Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 98 (2012)

Approaching the chasm at depth four

Ankit Gupta
Microsoft Research India
t-ankitg@microsoft.com

Pritish Kamath
Microsoft Research India
t-pritk@microsoft.com

Neeraj Kayal
Microsoft Research India
neeraka@microsoft.com

Ramprasad Saptharishi
Chennai Mathematical Institute
ramprasad@cmi.ac.in

September 10, 2012

Abstract

Agrawal-Vinay [AV08] and Koiran [Koi12] have recently shown that an $\exp(\omega(\sqrt{n}\log^2 n))$ lower bound for depth four homogeneous circuits computing the permanent with bottom layer of × gates having fanin bounded by $\sqrt{n}$ translates to super-polynomial lower bound for a general arithmetic circuits computing the permanent. Motivated by this, we examine the complexity of computing the permanent and determinant via such homogeneous depth four circuits with bounded bottom fanin.

We show here that any homogeneous depth four arithmetic circuit with bottom fanin bounded by $\sqrt{n}$ computing the permanent (or the determinant) must be of size $\exp(\Omega(\sqrt{n}))$.

1 Introduction

Background. The most natural and intuitive way to compute a polynomial is via an arithmetic circuit. In this model the inputs are variables $x_1, x_2, \ldots, x_n$ and the computation is performed using the operations $+, \times$. We typically allow arbitrary constants from a field $\mathbb{F}$ on the incoming edges to a $+$ gate so that the output of a $+$ gate is an arbitrary $\mathbb{F}$-linear combination of its inputs. The complexity measures associated with circuits are size and depth, which capture the number of operations and the maximal distance between an input and the output.

Recall that the permanent is an n^2-variate homogeneous[1]polynomial of degree n defined as:

$$\mathsf{Perm}_n \quad = \quad \sum_{\sigma \in S_n} \prod_{i=1}^{n} x_{i\sigma(i)}$$

[1] A multivariate polynomial is said to be homogeneous if all its monomials have the same total degree.

ISSN 1433-8092

The permanent, by virtue of being complete for the class VNP (an algebraic analogue of the class NP, defined in [Val79]), occupies a central position in the study of the complexity of counting problems. The best known circuit for the permanent is actually a depth three homogeneous circuit of size $O(n^2 \cdot 2^n)$ and is called the Ryser's formula. Its illustrious sibling, the determinant, is widely believed to be comparatively easy, being complete for a subclass of VP (an algebraic analogue of P, also defined in [Val79]). It is conjectured (cf. [AV08]) that any arithmetic circuit computing the $n \times n$ permanent must be of $\exp(n)$ size. Meanwhile, the arithmetic complexity of computing the determinant equals $\tilde{O}(n^\omega)$, where ω is the exponent of matrix multiplication. Resolving the arithmetic complexity of computing the permanent and the determinant (i.e. determining the exponent of matrix multiplication) are two of the most fascinating open problems of our times.

Prior Work. Lower bounds have been obtained earlier for depth three arithmetic circuits (with some restrictions) and constant depth multilinear circuits. Specifically, Nisan and Wigderson [NW97] showed that any homogeneous depth three circuit computing the permanent (also the determinant) must be of exponential size. Following that, Grigoriev and Karpinski [GK98] showed that any depth three arithmetic circuit over a finite field computing the permanent (also the determinant) requires exponential size but proving lower bounds for depth three circuits over fields of characteristic zero (or even over the algebraic closure of a finite field) remains an outstanding open problem. In this direction Shpilka and Wigderson [SW01] proved quadratic lower bounds for depth three circuits over arbitrary fields (without the homogeneity restriction). Meanwhile, Raz [Raz09] showed that any multilinear formula computing the permanent (also the determinant) must be of superpolynomial size. Following this, Raz and Yehudayoff [RY08] proved exponential lower bounds for constant depth multilinear circuits.

The model. In this work, we focus our attention on depth four homogeneous[2] arithmetic circuits with bottom fanin bounded by a parameter t which we denote by $\Sigma\Pi\Sigma\Pi^{[\mathrm{hom}]}(t)$. A $\Sigma\Pi\Sigma\Pi^{[\mathrm{hom}]}(t)$ circuit computes a polynomial of the form

$$C \quad = \quad \sum_{i=1}^{s} (Q_{i1} \cdot Q_{i2} \cdot \ldots \cdot Q_{id}) \tag{1}$$

where each Q_{ij} is homogeneous polynomial of degree bounded by t, and every summand has the same degree. Our motivation for investigating representations of the form (1) stems from a recent result of Agrawal and Vinay [AV08], and a subsequent strengthening by Koiran [Koi12].

