℓ over the same set of N variables. Then, the cardinality of M is at least $\left(sB - s^2\binom{N+\ell-\Delta}{N}\right)$ 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 ℓ. It is easy to see that $|B_i| = \binom{N+\ell}{N}$. We would like to estimate $|\cup_i B_i|$. Using the principle of inclusion and exclusion, we get $|\cup_{i=1}^s B_i| \ge \sum_{i\in[s]} |B_i| - \sum_{i,j\in[s],i\ne j} |B_i \cap B_j|$.

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

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 ℓ is lower bounded as follows.

$$|\cup_{i=1}^s B_i| \ge sB - s^2\binom{N+\ell-\Delta}{N} = sB\left(1 - \frac{s}{B}\binom{N+\ell-\Delta}{N}\right)$$

□

We use the above lemma to prove Theorem 1.2. Even though we prove the bounds against $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ circuits computing $n^{O(1)}$-variate polynomials of degree n, we state the following theorem with some generality in terms of the parameter t.

Theorem 3.2. *Let f be a $N = d^{O(1)}$-variate polynomial of degree d. Let there be at least $d^{\delta k}$ (δ is any constant > 0) different polynomials in $\langle \partial^{=k}(f)\rangle$ for $k = \varepsilon\frac{d}{t}$ such that any two of their leading monomials have a distance of at least $\Delta \ge \frac{d}{c}$ for any constant $c > 1$, and $0 < \varepsilon < \frac{1}{40c}$. Then any depth-4 $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuit that computes f must be of size $e^{\Omega_{\delta,c}(\frac{d}{t}\ln N)}$.*

Proof. Consider a set of $s = d^{\delta k}$ polynomials $f_1, f_2, \ldots, f_s \in \langle \partial^{=k}(f) \rangle$ such that $\mathsf{dist}(\mathsf{LM}(f_i), \mathsf{LM}(f_j)) \geq d/c$ for all $i \neq j$. We denote by m_i, the leading monomial $\mathsf{LM}(f_i)$. We now invoke Lemma 3.1 with the parameters $s = d^{\delta k}, \Delta = d/c$. Let N be the number of variables in f. From Lemma 3.1, we know that $|\cup_{i=1}^{s} B_i| \geq sB\left(1 - \frac{s}{B}\binom{N+\ell-\Delta}{N}\right)$. To get a good lower bound for $|\cup_{i=1}^{s} B_i|$, we need to upper bound $\frac{s}{B}\binom{N+\ell-\Delta}{N}$. Let us bound it by an inverse polynomial in n by suitably choosing ℓ. We set $\frac{s\binom{N+\ell-\Delta}{N}}{\binom{N+\ell}{N}} \leq \frac{1}{p(d)}$ where $p(d)$ is a polynomial in d.

After simplification, we get $s\frac{(N+\ell-\Delta)!}{(N+\ell)!}\frac{\ell!}{(\ell-\Delta)!} \leq \frac{1}{p(d)}$. Using Lemma 2.1 we tightly estimate the subsequent computations. In particular, we always choose the parameter ℓ such that $\Delta^2 = o(N+\ell)$. This also shows that the error term given by Lemma 2.1 is always asymptotically zero and we need not worry about it.

We now apply Lemma 2.1 to derive $s\left(\frac{\ell}{N+\ell}\right)^{\Delta} \leq \frac{1}{p(d)}$ or equivalently $s\left(\frac{1}{1+\frac{N}{\ell}}\right)^{\Delta} \leq \frac{1}{p(d)}$. We use the inequality $1+x > e^{x/2}$ for $0 < x < 1$ to lower bound $\left(1+\frac{N}{\ell}\right)^{\Delta}$ by $e^{\frac{N\Delta}{2l}}$. Thus, it is enough to choose ℓ in a way that $s \cdot p(d) \leq e^{\frac{N\Delta}{2\ell}}$ or equivalently $\ell \leq \frac{N\Delta}{2\ln(s \cdot p(d))}$. By fixing $p(d) = d^2$ and substituting for the parameters k and Δ , we get $\ell \leq \frac{Nt}{4c\delta\varepsilon \ln d}$. From Lemma 2.5, we get that the dimension of $\langle \partial^{=k} f \rangle_{\leq \ell} \geq (1 - \frac{1}{d^2})\, s\binom{N+\ell}{N}$.

