Electronic Colloquium on Computational Complexity, Report No. 73 (2015)

Lower Bounds for Sums of Products of Low arity Polynomials

Neeraj Kayal
Microsoft Research India
neeraka@microsoft.com

Chandan Saha
Indian Institute of Science
chandan@csa.iisc.ernet.in

April 25, 2015

Abstract

We prove an exponential lower bound for expressing a polynomial as a sum of product of low arity polynomials. Specifically, we show that for the iterated matrix multiplication polynomial, $\mathrm{IMM}_{d,n}$ (corresponding to the product of d matrices of size $n \times n$ each), any expression of the form

$$\mathrm{IMM}_{d,n} = \sum_{i=1}^{s} \prod_{j=1}^{m} Q_{ij},$$

where the Q_{ij}'s are of arity at most $t \leq \sqrt{d}$ (i.e. each Q_{ij} depends on at most t variables), the number of summands s must be at least $d^{\Omega\left(\frac{d}{t}\right)}$.

A special case of this problem where the Q_{ij}'s are further restricted to have degree at most one was posed as an open problem by Shpilka and Wigderson [SW99] and recently resolved in [KS15a]. We show that a refinement of the argument in [KS15a] yields the above-mentioned lower bound on s, *regardless of the degrees of the Q_{ij}'s* (and also regardless of m, the number of factors in each summand). Lower bounds for the same model were also obtained in an almost simultaneous but independent work by Kumar and Saraf [KS15b].

ISSN 1433-8092

1 Introduction

The most natural way to compute a polynomial is via a sequence of basic arithmetic operations $+, -, \times$. This is formalized via the model of arithmetic circuits: an arithmetic circuit contains addition $(+)$ and multiplication $(\times)$ gates and it naturally computes a polynomial in the input variables over some underlying field. We typically allow the input edges to a $+$ gate to be labeled with arbitrary constants from the underlying field $\mathbb{F}$ so that a $+$ gate can in fact compute an arbitrary $\mathbb{F}$-linear combination of its inputs. A major open problem is to prove superpolynomial lower bounds for the size of a circuit computing some explicit polynomial. As a possible stepping stone, researchers have focused on restricted (but still nontrivial and interesting) subclasses of arithmetic circuits. In particular, circuits of low depth[1] are interesting for they correspond to computation which is highly parallel. An extra motivation for investigating lower bounds for low depth arithmetic circuits is provided by the depth reduction results initiated by the work of Hyafil [Hya79] with a significant improvement by Valiant et al. in [VSBR83] and subsequent refinements and improvements in [AJMV98, AV08, Koi12, GKKS13a, Tav13]. Lower bounds for low depth circuits enhance our understanding of the limits of depth reduction and *strong enough* lower bounds for low depth circuits would even imply lower bounds for general arithmetic circuits[2]. Note that proving lower bounds for circuit classes is considered to be among the most challenging problems in theoretical computer science and in particular, proving superpolynomial lower bounds for bounded depth arithmetic circuits remains an outstanding open problem.

Previous work on lower bounds. Arithmetic circuits being the most natural model for computing polynomials have been widely investigated and lower bounds are known for many restricted subclasses of arithmetic circuits such as for monotone circuits [JS82], for noncommutative formulas and branching programs [Nis91], for multilinear formulas [Raz09] and for regular formulas [KSS14]. See [SY10] for a more comprehensive survey of this area. Most relevant to the present work is a line of research on lower bounds for low depth arithmetic formulas[3].

Lower bounds for low depth formulas. If a formula $\mathcal{C}$ of depth Δ computes a polynomial $f(\mathbf{x})$ of degree d on N variables[4] then it corresponds to an identity of the form[5]

$$f(\mathbf{x}) = \sum_{i=1}^{s} \prod_{j=1}^{m} Q_{ij}, \tag{1}$$

where the Q_{ij}'s are *simpler polynomials* in the sense that they are computed by (small) formulas of depth $(\Delta - 2)$. Nisan and Wigderson [NW97] considered the case where each Q_{ij} is of

[1] Recall that the depth of a circuit is the maximum length of any path in the circuit.

[2] For example, the result of [Tav13] states that a lower bound of $N^{\omega(\sqrt{d})}$ for homogeneous depth four circuits computing an N-variate polynomial $f(\mathbf{x})$ of degree d implies that f has no $\mathsf{poly}(N)$-sized arithmetic circuits.

[3] Constant depth circuits can be simulated by constant depth formulas incurring only a polynomial overhead and hence the two terms - constant depth circuits and constant depth formulas - are often used in an interchangeable manner.

[4] The polynomial f which is supposed to be computed by our circuit is also referred to as the target polynomial or the output polynomial for $\mathcal{C}$.