[2] An arithmetic circuit is said to be *homogeneous* if the polynomial computed at every internal node of the circuit is a homogeneous polynomial. It is a folklore result (cf. the survey by Shpilka and Yehudayoff [SY10]) that as far as computation by polynomial-sized arithmetic circuits of unbounded depth is concerned one can assume without loss of generality that the circuit is homogeneous. Specifically, if a homogeneous polynomial f of degree d can be computed by an (unbounded depth) arithmetic circuit of size s, then it can also be computed by a *homogeneous* circuit of size $O(d^2 \cdot s)$.

Theorem 1. *[AV08, Koi12] If there is a polynomial sized arithmetic circuit computing* Perm_n*, then there is a* $2^{O(\sqrt{n}\log^2 n)}$*-sized* $\Sigma\Pi\Sigma\Pi^{[\text{hom}]}(\sqrt{n})$*-circuit computing* Perm_n*.*

The contrapositive of the above statement is that it suffices to show a $2^{\omega(\sqrt{n}\log^2 n)}$ lower bound for $\Sigma\Pi\Sigma\Pi^{[\text{hom}]}(\sqrt{n})$ circuits computing the Perm_n to prove a super-polynomial circuit lower bound. Thus, a good enough lower bound for $\Sigma\Pi\Sigma\Pi^{[\text{hom}]}(\sqrt{n})$ circuits would imply super-polynomial lower bounds for Perm_n. In this paper, we give a lower bound for the permanent (or determinant) that comes very close to the above threshold.

Theorem 2. *Any* $\Sigma\Pi\Sigma\Pi^{[\text{hom}]}(t)$ *that computes* Perm_n *(or* Det_n*) must have size* $\exp\left(\Omega\left(\frac{n}{t}\right)\right)$*.*

Remark: The results of Agrawal-Vinay [AV08] and Koiran [Koi12] depth-reduce any polynomial sized circuit computing a degree n polynomial to a $\Sigma\Pi\Sigma\Pi^{[\text{hom}]}(t)$ formula with top fanin $\exp\left(\frac{n}{t}\log^2 n\right)$. The above theorem infact is a bound on the top fanin of $\Sigma\Pi\Sigma\Pi^{[\text{hom}]}(t)$ circuits computing the permanent or determinant. In Section 6, we prove a prove a generalization of Theorem 2 by extending the lowerbound for all circuits that are sums of arbitrary powers of $O\left(\frac{n}{t}\right)$-many degree t polynomials. Further, our proofs are completely elementary and self-contained but it is possible that using some more sophisticated theorems from algebraic geometry, the bounds that we obtain can be improved. Also, though the above theorem gives a lower bound for both the determinant and permanent, there is a subtle difference between the two and we discuss this in Section 7.

2 Basic Idea and Outline

Our key idea is to exploit the *shifted derivatives of a polynomial* - a notion that we now define. Let $\mathbb{F}$ be a field and $\mathbb{F}[\mathbf{x}]$ be the set of polynomials over $\mathbb{F}$ in the set of variables $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. For an n-tuple $\mathbf{i} = (i_1, i_2, \ldots, i_n) \in \mathbb{Z}^n_{\geq 0}$, $\mathbf{x}^{\mathbf{i}}$ denotes the monomial $(x_1^{i_1} \cdot x_2^{i_2} \cdot \ldots \cdot x_n^{i_n})$ which has degree $|\mathbf{i}| \stackrel{\text{def}}{=} (i_1 + i_2 + \ldots + i_n)$. $\partial^{\mathbf{i}} f$ denotes the partial derivative of f with respect to the monomial $\mathbf{x}^{\mathbf{i}}$,

$$\partial^{\mathbf{i}} f \quad \stackrel{\text{def}}{=} \quad \frac{\partial^{i_1}}{\partial x_1^{i_1}}\left(\frac{\partial^{i_2}}{\partial x_2^{i_2}}\left(\cdots\left(\frac{\partial^{i_n} f}{\partial x_n^{i_n}}\right)\cdots\right)\right).$$

For a finite subset of polynomials $S \subseteq \mathbb{F}[\mathbf{x}]$, the $\mathbb{F}$-span of S, denoted $\mathbb{F}\text{-span}\,(S)$, is the set of all possible $\mathbb{F}$-linear combinations of polynomials in S. i.e.

$$\mathbb{F}\text{-span}\,(S) \quad \stackrel{\text{def}}{=} \quad \left\{\sum_{i=1}^{|S|} \alpha_i \cdot f_i \ : \ \alpha_i \in \mathbb{F}, \quad f_i \in S\right\}.$$

With these notational preliminaries in hand, we are now ready to define our key concept.