Combining this with Lemma 2.6, we get $s' \geq \frac{(1-\frac{1}{d^2})s\binom{N+l}{N}}{\binom{D}{k}\binom{N+l+k(t-1)}{N}}$. Suppose we choose ℓ such that $(kt-k)^2 = o(\ell)$. Then, by applying Lemma 2.1 we can easily show the following.

$$s' \geq \frac{s\left(1-\frac{1}{d^2}\right)}{\binom{D}{k}(1+\frac{N}{l})^{(kt-k)}} \geq \frac{d^{\delta k}\left(1-\frac{1}{d^2}\right)}{\binom{D}{k}e^{\frac{N}{\ell}kt}}$$

Since $D = O(d/t)$ and $k = \varepsilon d/t$, we can estimate $\binom{D}{k}$ to be $e^{O_{\varepsilon}\left(\frac{d}{t}\right)}$ by Shannon's entropy estimate for binomial coefficients. To get the required lower bound it is sufficient to choose ℓ such that $\frac{Nkt}{\ell} < (0.1)\delta k \ln d$. By comparing the lower and upper bounds of ℓ, we can fix ε such that $\varepsilon < \frac{1}{40c}$. Since ε depends only on c, we can infer that $s' = e^{\Omega_{\delta,c}\left(\frac{d}{t}\ln d\right)} = e^{\Omega_{\delta,c}\left(\frac{d}{t}\ln N\right)}$. □

The above proof clearly goes through even if we set $\frac{Nkt}{\ell} < \mu\delta k \ln d$ for any $0 < \mu < 1$, and choose $\varepsilon < \frac{\mu}{4c}$.

4 Lower bounds for explicit polynomials

In this section we shall apply Theorem 3.2 to two explicit polynomials, $\mathsf{NW}_{n,r}$ which is a polynomial in VNP and $\mathsf{IMM}_{n,d}$ 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 $\mathsf{NW}_{n,r}$ polynomial by a simple application of Theorem 3.2 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 $\mathsf{NW}_{n,r}(X)$ must be of size $2^{\Omega(\sqrt{n}\log n)}$ where $r = \varepsilon\sqrt{n}$.*

Proof. Recall that $\mathsf{NW}_{n,\varepsilon}(X) = \sum_{a(z)\in\mathbb{F}[z]} x_{1,a(1)}x_{2,a(2)}\ldots x_{n,a(n)}$ where $\mathbb{F}$ is a finite field of size n and $a(z)$ is a univariate polynomial of degree $\leq r-1$ where $r = \varepsilon\sqrt{n}$. Notice that any two monomials can intersect in at most $r-1$ variables.

We differentiate the polynomial $\mathsf{NW}_{n,r}(X)$ with respect to the first $k = r = \varepsilon\sqrt{n}$ 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 Δ between any two monomials $\geq n - 2k > n/2$. So to get the required lower bound we invoke Theorem 3.2 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 $\mathsf{IMM}_{n,n}$.

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

Proof. Recall that $\mathsf{IMM}_{n,n}(X) = \sum_{i_1,i_2,\ldots,i_{n-1}\in[n]} x^{(1)}_{1,i_1} x^{(2)}_{i_1,i_2} \ldots x^{(n-1)}_{i_{(n-2)},i_{(n-1)}} x^{(n)}_{i_{(n-1)},1}$. It is a polynomial over $(n-2)n^2+2n$ variables. We fix the following lexicographic ordering on the variables of the set of matrices $\{X_1, X_2, \ldots, X_n\}$ as follows: $X_1 \succ X_2 \succ X_3 \succ \ldots \succ X_n$ and in any X_i the ordering is $x^{(i)}_{1,1} \succ x^{(i)}_{1,2} \succ \ldots \succ x^{(i)}_{1,n} \succ \ldots \succ x^{(i)}_{n,1} \ldots \succ x^{(i)}_{n,n}$.