[5] Most multivariate polynomials of interest are irreducible and in such a case we can assume without loss of generality that the output gate of the formula is a $+$ gate. By merging any $+$ gate adjacent to the output gate into a single $+$ gate (with slightly larger fanin), we can assume without loss of generality that all the gates feeding into the output gate are $\times$ gates.

degree one (so that $\mathcal{C}$ overall corresponds to depth three circuits) and obtained a lower bound of roughly[6] $\left(\frac{N}{m}\right)^{\Omega(d)}$. Recently, [GKKS13b, KSS14] (building on [Kay12]) generalized it to Q_{ij}'s of low degree (degree at most t) and obtained a lower bound of[7] roughly $\left(\frac{N}{m}\right)^{\Omega\left(\frac{d}{t}\right)}$ (with subsequent improvements in the complexity of the target polynomial by [FLMS14, KS14a]). Building on this, [KLSS14b, KS14c, KLSS14a] obtained a lower bound of roughly[8] $\left(\frac{N}{m}\right)^{\frac{d \cdot \log d}{\log t}}$ for the case when the Q_{ij}'s are t-sparse polynomials (with subsequent improvements in the complexity of the target polynomial in [KS14b]). Now note that when the Q_{ij}'s have degree one (so that $\mathcal{C}$ corresponds to depth three circuits), the overall size of the circuit is $O(s \cdot m \cdot N)$ while the lower bound of [NW97] degrades rapidly as m increases and becomes trivial as m exceeds the number of variables N. Shpilka and Wigderson [SW99] pointed out[9] that the techniques available at that time did not seem to yield lower bounds (for large m) even when the Q_{ij}'s are degree one polynomials of bounded arity[10], say of arity $t = O(1)$ and posed this case as an open problem. Recently, [KS15a] showed that the complexity measure developed in [KLSS14a] (to handle the case of sparse Q_{ij}'s) also applies to the problem posed in [SW99] and yields a lower bound[11] of roughly $N^{\Omega\left(\frac{d}{t}\right)}$ (regardless of the number of factors m in each term). Also, it follows[12] from the work of [GKKS13a] that any asymptotic improvement in the exponent of the lower bound in [KS15a] would lead to superpolynomial lower bounds for general arithmetic circuits. While we do not improve the asymptotics of the lower bound in [KS15a], we show here that a refinement of that argument can be used to remove the degree restriction on the Q_{ij}'s without incurring any asymptotic loss in the lower bound.

Our results. The *target polynomial* to which our lower bound applies is the iterated matrix multiplication polynomial which we now define.

Definition 1. The Iterated Matrix Multiplication Polynomial. *Fix any $d, n \in \mathbb{N}$ with $d, n \geq 2$. Define sets of variables $X_1, X_2, \ldots, X_d$ as follows. If $p \in \{1, d\}$, $X_p = \{x_{j,p}\}$ is a set of n variables; otherwise $X_p = \{x_{i,j,p}\}$ is a set of n^2 variables. We think of X_1 and X_d as row and column vectors of variables respectively and of X_p (for $p \in \{2, 3, \ldots, (d-1)\}$ as $n \times n$ matrices of variables. Now we define $\mathrm{IMM}_{d,n}$ as (the unique entry of) the product of the matrices $X_1, X_2, \ldots, X_d$. Formally,*

$$\mathrm{IMM}_{d,n} = \sum_{j_1, j_2, \ldots, j_{d-1}} x_{j_1,1} \cdot x_{j_1,j_2,2} \cdot \ldots \cdot x_{j_{d-2},j_{d-1},d-1} \cdot x_{j_{d-1},d}.$$

[6] for $N \gg d$

[7] for $N \gg d$.

[8] For the situation where $N \geq d^2$. The lower bound is independent of the degree of the Q_{ij}'s but as the lower bound expression itself indicates, it degrades rapidly with m.

[9] [SW99] were motivated in part by the fact that the elementary symmetric polynomials (which are some sort of arithmetic analogs of threshold gates) can be computed by $O(N^2)$-sized expressions of the form (1) wherein the Q_{ij}'s have degree one and arity also one.

[10] Recall that the arity of a function is the number of variables on which it depends. In particular, for a polynomial $f(x_1, x_2, \ldots, x_N)$ its arity is the number of variables x_i such that $\deg_{x_i}(f) > 0$.

[11] For arity t upto $O(\sqrt{d})$ and for $N \gg d$.

[12] It was pointed to us by Ramprasad Saptharishi (personal communication) that the depth reduction result in [GKKS13a] implies in particular that if an N-variate polynomial $f(\mathbf{x})$ of degree d can be computed by $\mathsf{poly}(N)$-sized arithmetic circuits then it admits a representation of the form (1) wherein the Q_{ij}'s have degree one, arity at most $\sqrt{d}$ and $s, m \leq N^{O(\sqrt{d})}$.

Throughout the rest of this paper, we will use $N = 2n + (d-2)n^2$ to denote the number of variables in $\mathrm{IMM}_{d,n}$ and $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ to denote the tuple of N variables involved in $\mathrm{IMM}_{d,n}$.