Definition 1 (Shifted Derivatives). *Let* $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ *be a multivariate polynomial. The* span of the ℓ-shifted k-th order derivatives *of* f*, denoted* $\langle \boldsymbol{\partial}^{=k} f\rangle_{\leq \ell}$*, is defined as*

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

$\langle \boldsymbol{\partial}^{=k} f \rangle_{\leq \ell}$ *forms an* $\mathbb{F}$*-vector space and we denote by* $\dim(\langle \boldsymbol{\partial}^{=k} f \rangle_{\leq \ell})$ *the dimension of this space.*

Recent work in arithmetic complexity has shown how $\langle \boldsymbol{\partial}^{=k} f \rangle_{\leq \ell}$ can give insights into the structure and complexity of f in ways that are sometimes surprising and unexpected. Kayal [Kay12a] showed that $\langle \boldsymbol{\partial}^{=1} f \rangle_{\leq 1}$ yields a lie algebra that can help efficiently determine if f is equivalent (via an affine change of variables) to the permanent (or determinant). For $\ell = \infty$, note that $\langle \boldsymbol{\partial}^{=k} f \rangle_{\leq \ell}$ is precisely the ideal generated by the k-th order derivatives of f. Gupta, Kayal and Qiao [GKQ12] recently exploited the structure of $\langle \boldsymbol{\partial}^{=1} f \rangle_{\leq \infty}$ to devise an efficient reconstruction algorithm for random arithmetic formulas. Note that the dimension of partial derivatives employed by Nisan and Wigderson [NW97] in their lower bound proofs corresponds to looking at $\dim(\langle \boldsymbol{\partial}^{=k} f \rangle_{\leq 0})$. Closer to the present application, Kayal [Kay12b] showed how $\dim(\langle \boldsymbol{\partial}^{=k} f \rangle_{\leq \ell})$ (for suitably chosen ℓ and k) can be used to prove an exponential lower bound for representing a polynomial as a sum of powers of bounded degree polynomials. We show here that for suitably chosen values of ℓ and k, $\dim(\langle \boldsymbol{\partial}^{=k} f \rangle_{\leq \ell})$ is comparatively small when f is computed by a $\Sigma\Pi\Sigma\Pi^{[\text{hom}]}(t)$ circuit (Corollary 9). Meanwhile $\dim(\langle \boldsymbol{\partial}^{=k} \mathsf{Perm}_n \rangle_{\leq \ell})$ is relatively large (Corollary 16). This gives the lower bound.

Outline of the rest of the paper. We execute this idea in the rest of the paper as follows. In Section 4 we give an upper bound on $\langle \boldsymbol{\partial}^{=k} C \rangle_{\leq \ell}$ for C being a polynomial computed by a $\Sigma\Pi\Sigma\Pi^{[\text{hom}]}(t)$ circuit, i.e. when C is of the form given in equation (1). In Section 5, we give a lower bound estimate for $\dim(\langle \boldsymbol{\partial}^{=k} \mathsf{Perm}_n \rangle_{\leq \ell})$. We then combine these bounds to obtain a proof of our main theorem in Section 6. Finally, in Section 7, we conclude by discussing the possibility of improving the estimates for $\dim(\langle \boldsymbol{\partial}^{=k} \mathsf{Perm}_n \rangle_{\leq \ell})$ obtained here.

3 Preliminaries

Notation. We will use $[n]$ to denote the set $\{1, \cdots, n\}$ for any $n \geq 1$. $\mathbf{x}_n$ denotes the set of variables $\{x_1, x_2, \cdots, x_n\}$. However, when the context is clear, we would use just $\mathbf{x}$ instead of $\mathbf{x}_n$. Similarly for $\mathbf{y}$, $\mathbf{z}$, etc. We use $\boldsymbol{\partial}^{=k} f$ to denote the set of all k-th order partial derivatives of f. If $S \subseteq \mathbb{F}[\mathbf{x}]$, then,

$$\mathbf{x}^{\leq \ell} \cdot S \quad \stackrel{\text{def}}{=} \quad \left\{ \mathbf{x}^{\mathbf{i}} \cdot f \;:\; f \in S \text{ and } |\mathbf{i}| \leq \ell \right\}$$

Useful asymptotic estimates and inequalities. We now collect together some useful estimates for binomial coefficients that follow from Stirling's formula.