Choose a prime p such that $\frac{n}{2} \leq p \leq n$. Consider the set of univariate polynomials $a(z) \in \mathbb{F}_p[z]$ of degree at most $(k-1)$ for $k = \varepsilon\sqrt{n}$ where ε is a small constant to be fixed later in the analysis. Consider a set of $2k$ of the matrices $X_2, X_{3+\frac{n}{4k}}, \ldots, X_{2k+1+\frac{(2k-1)n}{4k}}$ such that they are $n/4k$ 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 $S_a = \{x^{(2)}_{1,a(1)}, x^{(3+\frac{n}{4k})}_{2,a(2)}, \ldots, x^{(2k+1+\frac{(2k-1)n}{4k})}_{2k,a(2k)}\}$. Number of such sets is at least $\left(\frac{n}{2}\right)^k$ and $|S_a \cap S_b| < k$ for $a \neq b$. Now we consider a polynomial $f(X)$ which is a restriction of the polynomial $\mathsf{IMM}_{n,n}(X)$. By restriction, we simply mean that a few variables of $\mathsf{IMM}_{n,n}(X)$ are fixed to some elements from the field and the rest of the variables are left untouched. We define the restriction as follows.

$$x^{(q)}_{i,j} = 0 \text{ if } r + \frac{(r-2)n}{4k} < q < (r+1) + \frac{(r-1)n}{4k} - 1 \text{ for } 2 \leq r \leq 2k \text{ and } i \neq j.$$

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 $a(z) \in \mathbb{F}[z]$. Consider the leading monomial of the derivatives with respect to the sets S_a for all $a(z) \in \mathbb{F}[z]$. Since $|S_a \cap S_b| < k$, 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 3.2. Notice that $f(X)$ is a $n^{O(1)}$-variate polynomial of degree n such that there are at least $(n/2)^k > n^{\frac{1}{4}(2k)}$ 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 3.2. A simple calculation shows that the parameter ε can be fixed to something $< 1/320$.

Since $f(X)$ is a restriction of $\mathsf{IMM}_{n,n}(X)$, any lower bound for $f(X)$ is a lower bound for $\mathsf{IMM}_{n,n}(X)$ too. Otherwise, if $\mathsf{IMM}_{n,n}(X)$ 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 $\Gamma_k(f) := \dim\left(\partial^{=k} f\right)$.* ◊

If a polynomial $f = \mathsf{Det}_m(A(X))$ then we need $\Gamma_k(\mathsf{Det}(A(X)))$ must be at least $\Gamma_k(f)$. Let us first obtain a lower bound on the derivative space of $\mathsf{Det}_m(A(X))$.

Derivative space of Det_m polynomial

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

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

Since the entries of $A(X)$ are linear polynomials, $\frac{\partial(A(X))_{s,t}}{\partial x_{i,j}}$ is a constant. Generalizing this, we get that

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

This implies that the span of the partial derivative space of $\mathsf{Det}_m(A(X))$, of order k, is a subset of the span of the kth order partial derivative space (with respect to Y) of $\mathsf{Det}_m(Y)$. More formally,

$$\mathbb{F}\text{-span}\left\{\partial^{=k}\,\mathsf{Det}_m(A(X))\right\} \subseteq \mathbb{F}\text{-span}\left\{(\partial_S\,\mathsf{Det}_m(Y))|_{y_{ij}=A_{ij};i,j\in[m]} : S\subseteq Y \;\&\; |S|=k\right\}.$$

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

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

Invoking the Proposition 5.2 and from the discussion above, we get that

$$\Gamma_k\left(\mathsf{Det}_m(A(X))\right) \le \Gamma_k\left(\mathsf{Det}_m(Y)\right) = \binom{m}{k}^2.$$

Derivative space of Nisan-Wigderson polynomial

Let us recall that $\mathsf{NW}_{n,\varepsilon n}(X) = \sum_{a(z)\in\mathbb{F}[z]} x_{1,a(1)}x_{2,a(2)}\cdots x_{n,a(n)}$ 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 $\mathsf{NW}_{n,\varepsilon n}(X)$ 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, $\Gamma_k\left(\mathsf{NW}_{n,\varepsilon n}\right) \ge n^{\varepsilon n}$.

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

Proof. If the dimension of the partial derivative space of the $\mathsf{Det}_m(A(X))$ is less than the dimension of the partial derivative space of the $\mathsf{NW}_{n,\varepsilon n}(X)$ polynomial, then $\mathsf{NW}_{n,\varepsilon n}(X) = \mathsf{Det}_m(A(X))$ can not hold true. Thus for $k=\varepsilon n$,

$$\binom{m}{k}^2 \le n^{\varepsilon n}$$

$$\left(\frac{e\cdot m}{k}\right)^{2k} \ge n^{\varepsilon n}$$

$$m \ge \frac{\varepsilon n\cdot\sqrt{n}}{e} = \Omega(n^{1.5})$$

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

Derivative space of the iterated matrix multiplication polynomial

The iterated matrix multiplication polynomial is defined over the disjoint sets of variables $X_1, X_2, \cdots, X_d$.