We consider representations of the form

$$\mathrm{IMM}_{d,n}(\mathbf{x}) = \sum_{i=1}^{s} Q_{i1}(\mathbf{x}) \cdot Q_{i2}(\mathbf{x}) \cdot \ldots \cdot Q_{im}(\mathbf{x}), \quad \text{where arity of } Q_{ij} \leq t \ \forall\ i, j. \tag{2}$$

and give an exponential lower bound on the number of summands s in any such representation. Formally, we have:

Theorem 1. Lower Bound for sums of product of low arity polynomials. *Let $\mathbb{F}$ be a field and let $\mathrm{IMM}_{d,n}$ be the polynomial corresponding to the iterated product of d matrices of size $n \times n$ each as defined above. Then for $n = d^5$, in any expression of the form*

$$\mathrm{IMM}_{n,d} = \sum_{i=1}^{s} Q_{i1} \cdot Q_{i2} \cdot \ldots \cdot Q_{im},$$

where each Q_{ij} is a polynomial of arity at most $t \leq \sqrt{d}$ (i.e. Q_{ij} depends on at most t variables), the number of summands s must be at least $d^{\Omega\left(\frac{d}{t}\right)}$.

As indicated above, our argument is a refined (and a bit more subtle) version of the argument in [KS15a] and we highlight the difference in the proofs in more detail in remark 10, after giving the proof of our main theorem.

A recent independent result. Very recently, an independent and almost simultaneous piece of work by Kumar and Saraf [KS15b], has results and techniques which are very similar to this work. In particular both employ the complexity measure called the dimension of projected shifted partials (DPSP for short) for proving the lower bound in this model. There are a few (perhaps relatively minor) differences though. [KS15b] seek to allow the arity of the Q_{ij}'s to be as large as possible while still obtaining a nontrivial lower bound and so they chose the target polynomial to be the Nisan-Wigderson design polynomial from [KS15a] and this allows them, over a field $\mathbb{F}$ of characteristic zero, to have the arity of the Q_{ij}'s to be as large as N^{μ} for any constant $\mu < 1$. On the other hand, we wanted our target polynomial to be easy to compute and the lower bound to hold over any field and hence choose the target polynomial to be the iterated-matrix multiplication polynomial which is easy to compute and for which a lower bound on the DPSP-complexity over any field $\mathbb{F}$ was obtained in an earlier work by Kumar and Saraf [KS14b]. This then yields a lower bound over any field $\mathbb{F}$ but wherein the arity of the Q_{ij}'s is substantially smaller compared to [KS15b]. The proofs are also similar but there are some small technical differences which then allow our lower bound to be independent of m. Meanwhile, Kumar and Saraf [KS15b] go on to apply the lower bound techniques to do identity testing of similar circuits and obtain further nice results in that direction.

2 Preliminaries

Homogeneous components of polynomials. Let $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be a polynomial of degree d. The r-th homogeneous component of f, which we denote by $f^{[r]}$ is the sum of all monomials of (total) degree exactly r with corresponding coefficients from f, i.e. the coefficient of any monomial $\boldsymbol{m}$ in $f^{[r]}$ is the same as the coefficient of $\boldsymbol{m}$ in f if $\boldsymbol{m}$ has degree r and is zero otherwise. Note that any f admits a unique decomposition into homogeneous components as follows:

$$f(\mathbf{x}) = f^{[0]} + f^{[1]} + \ldots + f^{[d-1]} + f^{[d]}.$$

A numerical estimate. The following numerical estimate from [GKKS13b] will be useful.

Lemma 2. *Let $a(n)$, $b(n)$, $c(n)$: $\mathbb{Z}_{>0} \mapsto \mathbb{Z}$ be integer valued functions such that $(|b| + |c|) = o(a)$. Then*

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

The proof is a simple application of Stirling's formula for estimating factorials.

3 Proof of theorem 1

In this section, we give a proof our main theorem, namely theorem 1. We want to prove lower bounds for representations of the form

$$\mathrm{IMM}_{d,n}(\mathbf{x}) = \sum_{i=1}^{s} Q_{i1}(\mathbf{x}) \cdot Q_{i2}(\mathbf{x}) \cdot \ldots \cdot Q_{im}(\mathbf{x}), \quad \text{where arity of } Q_{ij} \le t \;\forall\; i,j. \tag{3}$$

We give an outline of the proof before filling in the details.

3.1 Proof Outline

We first do some preprocessing to ensure that each Q_{ij} has a nonzero constant term as follows. Starting from equation (3), we do a random translation $\mathbf{x} \mapsto \mathbf{x} + \mathbf{a}$ to obtain another identity of the form

$$\mathrm{IMM}_{d,n}(\mathbf{x}) = \sum_{i=1}^{s} T_i^{[d]}, \text{ where } T_i = \prod_{j=1}^{m} \hat{Q}_{ij}(\mathbf{x}) \text{ and } \hat{Q}_{ij}(\mathbf{0}) \neq 0, \mathrm{arity}(\hat{Q}_{ij}) \le t. \tag{4}$$

for all $1 \le i \le s$, $1 \le j \le m$. We then employ a suitable complexity measure $\mathsf{DPSP} : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{R}_{\ge 0}$ and compare the value of DPSP for the two sides of an identity of the form (4) to obtain the lower bound. In a bit more detail we have:

1. **[Step 1]:** We formally define the DPSP measure and note down its basic properties, including the fact that DPSP is **sub-additive**, i.e. for *any* two polynomials f and g and for all $\alpha, \beta \in \mathbb{F}$ it holds that $\mathsf{DPSP}(\alpha \cdot f + \beta \cdot g) \le \mathsf{DPSP}(f) + \mathsf{DPSP}(g)$.