Definition 2. *The binary entropy function* H_2 *is defined as*

$$H_2(x) \quad = \quad -x \cdot \log_2(x) - (1-x) \cdot \log_2(1-x)$$

The natural-log version of the entropy function, denoted by H_e *is defined analogously as*

$$H_e(x) \quad = \quad -x \cdot \ln(x) - (1-x) \ln(1-x)$$

Lemma 3. *For any $0 < x < 1$, we have $x \ln \frac{1}{x} \leq H_e(x) \leq x \ln \frac{1}{x} + x$.*

Proposition 4 (Stirling's Formula, cf. [Rom])**.** $\ln(n!) \quad = \quad n \ln n - n + O(\ln n)$

Stirling's formula can be used to obtain the following estimates (proofs of which are in Appendix A).

Lemma 5. *Let $a(n)$, $f(n)$, $g(n) : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be integer valued function such that $(f+g) = o(a)$. Then,*

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

Lemma 6. *For any constants $\alpha \geq \beta > 0$,*

$$\ln \binom{\alpha n}{\beta n} \quad = \quad aH_e\left(\frac{\beta}{\alpha}\right) n + O(\ln n)$$

4 Upper bounding the dimension of shifted partials of $\Sigma\Pi\Sigma\Pi^{[\mathrm{hom}]}(t)$ circuits

In this section we give an upper bound on $\dim(\langle \boldsymbol{\partial}^{=k} C\rangle_{\leq \ell})$ when C is computed by a depth four circuit, i.e. C is of the form given in equation (1). We begin by noting that $\dim(\langle \boldsymbol{\partial}^{=k} f\rangle_{\leq \ell})$ is sub-additive.

Proposition 7. *For all $k, \ell \geq 0$, we have $\dim(\langle \boldsymbol{\partial}^{=k}(f+g)\rangle_{\leq \ell}) \quad \leq \quad \dim(\langle \boldsymbol{\partial}^{=k} f\rangle_{\leq \ell}) + \dim(\langle \boldsymbol{\partial}^{=k} g\rangle_{\leq \ell})$.*

Proof. By linearity of partial derivatives, we have $\mathbf{x}^{\mathbf{i}} \cdot \partial^{\mathbf{j}}(f+g) = \mathbf{x}^{\mathbf{i}} \cdot \partial^{\mathbf{j}} f + \mathbf{x}^{\mathbf{i}} \cdot \partial^{\mathbf{j}} g$. Hence,

$$\mathbf{x}^{\leq \ell} \cdot \boldsymbol{\partial}^{=k}(f+g) \quad \subseteq \quad \mathbb{F}\text{-span}\left(\left(\mathbf{x}^{\leq \ell} \cdot \boldsymbol{\partial}^{=k} f\right) \cup \left(\mathbf{x}^{\leq \ell} \cdot \boldsymbol{\partial}^{=k} g\right)\right)$$

The proposition follows. □

Let C be a depth-4 circuit computing a polynomial of the form[3]

$$C \quad = \quad \sum_{i=1}^{s} Q_{i1}^{e_{i1}} \cdot Q_{i2}^{e_{i2}} \ldots Q_{id}^{e_{id}} \quad \text{where } \deg(Q_{ij}) \leq t.$$

By Proposition 7, it suffices to understand the growth of $\dim(\langle \boldsymbol{\partial}^{=k} C\rangle_{\leq \ell})$ of a single term $(Q_1^{e_1} \ldots Q_d^{e_d})$.

Proposition 8. *If $f = Q_1^{e_1} \ldots Q_d^{e_d}$ where each $Q_i \in \mathbb{F}[\mathbf{x}_N]$ is a polynomial of degree bounded by t. Then, for any $\ell \geq 0$,*

$$\dim(\langle \boldsymbol{\partial}^{=k} f\rangle_{\leq \ell}) \quad \leq \quad \binom{d+k-1}{k}\binom{N+(t-1)k+\ell}{N}$$

[3] This is slightly more general than the form described in Equation (1).