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

[3] A much stronger statement about the determinantal ideal can be found in (Theorem 22) [GKKS13] and the references therein.

We will lower bound $\Gamma_k(\mathsf{IMM}_{n,d}(X))$ by the dimension of a specific subspace of the derivative space of $\mathsf{IMM}_{n,d}(X)$. That is dimension of the entire derivative space is lower bounded by the dimension of the subspace that we will now consider. For some distinct elements $J = \{j_1, j_2, \cdots, j_k\}$ such that $|j_s - j_t| > 2$ for any $s, t \in [k]$, consider the sets of variables $X_{j_1}, X_{j_2}, \cdots, X_{j_k}$. Let us consider the set of monomials M of degree k such that for any monomial $m \in M$, $\left|\mathrm{var}(m) \cap X_{j_t}\right| = 1$ for all $t \in [k]$, and all suitable sets J. It is easy to see that the partial derivatives with respect to the monomials in M are pairwise distinct. The number of ways of picking such a suitable set J is $\binom{d-k+1}{k}$. The number of monomials of degree k in M corresponding to a particular set J is n^{2k}. Thus,

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

Since we need $\Gamma_k(\mathsf{IMM}_{n,d}(X)) \leq \Gamma_k(\mathsf{Det}_m(A(X)))$. Thus for $k = \delta d$ for a suitable $\delta \in (0, 1)$,

$$\begin{aligned}
\binom{m}{k}^2 &\geq \binom{d-k+1}{k} \cdot n^{2k} \\
\left(\frac{e \cdot m}{k}\right)^{2k} &\geq \frac{d-k+1}{k}^k \cdot n^{2k} \\
\frac{e \cdot m}{k} &\geq \sqrt{\frac{d-k+1}{k}} \cdot n \\
m &\geq e^{-1} n \cdot \sqrt{k(d-k+1)} = \Omega(dn).
\end{aligned}$$

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

6 Determinantal complexity lower bounds via the Hessian

Approach of Mignon and Ressayre

We start by recalling a few facts from [CCL08]. Let f be the target polynomial over N variables. Let $A_{k,\ell}(X)$, $1 \leq k, \ell \leq m$ be the affine linear polynomials over $\mathbb{F}[X]$ such that $f(X) = \det((A_{k,\ell}(X))_{1 \leq k,\ell \leq m})$. Consider a point $X_0 \in \mathbb{F}^N$ such that $f(X_0) = 0$. 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, $(A_{k,\ell}(X))_{1 \leq k,\ell \leq m} = (L_{k,\ell}(X - X_0))_{1 \leq k,\ell \leq m} + \mathrm{Y}_0$. If $f(X_0) = 0$ then $\det(\mathrm{Y}_0) = 0$. Let C and D be two non-singular matrices such that $\mathrm{C}\mathrm{Y}_0\mathrm{D}$ is a diagonal matrix.

$$\mathrm{C}\mathrm{Y}_0\mathrm{D} = \begin{pmatrix} 0 & 0 \\ 0 & \mathrm{I}_s \end{pmatrix}$$

Since $\det(\mathrm{Y}_0) = 0$, $s < m$. It is enough to assume that $s = m - 1$. Since the first row and the first column of $\mathrm{C}\mathrm{Y}_0\mathrm{D}$ are zero, we may multiply $\mathrm{C}\mathrm{Y}_0\mathrm{D}$ by $\mathrm{diag}(\det(\mathrm{C})^{-1}, 1, \cdots, 1)$ and $\mathrm{diag}(\det(\mathrm{D})^{-1}, 1, \cdots, 1)$ on the left and the right side. Without loss of generality, we may assume that $\det(\mathrm{C}) = \det(\mathrm{D}) = 1$. 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

$$f(X) = \det((L_{k,\ell}(X - X_0)_{1 \leq k,\ell \leq m} + \mathrm{Y}_0))$$

where $\mathrm{Y}_0 = \mathrm{diag}(0, 1, \cdots, 1)$.