2. **[Step 2]:** For all $1 \le i \le s$ we have $\mathsf{DPSP}(T_i^{[d]})$ is *relatively small* (see lemma 8 for the quantitative bound).

3. **[Step 3]:** $\mathsf{DPSP}(IMM_{d,n})$ is *relatively large* (see lemma 9 for the quantitative bound).

We then fit these pieces together in section 3.6 to obtain the lower bound given in theorem 1.

3.2 The preprocessing step

Consider the identity given by equation (3). First note that we can assume without loss of generality that every Q_{ij} is a nonzero polynomial (else the corresponding term vanishes and we obtain a similar but smaller identity). Now apply the translation $\mathbf{x} \mapsto \mathbf{x} + \mathbf{a}$. We then have:

$$\mathrm{IMM}_{d,n}(\mathbf{x}+\mathbf{a}) = \sum_{i=1}^{s} Q_{i1}(\mathbf{x}+\mathbf{a}) \cdot Q_{i2}(\mathbf{x}+\mathbf{a}) \cdot \ldots \cdot Q_{im}(\mathbf{x}+\mathbf{a}).$$

Let $\hat{Q}_{ij}(\mathbf{x}) \stackrel{\text{def}}{=} Q_{ij}(\mathbf{x}+\mathbf{a})$ and let $T_i(\mathbf{x}) \stackrel{\text{def}}{=} \prod_{j=1}^{m} \hat{Q}_{ij}(\mathbf{x})$. Note that for each i, j we have $\mathrm{arity}(\hat{Q}_{ij}) = \mathrm{arity}(Q_{ij}) \leq t$ and moreover that $\hat{Q}_{ij}(\mathbf{0}) = Q_{ij}(\mathbf{a})$. By a simple application of the DeMillo-Lipton-Schwartz-Zippel lemma we get that for *most* points $\mathbf{a} \in \mathbb{F}^N$, it holds that $\hat{Q}_{ij}(\mathbf{0}) = Q_{ij}(\mathbf{a}) \neq 0$. Fix a point $\mathbf{a} \in \mathbb{F}^N$ such that $\hat{Q}_{ij}(\mathbf{0}) \neq 0$ for all $1 \leq i \leq s$ and $1 \leq j \leq m$. Let $T_i \stackrel{\text{def}}{=} \prod_{j \in [m]} \hat{Q}_{ij}$. Then we have

$$\mathrm{IMM}_{d,n}(\mathbf{x}+\mathbf{a}) = \sum_{i=1}^{s} T_i(\mathbf{x}).$$

Comparing the homogeneous components of degree d on the two sides of the above identity we have:

$$\mathrm{IMM}_{d,n}^{[d]}(\mathbf{x}+\mathbf{a}) = \sum_{i=1}^{s} T_i^{[d]}(\mathbf{x}).$$

Now since $\mathrm{IMM}_{d,n}$ is homogeneous of degree d to begin with, we have that $\mathrm{IMM}_{d,n}^{[d]}(\mathbf{x}+\mathbf{a}) = \mathrm{IMM}_{d,n}(\mathbf{x})$. This yields an identity of the form given in equation (4), as required.

3.3 Step 1: A subadditive complexity measure

The complexity measure that we employ here, called *projected shifted partials* was first defined in [KLSS14a] and subsequently used again in [KS14b, KS15a, BC15]. Let $\boldsymbol{m} = x_{i_1} \cdots x_{i_k}$ be a monomial in $\mathbf{x}$. Denote $\frac{\partial^k}{\partial x_{i_1} \cdots \partial x_{i_k}} f$ by $\partial_{\boldsymbol{m}} f$ and define

$$\boldsymbol{\partial}_{\mathrm{ML}}^{=k} f := \{\partial_{\boldsymbol{m}} f \mid \boldsymbol{m} \text{ is a multilinear monomial of degree } k\}$$