Proof. Let $\mathbf{j} \in \mathbb{Z}_{\geq 0}^{d}$ be any d-tuple satisfying $|\mathbf{j}| = k$. We observe by repeated use of the product rule that

$$\partial^{\mathbf{j}}\left(Q_1^{e_1}\dots Q_d^{e_d}\right) \quad = \quad \sum_{\mathbf{j}_1+\cdots+\mathbf{j}_d=\mathbf{j}} \left(\partial^{\mathbf{j}_1} Q_1^{e_1}\right)\dots\left(\partial^{\mathbf{j}_d} Q_d^{e_d}\right)$$

Hence, each term in the above sum can be written as $\left(Q_1^{e_1-j_1}\dots Q_d^{e_d-j_d}\right)\cdot\tilde{Q}$ where $\sum j_i = k$ and $\tilde{Q}$ has degree at most $(tk-k)$. Thus, every element of $\mathbf{x}^{\leq \ell}\boldsymbol{\partial}^{=k}(Q_1^{e_1}\cdots Q_d^{e_d})$ can be written as a linear combination of $\left(Q_1^{e_1-j_1}\dots Q_d^{e_d-j_d}\right)\mathbf{x}^{\mathbf{r}}$ where $\sum j_i = k$ and $\mathbf{x}^{\mathbf{r}}$ is a monomial of degree at most $\ell+(t-1)k$. The total number of monomials of degree at most $\ell+(t-1)k$ over N variables is $\binom{N+(t-1)k+\ell}{N}$, and the total number of choices for $j_1+\cdots+j_d = k$ is $\binom{d+k-1}{k}$. Hence we obtain,

$$\dim(\langle\boldsymbol{\partial}^{=k}(Q_1^{e_1}\cdots Q_d^{e_d})\rangle_{\leq\ell}) \quad \leq \quad \binom{d+k-1}{k}\binom{N+(t-1)k+\ell}{N}$$

□

The following corollary follows directly from the above observation via sub-additivity.

Corollary 9. *If $C = \sum_{i=1}^{s}\prod_{j=1}^{d} Q_{ij}^{e_{ij}}$ where each $Q_{ij} \in \mathbb{F}[\mathbf{x}_N]$ is a polynomial of degree bounded by t, then for any $k \leq d$*

$$\dim(\langle\boldsymbol{\partial}^{=k}(C)\rangle_{\leq\ell}) \quad \leq \quad s\cdot\binom{d+k-1}{k}\binom{N+(t-1)k+\ell}{N}$$

In the next section we give a reasonable lower bound for $\dim(\langle\boldsymbol{\partial}^{=k}(\mathsf{Perm}_n)\rangle_{\leq\ell})$ for suitable choice of parameters k and ℓ.

5 Lower a bounding the dimension of shifted partials of the Permanent

Reducing dimension computation to counting leading monomials. In this section, we shall present a lower bound for $\dim(\langle\boldsymbol{\partial}^{=k}(\mathsf{Perm}_n)\rangle_{\leq\ell})$. Let $\succ$ be any admissible monomial ordering[4]. Recall that the leading monomial of a polynomial $f \in \mathbb{F}[\mathbf{x}]$, denoted $\mathsf{LM}(f)$, is the largest monomial $\mathbf{x}^{\mathbf{i}}$ under the ordering $\succ$.

Proposition 10. *Let $S \subseteq \mathbb{F}[\mathbf{x}]$ be any finite set of polynomials. Then*

$$\dim(\mathbb{F}\text{-}span\,(S)) = \#\{\mathsf{LM}(f)\ :\ f \in \mathbb{F}\text{-}span\,(S)\}.$$

The proof is a simple application of Gaussian elimination. As a corollary we obtain

[4] For more on monomial orderings and their applications in algebraic geometry, we refer the interested reader to Chapter 2 of the text by Cox, Little and O'Shea [CLO07]

Corollary 11. *For any polynomial $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ we have*

$$\dim(\langle \boldsymbol{\partial}^{=k} f\rangle_{\leq \ell}) \quad \geq \quad \#\{\mathbf{x}^{\mathbf{i}} \cdot \mathsf{LM}(\boldsymbol{\partial}^{\mathbf{j}} f) \ : \ \mathbf{i}, \mathbf{j} \in \mathbb{Z}_{\geq 0}^{|\mathbf{x}|}, \ |\mathbf{i}| \leq \ell \quad \text{and} \quad |\mathbf{j}| = k\}$$

The lower bound given by this corollary is usually a severe underestimate but fortunately even this will suffice for our purpose for the case when $f = \mathsf{Perm}_n$.