We use $\mathrm{Hess}_f(X)$ to denote the Hessian matrix of the polynomial f and is defined as follows.

$$\mathrm{Hess}_f(X) = (H_{s;ij,t;k\ell}(X))_{1 \leq i,j \leq n, 1 \leq s,t \leq d} \text{ such that } H_{s;ij,t;k\ell}(X) = \frac{\partial^2 f(X)}{\partial x_{i,j}^{(s)} \partial x_{k,\ell}^{(t)}}$$

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

By taking second order derivatives and evaluating the Hessian matrices of $f(X)$ and $\det((A_{k,\ell}(X))_{1\leq k,\ell\leq m})$ at X_0, we obtain $\mathrm{Hess}_f(X_0) = \mathrm{L}\,\mathrm{Hess}_{\det}(\mathrm{Y}_0)\mathrm{L}^T$ where L is a $N \times m^2$ matrix with entries from the field. It follows that $\mathrm{rank}(\mathrm{Hess}_f(X_0)) \leq \mathrm{rank}(\mathrm{Hess}_{\det}(\mathrm{Y}_0))$. It was observed in the earlier work of [MR04] and [CCL08] that it is relatively easy to get an upper bound for $\mathrm{rank}(\mathrm{Hess}_{\det}(\mathrm{Y}_0))$. The main task is to construct a point X_0 such that $f(X_0) = 0$, yet the rank of $\mathrm{Hess}_f(X_0)$ is high.

Determinantal complexity of $\mathsf{IMM}_{n,d}$

We shall fix our target polynomial to be $\mathsf{IMM}_{n,d}$ where $N = n^2d$. We give an explicit construction of a point $X_0 \in \mathbb{F}^{n^2d}$ such that $\mathsf{IMM}_{n,d}(X_0) = 0$ and $\mathrm{rank}(\mathrm{Hess}_{\mathsf{IMM}_{n,d}}(X_0)) \geq d(n-1)$. First for the sake of completeness, we briefly recall the upper bound argument for the rank of $\mathrm{Hess}_{\det}(\mathrm{Y}_0)$ from Section 2.1 of [CCL08].

Upper bound for the rank of $\mathrm{Hess}_{\det}(\mathrm{Y}_0)$

When we take a partial derivative of the determinant polynomial with respect to the variable $x_{i,j}$, the result is a minor that is obtained by striking out the row i and column j. The second order derivative of $\det(\mathrm{Y})$ 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 non-zero entries in $\mathrm{Hess}_{\mathsf{Det}}(\mathrm{Y}_0)$ are obtained only if $1 \in \{i,k\}$ and $1 \in \{j,\ell\}$ and thus $(ij,k\ell)$ are of the form $(11,tt)$ or $(t1,1t)$ or $(1t,t1)$ for any $t > 1$. Thus, $\mathrm{rank}(\mathrm{Hess}_{\det}(\mathrm{Y}_0)) = 2m$.

Lower bound for the rank of $\mathrm{Hess}_{\mathsf{IMM}_{n,d}}(X_0)$

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 $\mathsf{IMM}_{n,d}(X_0) = 0$ and $\mathrm{rank}(\mathrm{Hess}_{\mathsf{IMM}_{n,d}}(X_0)) \geq d(n-1)$. Since $\mathrm{rank}(\mathrm{Hess}_{\det}(\mathrm{Y}_0)) = 2m$ and $\mathrm{rank}(\mathrm{Hess}_{\mathsf{IMM}_{n,d}}(X_0)) \leq \mathrm{rank}(\mathrm{Hess}_{\det}(\mathrm{Y}_0))$, we get that $m = d(n-1)/2$. As mentioned in the introduction, the determinantal complexity of $\mathsf{IMM}_{n,d}(X)$ 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 $\mathsf{IMM}_{n,d}(X_0) = 0$ and $\mathrm{rank}(\mathrm{Hess}_{\mathsf{IMM}_{n,d}}(X_0)) \geq d(n-1)$. 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 maintain that the entries indexed by the indices $(1,2),(1,3),\cdots,(1,n)$ 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 $\mathsf{IMM}_{n,2}(X) = \sum_{i=1}^{n} x_{1,i}^{(1)}x_{i,1}^{(2)}$. It is easy to observe that the rank of the corresponding Hessian matrix is $2n > 2(n-1)$ at any point since each non-zero entry of the Hessian matrix is 1 and the structure of the Hessian matrix is the following: ,

$$\mathrm{Hess}_{\mathsf{IMM}_{n,2}}(X) = \begin{bmatrix} 0 & B_{1,2} \\ B_{2,1} & 0 \end{bmatrix}$$

where $B_{2,1} = B_{1,2}^T$ and the matrix $B_{1,2}$ is formally described as

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

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

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 $\mathsf{IMM}_{n,(d+1)}(X)$.