We will refer to $\boldsymbol{\partial}_{\mathrm{ML}}^{=k} f$ as the set of all *multilinear* k-th order partial derivatives of $f \in \mathbb{F}[\mathbf{x}]$. Let $\mathbf{x}^{=\ell}$ be the set of all multilinear monomials in $\mathbf{x}$ of degree equal to ℓ. We denote by $\mathbf{x}^{=\ell} \cdot \boldsymbol{\partial}_{\mathrm{ML}}^{=k} f$ the set of all polynomials of the form $\boldsymbol{m} \cdot g$ where $\boldsymbol{m} \in \mathbf{x}^{=\ell}$ and $g \in \boldsymbol{\partial}_{\mathrm{ML}}^{=k} f$. Define a map $\pi : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{F}[\mathbf{x}]$ such that when π acts on a polynomial f, it retains only and exactly the *multilinear monomials* of f. More precisely, let M_f be the set of all monomials with nonzero coefficients in f. Then, $\pi(f) := \sum_u c_u \boldsymbol{m}_u$ where $\boldsymbol{m}_u$ is a multilinear monomial in M_f and coefficient of $\boldsymbol{m}_u$ in f is c_u. Naturally, π is a linear map, i.e. $\pi(af + bg) = a \cdot \pi(f) + b \cdot \pi(g)$ for every $a, b \in \mathbb{F}$ and $f, g \in \mathbb{F}[\mathbf{x}]$. The definition of π extends naturally to sets of polynomials: For $A \subseteq \mathbb{F}[\mathbf{x}]$, let $\pi(A) := \{\pi(f) \mid f \in A\}$. For integers k and ℓ, the space of projected shifted partials of f is the linear span (i.e. $\mathbb{F}$-span) of the polynomials in $\pi(\mathbf{x}^{=\ell} \cdot \boldsymbol{\partial}_{\mathrm{ML}}^{=k} f)$. The measure we use is the dimension of this space of projected shifted partials, denoted by $\mathsf{DPSP}_{k,\ell}$ (or simply DPSP assuming parameters k and ℓ are fixed suitably):

$$\mathsf{DPSP}_{k,\ell}(f) := \dim(\pi(\mathbf{x}^{=\ell} \cdot \boldsymbol{\partial}_{\mathrm{ML}}^{=k} f)).$$

Observe that the measure $\mathsf{DPSP}_{k,\ell}$ obeys subadditivity, i.e.

Proposition 3. (in [KLSS14a, KS14b, KS15a]): *For any pair of polynomials $f, g \in \mathbb{F}[\mathbf{x}]$ and for all integers $k, \ell \geq 0$ and for all $\alpha, \beta \in \mathbb{F}$ it holds that:*

$$\mathsf{DPSP}_{k,\ell}(\alpha \cdot f + \beta \cdot g) \leq \mathsf{DPSP}_{k,\ell}(f) + \mathsf{DPSP}_{k,\ell}(g).$$

In particular, this means that in order to bound how large our measure can be for the right hand side of the equation (4), it suffices to obtain an upper bound on a single term $T_i^{[d]}$, where the corresponding T_i is a product of polynomials of low arity. We will also use the following property of DPSP in order to obtain an upper bound on $\mathsf{DPSP}(T_i^{[d]})$.

Proposition 4. *For any polynomial $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ and for all $k, \ell, r \geq 0$ we have*

$$\mathsf{DPSP}_{k,\ell}(f^{[r]}) \leq \mathsf{DPSP}_{k,\ell}(f)$$

Proof. For a set of polynomials $A \subseteq \mathbb{F}[\mathbf{x}]$, let

$$A^{[r]} \overset{\text{def}}{=} \{f^{[r]} \ : \ f \in A\}.$$

From the definition above, it follows that

$$\pi(\mathbf{x}^{=\ell} \cdot \boldsymbol{\partial}_{\text{ML}}^{=k} f^{[r]}) = \pi(\mathbf{x}^{=\ell} \cdot \boldsymbol{\partial}_{\text{ML}}^{=k} f)^{[r-k+\ell]}.$$

Consequently

$$\begin{aligned} \mathsf{DPSP}_{k,\ell}(f^{[r]}) &= \dim\left(\mathbb{F}\text{-span}\left(\pi(\mathbf{x}^{=\ell} \cdot \boldsymbol{\partial}_{\text{ML}}^{=k} f^{[r]})\right)\right) \\ &= \dim\left(\mathbb{F}\text{-span}\left(\pi(\mathbf{x}^{=\ell} \cdot \boldsymbol{\partial}_{\text{ML}}^{=k} f)^{[r-k+\ell]}\right)\right) \\ &\leq \dim\left(\mathbb{F}\text{-span}\left(\pi(\mathbf{x}^{=\ell} \cdot \boldsymbol{\partial}_{\text{ML}}^{=k} f)\right)\right) \end{aligned}$$

where the last inequality follows from the fact that

$$\dim(\mathbb{F}\text{-span}\left(A^{[r]}\right)) \leq \dim(\mathbb{F}\text{-span}\,(A)) \text{ for any } A \subseteq \mathbb{F}[\mathbf{x}] \quad \text{and any } r \geq 0.$$

□

Remark 5. We note in passing that a similar statement holds in much more generality. Suppose that $\pi : \mathbb{F}[\mathbf{x}] \mapsto \mathbb{R}_{\geq 0}$ is *any* complexity measure that is subadditive, i.e.

$$\pi(\alpha \cdot f + \beta \cdot g) \leq \pi(f) + \pi(g), \quad \text{for any } \alpha, \beta \in \mathbb{F}$$

and scale-invariant, i.e.