Reduction to counting monomials with increasing subsequences. Let's fix $\succ$ to be the lexicographic monomial ordering induced by the following ordering on the variables: $x_{11} \succ \cdots \succ x_{1n} \succ x_{21} \succ \cdots \succ x_{nn}$. Now note that any partial derivative of Perm_n is just the corresponding permanental minor (or just 'P-minor'). Hence by the above corollary we have

$$\dim(\langle \boldsymbol{\partial}^{=k}(\mathsf{Perm}_n)\rangle_{\leq \ell}) \quad \geq \quad \#\left\{\mathbf{x}^{\mathbf{i}} \cdot \mathsf{LM}(M) \ : \ \begin{array}{l} \mathbf{x}^{\mathbf{i}} \text{ is a monomial of degree at most } \ell \text{ and} \\ M \text{ is an } (n-k) \times (n-k) \text{ P-minor} \end{array}\right\}$$

Now note that the leading monomial under $\succ$ of any $(n-k) \times (n-k)$ P-minor M is just the product of the variables along the principal diagonal of M. Now if the variables along the principal minor of M are $(x_{i_1 j_1}, \cdots, x_{i_{n-k} j_{n-k}})$ then the indices satisfy

$$i_1 < i_2 < \ldots < i_{n-k} \quad \text{and } j_1 < j_2 < \ldots < j_{n-k}$$

This naturally leads to the following definition.

Definition 3. *We shall refer to a sequence of variables $(x_{i_1 j_1}, \cdots, x_{i_t j_t})$ as an* increasing sequence *if the indices satisfy*

$$i_1 < i_2 < \ldots < i_t \quad \text{and } j_1 < j_2 < \ldots < j_t.$$

We will say that a monomial $A = \mathbf{x}^{\mathbf{j}}$ contains an increasing sequence of length t *if there exists an increasing sequence $(x_{i_1 j_1}, \cdots, x_{i_t j_t})$ wherein every variable $x_{i_r j_r}$ ($r \in [t]$) divides A.*

In this terminology we would then say that the leading monomial of any $(n-k) \times (n-k)$ P-minor is exactly the product of the variables in an increasing sequence of length $(n-k)$. Consequently for any P-minor M of size $(n-k)$ we have that $\mathbf{x}^{\mathbf{i}} \cdot \mathsf{LM}(M)$ contains an increasing sequence of length $(n-k)$. Conversely, every monomial of degree at most $(n-k+\ell)$ that contain an increasing sequence of length $(n-k)$ can be written as the leading monomial of $\mathbf{x}^{\mathbf{i}} \cdot M$ for some monomial $\mathbf{x}^{\mathbf{i}}$ of degree at most ℓ and a $(n-k) \times (n-k)$ P-minor M. Hence we have:

Corollary 12. $\dim(\langle \boldsymbol{\partial}^{=k}(\mathsf{Perm}_n)\rangle_{\leq \ell})$ *is lower bounded by the number of distinct monomials of degree at most $(n-k+\ell)$ over n^2 variables that contain an increasing sequence of length $(n-k)$.*

In order to count the number of monomials of degree bounded by $n-k+\ell$ that contain an increasing sequence, we shall restrict ourselves to a very *small set* of variables to contribute the increasing sequence, and "fill-up" the remaining degree using the other variables. The "small set" that we consider here is just two diagonals – the principal diagonal and the one above it.

5.1 Restricting to two diagonals

We shall focus on the variables $D_{2,n} = \{x_{ii} \ : \ 1 \leq i \leq n\} \cup \left\{x_{i(i+1)} \ : \ 1 \leq i \leq n-1\right\}$.

Lemma 13. *Fix parameters $n, r, m \geq 0$, and let $S(n,r,m)$ be the number of size-m subsets of $D_{2,n}$ that contain an increasing sequence of length r. Then for any $d, m > 0$, the number of monomials of degree bounded by d containing an increasing sequence of length r is lower bounded by $S(n,r,m) \cdot \binom{n^2-2n+d+1}{n^2-2n+m+1}$.*

Proof. For any monomial, define the *support in $D_{2,n}$* as the set of variables in $D_{2,n}$ that divide it. Note that for any $S \subseteq D_{2,n}$, there are exactly $\binom{n^2-2n+1+|S|+d-|S|}{n^2-2n+1+|S|}$ monomials of degree d whose support in $D_{2,n}$ is S. Certainly two monomials having different supports in $D_{2,n}$ are not equal. Summing up over all sets size-m sets containing an increasing sequence of length r gives the required lowerbound. □

We now are faced with the job of lowerbounding the number of size-s subsets of $D_{2,n}$ that contain an increasing sequence of length r. To do that, we shall first pick an increasing sequence of length r, and add more terms to form a size-m subset. In the process, the subset could have several increasing sequences of length r. To avoid double counting, we shall ensure that the *leading increasing sequence* (under the lexicographic ordering described in Section 5) remains invariant.