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

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

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

Thus, we have the following expression.

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

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

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

Each $B_{i,j}$ is a block of size $n^2 \times n^2$ and is indexed by the variables 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 non-zero rows in $B_{(d+1),d}$ are indexed by the variables $x^{(d+1)}_{1,1}, x^{(d+1)}_{2,1}, \cdots, x^{(d+1)}_{n,1}$. The potential non-zero entries for the row $x^{(d+1)}_{1,1}$ are indexed by the columns $x^{(d)}_{1,1}, x^{(d)}_{2,1}, \cdots, x^{(d)}_{n,1}$. Similarly the potential non-zero entries for the row $x^{(d+1)}_{2,1}$ are indexed by the columns $x^{(d)}_{1,2}, x^{(d)}_{2,2}, \cdots, x^{(d)}_{n,2}$ and so on.

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

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

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

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

$Q_1, Q_2, \cdots, Q_n$ are also the entries indexed by the indices $(1,1), (1,2), \cdots, (1,n)$ of the matrix obtained after multiplying the first $(d-1)$ matrices. By induction hypothesis, we know that the entries indexed by the indices $(1,2), \cdots, (1,n)$ are not all zero at the point X_0, the zero of the polynomial $\mathsf{IMM}_{n,d}(X)$. This also makes the rows indexed by the variables $x^{(d+1)}_{1,1}, x^{(d+1)}_{2,1}, \cdots, x^{(d+1)}_{n,1}$ linearly independent. It is important to note that $P_{1,1}(X) = \mathsf{IMM}_{n,d}(X)$.

Now, let us define a point such that it is a zero of the polynomial $\mathsf{IMM}_{n,(d+1)}(X)$. Let X_0 be the zero of the polynomial $P_{1,1}(X) = \mathsf{IMM}_{n,d}(X)$. Now to construct the new point, we inductively 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 $\mathsf{IMM}_{n,(d+1)}(X)$, as X_0 as well.

Now, consider the first $d \times d$ blocks of the Hessian matrix $\mathrm{Hess}_{\mathsf{IMM}_{n,(d+1)}}(X_0)$. It precisely represents the Hessian matrix of $P_{1,1}(X)$ which is also the Hessian matrix of the polynomial $\mathsf{IMM}_{n,d}(X)$ at the point X_0. This can be easily seen from the setting of the variables $x_{1,1}^{(d+1)} = 1$ and $x_{2,1}^{(d+1)} = x_{3,1}^{(d+1)} = \cdots = x_{n,1}^{(d+1)} = 0$. By induction hypothesis, the rank of this minor of $\mathrm{Hess}_{\mathsf{IMM}_{n,(d+1)}}(X_0)$ is at least $d(n-1)$. The only non-zero entries in the columns indexed by the variable set $X^{(d)}$ are indexed by the variables $x_{1,1}^{(d)}, x_{2,1}^{(d)}, \cdots, x_{n,1}^{(d)}$. This is because the other variables of X_d do not appear in $\mathsf{IMM}_{n,d}(X)$. The row in $B_{(d+1),d}$ indexed by $x_{1,1}^{(d+1)}$ is the only row that interferes with any of the rows of $B_{1,d}, B_{2,d}, \cdots, B_{d,d}$. The rows indexed by the variables $x_{2,1}^{(d+1)}, x_{3,1}^{(d+1)}, \cdots, x_{n,1}^{(d+1)}$ in $B_{(d+1),d}$ are linearly independent of the rows of $B_{1,d}, B_{2,d}, \cdots, B_{d,d}$. Hence the rank of $\mathrm{Hess}_{\mathsf{IMM}_{n,(d+1)}}$ at the point described is $\geq (d+1)(n-1)$.