$$\pi(f(\alpha \cdot x_1, \alpha \cdot x_2, \ldots, \alpha \cdot x_N) = \pi(f(x_1, x_2, \ldots, x_N)) \quad \text{for any } \alpha \in \mathbb{F} \setminus \{0\}.$$

Then for all inetgers $r \geq 0$ we have

$$\pi(f^{[r]}) \leq (1 + \deg(f)) \cdot \pi(f).$$

This is because any homogeneous component of f can be expressed as a linear combination of $(\deg(f) + 1)$-many scaled versions of $f(\mathbf{x})$ (polynomials of the form $f(\alpha \cdot \mathbf{x})$ for $\alpha \in \mathbb{F}$). The advantage of DPSP is that we do not incur even the $(1 + \deg(f))$-factor in going from f to its homogeneous components.

3.4 Step 2: Upper bounding DPSP for a single term.

Our upper bound for the DPSP-complexity of a term is a refined (and more subtle) version of the argument in [KS15a]. We first recall the elementary symmetric polynomials -

$$\mathrm{E}\textsc{Sym}_d(y_1, y_2, \ldots, y_m) \stackrel{\text{def}}{=} \sum_{\substack{S \subseteq \{1,2,\ldots,m\} \\ |S|=d}} \prod_{i \in S} y_i$$

is the elementary symmetric polynomial of degree d on the m formal variables $y_1, y_2, \ldots, y_m$. We will need the following lemma which is implicit in [KS15a].

Lemma 6. *Let $R_1(\mathbf{x}), R_2(\mathbf{x}), \ldots, R_m(\mathbf{x}) \in \mathbb{F}[x_1, x_2, \ldots, x_N]$ be any set of m polynomials each of arity at most t. Then for all $r \geq 0$, all $k \geq 0$ and all $\ell \leq \frac{N}{2} - 2kt$ we have*

$$\mathsf{DPSP}_{k,\ell}(\mathrm{E}\textsc{Sym}_r(R_1, R_2, \ldots, R_m)) \leq 3^{\sqrt{r}} \cdot \binom{\sqrt{2r}+k}{k} \cdot \binom{N}{\ell + 2kt}$$

Remark 7. This upper bound is obtained in [KS15a] by expressing $\mathrm{E}\textsc{Sym}_r(R_1, R_2, \ldots, R_m)$ as a sum of $(3^{\sqrt{r}})$-many sums of products of *low support* polynomials (i.e. polynomials in which every monomial has only a few distinct variables). For the bound stated here, we need to also observe that for bounding the DPSP-complexity of a product of low support polynomials $P = L_1 \cdot L_2 \cdot \ldots L_m$, it is essentially only the number of *distinct* L_j's that is important and in this case this quantity can be upper-bounded by $\sqrt{2r}$.

Lemma 8. *Consider a term*

$$T_i(\mathbf{x}) = \hat{Q}_{i1}(\mathbf{x}) \cdot \hat{Q}_{i2}(\mathbf{x}) \cdot \ldots \cdot \hat{Q}_{im}(\mathbf{x}),$$

where each $\hat{Q}_{ij}$ has arity at most t and $\hat{Q}_{ij}(\mathbf{0}) \neq 0$. Then for all $d \geq 0$ and all $k \geq 0$ and all $\ell \leq \frac{N}{2} - 2kt$, it holds that

$$\mathsf{DPSP}_{k,\ell}(T_i^{[d]}) \leq d \cdot 3^{\sqrt{d}} \cdot \binom{\sqrt{2d}+k}{k} \cdot \binom{N}{\ell + 2kt}.$$

Proof. Let $\hat{Q}_{ij}(\mathbf{x}) = \alpha_{ij} \cdot (1 + R_{ij}(\mathbf{x}))$, where $\alpha_{ij} \neq 0$ and $R_{ij}(\mathbf{x})$ consists only of monomials of (total) degree at least one. Let $\alpha \stackrel{\text{def}}{=} \prod_{j \in [m]} \alpha_{ij}$ and $\mathbf{R} \stackrel{\text{def}}{=} (R_{i1}(\mathbf{x}), R_{i2}(\mathbf{x}), \ldots, R_{im}(\mathbf{x}))$. We then have

$$\begin{aligned} T_i(\mathbf{x}) &= \alpha \cdot (1 + R_{i1}) \cdot (1 + R_{i2}) \cdot \ldots \cdot (1 + R_{m1}) \\ &= \alpha \cdot (1 + \mathrm{E}\textsc{Sym}_1(\mathbf{R}) + \mathrm{E}\textsc{Sym}_2(\mathbf{R}) + \ldots + \mathrm{E}\textsc{Sym}_m(\mathbf{R})) \end{aligned}$$