Lemma 14. *The number of length r increasing sequences in contained in $D_{2,n}$ is exactly $\binom{2n-r}{r}$.*

Proof. Consider the $(2n-1)$ variables in $D_{2,n}$ in the sequence $x_{11}, x_{12}, x_{21}, \ldots, x_{nn}$. Picking an increasing sequence of length r is the same as picking r of the $(2n-1)$ variables such that no two adjacent variables (in the above order) are chosen. This can be thought as distributing the $(2n-r-1)$ variables that won't be picked such that there is at least one between any two variables that are picked, and this is exactly equal to

$$\binom{(2n-r-1-(r-1))+(r+1)-1}{(r+1)-1} = \binom{2n-r}{r}$$

□

Lemma 15. *For every $n, r, m \geq 0$, we have that $S(n,r,m) \geq \binom{2n-r}{r}\binom{r-1}{m-r}$.*

Proof. For any variable x_{ij}, define its *companions* to be the variables to its right in the same row, or below it in the same column, i.e. $\{x_{ij'} \ : \ j' > j\} \cup \{x_{i'j} \ : \ i' > i\}$.
Fix an increasing sequence $Q = \{x_{i_1j_1}, \ldots, x_{i_rj_r}\} \subseteq D_{2,n}$. Let Q' be the set of all companions of variables in Q which are in $D_{2,n}$. The key observation is that adding elements of Q' to Q does not alter the leading increasing sequence. For any increasing sequence that uses elements of Q', replacing every $x_{i'j'} \in Q'$ by the corresponding $x_{ij} \in Q$ for which it is a companion for yields a "higher" increasing sequence. Hence adding any subset $T \subseteq Q'$ to Q does not alter the leading increasing sequence.

Note that every element of $D_{2,n}$ besides x_{nn} has exactly one companion in $D_{2,n}$. Hence, if Q is a length r increasing sequence, the set of companions its Q' has cardinality at least $(r-1)$. Since we are interested in size m subsets, there are at least $\binom{r-1}{m-r}$ ways of augmenting Q with a size-$(m-r)$ subset of Q'. By Lemma 14 there are $\binom{2n-r}{r}$ choices of Q to start with, the bound follows. □

By setting $d = \ell + n - k$ and $r = n - k$ in Lemma 15 and using Lemma 13 with these parameters, we get the following lowerbound for $\dim(\langle \boldsymbol{\partial}^{=k}\mathsf{Perm}_n\rangle_{\leq \ell})$ via Corollary 12.

Corollary 16. *For every $n, m, k, \ell \geq 0$,*

$$\dim(\langle \boldsymbol{\partial}^{=k}\mathsf{Perm}_n\rangle_{\leq \ell}) \;\geq\; \binom{n+k}{2k} \cdot \binom{n-k-1}{m-n+k} \cdot \binom{n^2-n+\ell-k+1}{n^2-2n+m+1}$$

6 Putting it all together

We are now ready to prove the main theorem, which is a stronger form of Theorem 2.

Theorem 17. *Let $t : \mathbb{Z}_{\geq 0} \to \mathbb{Z}_{\geq 0}$ be any increasing function such that $t(n) = o(n)$. Suppose C is a circuit of the form $C = \sum_{i=1}^{s} Q_{i1}^{e_{i1}} \cdots Q_{id}^{e_{id}}$ where each Q_{ij} is a polynomial of degree bounded by t, and $d = cn/t$ for some constant c. If C computes the polynomial Perm_n, then $s \geq \exp\left(\Omega\left(\frac{n}{t}\right)\right)$.*

Proof. From Corollary 9, $\dim(\langle \boldsymbol{\partial}^{=k}C\rangle_{\leq \ell})$ can be upper bounded as

$$\dim(\langle \boldsymbol{\partial}^{=k}(C)\rangle_{\leq \ell}) \quad \leq \quad s \cdot \binom{d+k-1}{k}\binom{n^2+\ell+(t-1)k}{n^2} \tag{2}$$