For the purpose of induction, we must verify that the entries indexed by the indices $(1,2), (1,3), \cdots, (1,n)$ of the matrix obtained after multiplying the first d matrices are not all zero at X_0. These entries are the polynomials $P_{1,2}, P_{1,3}, \cdots, P_{1,n}$. We shall express each of the polynomials in terms of $\mathrm{Q}_1, \mathrm{Q}_2, \cdots, \mathrm{Q}_n$ as follows.

$$P_{1j} = \mathrm{Q}_1 x_{1,j}^{(d)} + \mathrm{Q}_2 x_{2,j}^{(d)} + \cdots + \mathrm{Q}_n x_{n,j}^{(d)} \text{ for } 2 \leq j \leq n.$$

By induction hypothesis, we already know that $Q_2, Q_3, \cdots, Q_n$ are not all zero at X_0. Notice that the variables in $X^{(d)} \setminus \{x_{1,1}^{(d)}, x_{2,1}^{(d)}, \cdots, x_{n,1}^{(d)}\}$ 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 $P_{1,2}, P_{1,3}, \cdots, P_{1,n}$ are not all zero when evaluated at the point X_0 (in fact, we can make all of them non-zero). It is clear that we construct the point X_0 in polynomial time. This completes the proof. □

7 Formula size lower bound for $\mathsf{IMM}_{n,d}$

In this section, we shall prove a super-linear but subquadratic lower bound on the size of any formula that computes the $\mathsf{IMM}_{n,d}$ polynomial. The following proof is an adaptation of the proof strategy of Kalorkoti [Kal85]. Let us first recall the notion of algebraic independence and transcendence degree.

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

The transcendental degree *of the polynomials $f_1, f_2, \cdots, f_m \in \mathbb{F}[X]$, denoted by* $\mathrm{trdeg}(f_1, f_2, \cdots, f_m)$*, 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* $\mathrm{trdeg}_{X'}(f)$ *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 [Kal85] (cf. [Sap15]).

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

Theorem 7.4. *For all integers $n, d > 0$, any arithmetic formula computing the $\mathsf{IMM}_{n,d}(X)$ 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 $M_1, M_2, \cdots, M_d$ be the generic $n \times n$ matrices being multiplied in $\mathsf{IMM}_{n,d}(X)$ 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_i and M_{i+2} by grouping jth row of M_i and jth column of M_{i+2} together, for all $j \in [n]$. We shall denote such a set by $X_{i,j} = \{x^{(i)}_{j,1}, \ldots, x^{(i)}_{j,n}, x^{(i+2)}_{1,j}, \ldots, x^{(i+2)}_{n,j}\}$. The final partition of the variables is as follows.

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

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

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

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,

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

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

For the sake of contradiction, let us assume that there is a polynomial $g \in \mathbb{F}[Y]$ that annihilates the polynomials in $\mathscr{P}$. Consider two distinct monomials $m_1 = y_1^{\alpha_1} y_2^{\alpha_2} \cdots y_{n^2}^{\alpha_{n^2}}$ and $m_2 = y_1^{\beta_1} y_2^{\beta_2} \cdots y_{n^2}^{\beta_{n^2}}$ in g such that $\bar{\alpha} \neq \bar{\beta}$. Consider $m_1' = m_1|_{y_i \leftarrow P_i \in \mathscr{P}}$ and $m_2' = m_2|_{y_i \leftarrow P_i \in \mathscr{P}}$. We can see that $m_1' = P_1^{\left(\sum_{r\in[n^2]}\alpha_r\right)} \left(x^{(i+1)}_{1,1}\right)^{\alpha_1} \left(x^{(i+1)}_{1,2}\right)^{\alpha_2} \cdots \left(x^{(i+1)}_{n,n}\right)^{\alpha_{n2}}$ and $m_2' = P_1^{\left(\sum_{r\in[n^2]}\beta_r\right)} \left(x^{(i+1)}_{1,1}\right)^{\beta_1} \left(x^{(i+1)}_{1,2}\right)^{\beta_2} \cdots \left(x^{(i+1)}_{n,n}\right)^{\beta_{n2}}$.

W.l.o.g, let us assume that $\alpha_1 > \beta_1$. The overall degree of $x^{(i)}_{1,1}$ in m_1' is equal to α_1 and similarly the overall degree of the variable $x^{(i)}_{1,1}$ in m_2' is equal to β_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 can not annihilate the polynomials in $\mathscr{P}$. From Lemma 7.3, we get that the size of any arithmetic formula computing $\mathsf{IMM}_{n,d}(X)$ is of size at least $\sum_{i,j} \mathrm{trdeg}_{X_{i,j}}(\mathsf{IMM}_{n,d}(X)) = \Omega(dn^3)$.