Thus

$$\begin{aligned} T_i^{[d]}(\mathbf{x}) &= \alpha \cdot (\mathrm{E}\textsc{Sym}_1^{[d]}(\mathbf{R}) + \ldots + \mathrm{E}\textsc{Sym}_d^{[d]}(\mathbf{R}) + \ldots + \mathrm{E}\textsc{Sym}_m^{[d]}(\mathbf{R})) \\ &= \alpha \cdot (\mathrm{E}\textsc{Sym}_1^{[d]}(\mathbf{R}) + \mathrm{E}\textsc{Sym}_2^{[d]}(\mathbf{R}) + \ldots + \mathrm{E}\textsc{Sym}_d^{[d]}(\mathbf{R})) \end{aligned}$$

where the last equality holds for the following reason: since each $R_{ij}(\mathbf{x})$ consists of monomials degree at least one, it follows that any degree-r homogeneous expression in $\mathbf{R}$ (and $\mathrm{E}\textsc{Sym}_r(\mathbf{R})$

in particular) consists only of monomials of degree r or more and hence $\mathrm{ES}{\scriptstyle\mathrm{YM}}_r^{[d]}(\mathbf{R}) = 0$ for any $r > d$. By subadditivity of the DPSP-measure we have

$$\begin{aligned}
\mathsf{DPSP}_{k,\ell}(T_i^{[d]}) &\leq \sum_{r=1}^{d} \mathsf{DPSP}_{k,\ell}(\mathrm{ES}{\scriptstyle\mathrm{YM}}_r^{[d]}(\mathbf{R})) \\
&\leq \sum_{r=1}^{d} \mathsf{DPSP}_{k,\ell}(\mathrm{ES}{\scriptstyle\mathrm{YM}}_r(\mathbf{R})) \quad \text{(via Proposition 4)} \\
&\leq \sum_{r=1}^{d} 3^{\sqrt{r}} \cdot \binom{\sqrt{2r}+k}{k} \cdot \binom{N}{\ell+2kt} \quad \text{(via Lemma 6)} \\
&\leq d \cdot 3^{\sqrt{d}} \cdot \binom{\sqrt{2d}+k}{k} \cdot \binom{N}{\ell+2kt}, \quad \text{as required.}
\end{aligned}$$

This proves our upper bound on the DPSP-complexity of a term.

□

3.5 Step 3: Lower bounding DPSP for IMM.

The work by Kumar and Saraf [KS14b] studied the dimension of projected shifted partials of iterated matrix multiplication and they obtained a lower bound on DPSP-complexity of $\mathrm{IMM}_{d,n}$ over any field $\mathbb{F}$, for the appropriate choice of parameters.

Choice of parameters. We choose our parameters as follows.

$$n = d^5, \quad k = \frac{d}{32t}, \quad \ell = \frac{N}{2} \cdot \left(1 - \frac{d^{\frac{1}{32t}} - 1}{d^{\frac{1}{32t}} + 1}\right) \tag{5}$$

Lemma 9. implicit in [KS14b]. *Let $\mathbb{F}$ be any field. Then for the choice of parameters as in (5) above, we have*

$$\mathsf{DPSP}_{k,\ell}(\mathrm{IMM}_{d,n}) \geq \frac{1}{d^5} \cdot d^{\frac{7}{8}\cdot k} \cdot \binom{N}{\ell}$$

3.6 Putting everything together

With these bounds on the DPSP-complexity of the circuit and of $\mathrm{IMM}_{d,n}$ in our hand, we are ready to obtain a proof of theorem 1. Consider representations of the form given by equation (4). Comparing the DPSP-complexity of the two sides we have:

$$\begin{aligned}
\mathsf{DPSP}_{k,\ell}(\mathrm{IMM}_{d,n}) &= \mathsf{DPSP}_{k,\ell}(\sum_{i=}^{s} T_i^{[d]}) \\
&\leq \sum_{i=1}^{s} \mathsf{DPSP}_{k,\ell}(T_i^{[d]}) \quad \text{(via lemmas 3)} \\
&\leq s \cdot 3^{\sqrt{d}} \cdot \binom{\sqrt{2d}+k}{k} \cdot \binom{N}{\ell+2kt} \quad \text{(via lemma 8).}
\end{aligned}$$

so that

$$
\begin{aligned}
s &\geq \frac{\mathsf{DPSP}_{k,\ell}(\mathrm{IMM}_{d,n})}{3^{\sqrt{d}} \cdot \binom{\sqrt{2d}+k}{k} \cdot \binom{N}{\ell+2kt}} \\
s &\geq \frac{\frac{1}{d^5} \cdot d^{\frac{7}{8}\cdot k} \cdot \binom{N}{\ell}}{3^{\sqrt{d}} \cdot \binom{\sqrt{2d}+k}{k} \cdot \binom{N}{\ell+2kt}} \quad \text{(via lemma 9 and for } n, k \text{ as in equation (5))} \\
&= \frac{1}{d^5 \cdot 3^{\sqrt{d}}} \cdot \frac{d^{\frac{7}{8}\cdot k}}{\binom{\sqrt{2d}+k}{k}} \cdot \frac{(\ell+2kt)!}{(\ell)!} \cdot \frac{(N-\ell-2kt)!}{(N-\ell)!}
\end{aligned}
$$