□

Acknowledgements

We thank the anonymous reviewers of STACS 2014 for their invaluable feedback that helped improve the paper and take the current form.

References

[AV08] Manindra Agrawal and V Vinay. Arithmetic circuits: A chasm at depth four. In *Proceedings of Foundations of Computer Science (FOCS)*, pages 67–75. IEEE, 2008.

[Cai90] Jin-Yi Cai. A note on the determinant and permanent problem. *Information and Computation*, 84(1):119–127, 1990.

[CCL08] Jin-Yi Cai, Xi Chen, and Dong Li. A quadratic lower bound for the permanent and determinant problem over any characteristic$\neq 2$. In *Proceedings of Symposium on Theory of Computing*, pages 491–498. ACM, 2008.

[CKSV16] Suryajith Chillara, Mrinal Kumar, Ramprasad Saptharishi, and V. Vinay. The chasm at depth four, and tensor rank : Old results, new insights. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:96, 2016.

[FLMS14] Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. In *Proceedings of Symposium on Theory of Computing*, pages 128–135. ACM, 2014.

[GKKS13] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Approaching the chasm at depth four. In *Proceedings of the Conference on Computational Complexity (CCC)*, 2013.

[Jan11] Maurice Jansen. Lower bounds for the determinantal complexity of explicit low degree polynomials. *Theory of Computing Systems*, 49(2):343–354, 2011.

[Kal85] Kyriakos Kalorkoti. A Lower Bound for the Formula Size of Rational Functions. *SIAM Journal of Computing*, 14(3):678–687, 1985.

[Kay12] Neeraj Kayal. An exponential lower bound for the sum of powers of bounded degree polynomials. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:81, 2012.

[KLSS14] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An Exponential Lower Bound for Homogeneous Depth Four Arithmetic Circuits. In *Proceedings of Foundations of Computer Science (FOCS)*. IEEE, 2014.

[Koi12] Pascal Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theor. Comput. Sci.*, 448:56–65, 2012.

[KS14a] Mrinal Kumar and Shubhangi Saraf. The limits of depth reduction for arithmetic formulas: it's all about the top fan-in. In *Proceedings of Symposium on Theory of Computing*, pages 136–145. ACM, 2014.

[KS14b] Mrinal Kumar and Shubhangi Saraf. On the power of homogeneous depth 4 arithmetic circuits. In *Proceedings of Foundations of Computer Science (FOCS)*. IEEE, 2014.

[KS16a] Mrinal Kumar and Shubhangi Saraf. Arithmetic circuits with locally low algebraic rank. In *Proceedings of Conference on Computational Complexity (CCC)*, 2016.

[KS16b] Mrinal Kumar and Shubhangi Saraf. Sums of products of polynomials in few variables: Lower bounds and polynomial identity testing. In *Proceedings of Conference on Computational Complexity (CCC)*, 2016.

[KSS14] Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. In *Proceedings of Symposium on Theory of Computing*, pages 146–153. ACM, 2014.

[Mes89] Roy Meshulam. On two extremal matrix problems. *Linear Algebra and its Applications*, 114:261–271, 1989.

[MR04] Thierry Mignon and Nicolas Ressayre. A quadratic bound for the determinant and permanent problem. *International Mathematics Research Notices*, 2004(79):4241–4253, 2004.

[Nis91] Noam Nisan. Lower bounds for non-commutative computation. In *Proceedings of Symposium on Theory of Computing*, pages 410–418. ACM, 1991.

[NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994.

[Sap15] Ramprasad Saptharishi. A survey of lower bounds in arithmetic circuit complexity. Github survey, 2015.

[Tav13] Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. In *Proceedings of Mathematical Foundations of Computer Science (MFCS)*, pages 813–824, 2013.

[Val79] Leslie G Valiant. Completeness classes in algebra. In *Proceedings of Symposium on Theory of Computing (STOC)*, pages 249–261. ACM, 1979.

[vzG86] Joachim von zur Gathen. Permanent and determinant. In *Proceedings of Foundations of Computer Science (FOCS)*, pages 398–401. IEEE Computer Society, 1986.

[vzG87] Joachim von zur Gathen. Permanent and determinant. *Linear Algebra and its Applications*, 96:87–100, 1987.