Now plugging in the choice of parameters n, k, ℓ as given by (5) and then using lemma 2 to estimate the relevant ratios of factorials we get that for $t = O(\sqrt{d})$:

$$
\begin{aligned}
s &\geq \exp\left(\frac{7}{8} \cdot \frac{d \log d}{32t} - \frac{1}{16} \cdot \frac{d \log d}{32t} - O(\sqrt{d})\right) \\
&\geq d^{\Omega\left(\frac{d}{t}\right)} \quad \left(\text{for } t = O(\sqrt{d})\right).
\end{aligned}
$$

This completes the proof of theorem 1. □

Remark 10. Comparison with the proof in [KS15a]. Our proof here is very similar to the proof in [KS15a] but there is one crucial difference that we want to highlight. Consider a term

$$T = Q_1 \cdot Q_2 \cdot \ldots \cdot Q_m,$$

where each Q_j has low arity. In the case of [KS15a], the Q_j's are of degree 1 and hence have only two homogeneous components - the constant term and the linear part. In that case, [KS15a] gives an explicit way of writing $T^{[d]}$ in terms of the homogeneous components of the Q_j's and this in turn yields an expression for $T_i^{[d]}$ as a sum of a relatively small ($\exp(\sqrt{d})$-many) number of products of low support (homogeneous) polynomials. When the Q_j's have larger degree then each Q_j itself has many more homogeneous components and it is not clear whether $T^{[d]}$ can now also be written as a sum of a small number of products of low support polynomials. Fortunately, we are able to get around this difficulty by the somewhat indirect argument given in section 3.4.

References

[AJMV98] Eric Allender, Jia Jiao, Meena Mahajan, and V. Vinay. Non-Commutative Arithmetic Circuits: Depth Reduction and Size Lower Bounds. *Theor. Comput. Sci.*, 209(1-2):47–86, 1998.

[AV08] Manindra Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *FOCS*, pages 67–75, 2008.

[BC15] Suman Bera and Amit Chakrabarti. A depth-five lower bound for iterated matrix multiplication. In *Proooceedings of 30th Conference on Computational Complexity*, 2015.

[FLMS14] Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. In *STOC*, pages 128–135, 2014.

[GKKS13a] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic circuits: A chasm at depth three. In *Foundations of Computer Science (FOCS)*, pages 578–587, 2013.

[GKKS13b] Ankit Gupta, Neeraj Kayal, Pritish Kamath, and Ramprasad Saptharishi. Approaching the chasm at depth four. In *Conference on Computational Complexity (CCC)*, 2013.

[Hya79] Laurent Hyafil. On the parallel evaluation of multivariate polynomials. *SIAM J. Comput.*, 8(2):120–123, 1979.

[JS82] Mark Jerrum and Marc Snir. Some exact complexity results for straight-line computations over semirings. *Journal of the ACM*, 29(3):874–897, 1982.

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

[KLSS14a] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An Exponential Lower Bound for Homogeneous Depth Four Arithmetic Formulas. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, 2014.

[KLSS14b] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. Superpolynomial lower bounds for depth-4 homogeneous arithmetic formulas. In *STOC*, pages 119–127, 2014.

[Koi12] Pascal Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theoretical Computer Science*, 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 *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 136–145, 2014.

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

[KS14c] Mrinal Kumar and Shubhangi Saraf. Superpolynomial lower bounds for general homogeneous depth 4 arithmetic circuits. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014, Copenhagen, Denmark, July 8-11, 2014, Proceedings, Part I*, pages 751–762, 2014.

[KS15a] Neeraj Kayal and Chandan Saha. Lower bounds for depth three circuits with small bottom fanin. In *Prooceedings of 30th Conference on Computational Complexity*, 2015.

[KS15b] Mrinal Kumar and Shubhangi Saraf. Sums of products of polynomials in few variables : lower bounds and polynomial identity testing. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:71, 2015.

[KSS14] Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. In *STOC*, pages 146–153, 2014.

[Nis91] Noam Nisan. Lower bounds for non-commutative computation (extended abstract). In *STOC*, pages 410–418, 1991.

[NW97] Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1997. Available at http://www.math.ias.edu/~avi/PUBLICATIONS/MYPAPERS/NW96/final.pdf.

[Raz09] Ran Raz. Multi-linear formulas for permanent and determinant are of superpolynomial size. *J. ACM*, 56(2), 2009.

[SW99] Amir Shpilka and Avi Wigderson. Depth-3 arithmetic formulae over fields of characteristic zero. In *IEEE Conference on Computational Complexity*, pages 87–, 1999. Available at http://eccc.hpi-web.de/report/1999/023/.

[SY10] Amir Shpilka and Amir Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5(3-4):207–388, 2010.

[Tav13] Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. In *MFCS*, pages 813–824, 2013.

[VSBR83] L.G. Valiant, S. Skyum, S. Berkowitz, and C. Rackoff. Fast parallel computation of polynomials using few processors. *SIAM Journal on Computing*, 12(4):641–644, 1983.