Also, Corollary 16 gives a lower bound for $\dim(\langle \boldsymbol{\partial}^{=k}\mathsf{Perm}_n\rangle_{\leq \ell})$ (for any choice of m):

$$\dim(\langle \boldsymbol{\partial}^{=k}\mathsf{Perm}\rangle_{\leq \ell}) \quad \geq \quad \binom{n+k}{2k}\binom{n-k-1}{m-n+k}\binom{n^2-n+\ell-k+1}{n^2-2n+m+1} \tag{3}$$

Both these equations imply that

$$s \quad \geq \quad \frac{\binom{n+k}{2k}\binom{n-k-1}{m-n+k}\binom{n^2-n+\ell-k+1}{n^2-2n+m+1}}{\binom{d+k-1}{k}\binom{n^2+\ell+(t-1)k}{n^2}}$$

We shall set parameters as $\ell = n^2 t$, $m = 2n - (n/t)$ and $k = \varepsilon(n/t)$ (for an $\varepsilon > 0$ that shall be chosen shortly). The proofs of the following estimates for binomial coefficients are straightforward applications of Lemma 5 and Lemma 6, and are presented in the Appendix A.

Claim 18. *For the above choice of parameters:*

(a) $\ln\binom{n+k}{2k} = 2\varepsilon\left(\frac{n}{t}\right)\left(\ln\left(\frac{t}{2\varepsilon}\right)+1\right) \pm O\left(\frac{n}{t^2}\right)$

(b) $\ln \binom{n-k-1}{m-n+k} = (1-2\varepsilon)\left(\frac{n}{t}\right)\left(\ln\left(\frac{t}{1-2\varepsilon}\right)+1\right) \pm O\left(\frac{n}{t^2}\right)$

(c) $\ln \binom{cn/t+k-1}{k} = (c+\varepsilon)H_e\left(\frac{\varepsilon}{c+\varepsilon}\right)\cdot\left(\frac{n}{t}\right) \pm O\left(\ln n\right)$

(d) $\ln \frac{\binom{n^2-n+\ell-k+1}{n^2-2n+m+1}}{\binom{n^2+\ell+(t-1)k}{n^2}} = -\left(\frac{n}{t}\right)\ln t - (1+\varepsilon)\left(\frac{n}{t}\right) \pm O\left(\frac{n}{t^2}\right)$

Using this, we get

$$\begin{aligned}\ln s &\geq \left(2\varepsilon\ln\left(\frac{1}{2\varepsilon}\right) + (1-2\varepsilon)\ln\left(\frac{1}{1-2\varepsilon}\right) - \varepsilon - (c+\varepsilon)H_e\left(\frac{\varepsilon}{c+\varepsilon}\right)\right)\left(\frac{n}{t}\right) \ \pm\ O\left(\frac{n}{t^2}\right)\\ &= \left(H_e(2\varepsilon) - \varepsilon - (c+\varepsilon)H_e\left(\frac{\varepsilon}{c+\varepsilon}\right)\right)\left(\frac{n}{t}\right) \ \pm\ O\left(\frac{n}{t^2}\right)\end{aligned}$$

which after an application of Lemma 3 yields

$$\begin{aligned}\ln s &\geq \left(2\varepsilon\ln\frac{1}{2\varepsilon} - \varepsilon - \varepsilon\ln\left(\frac{c+\varepsilon}{\varepsilon}\right) - \varepsilon\right)\\ &= \left(\varepsilon\ln\frac{1}{\varepsilon} - \varepsilon\ln(4e^2(c+1))\right)\left(\frac{n}{t}\right) \pm +O\left(\frac{n}{t^2}\right)\end{aligned}$$

Choosing ε small enough gives $\ln s = O\left(\frac{n}{t}\right)$, i.e. $s \geq \exp\left(\Omega\left(\frac{n}{t}\right)\right)$ as claimed □

Remark. *Though the above theorem is stated for any increasing function $t(n)$, the result also holds when t is a constant. The choice of parameters in that case would be $\ell = n^2$, $m = 3(n-k)/2$ and $k = \varepsilon n$. Using similar estimates on the binomial coefficients, it can be shown that* $\log s = \Omega(n)$ *by choosing a small enough $\varepsilon > 0$.*

In order to apply the above theorem to prove Theorem 2, we need a bound on d (the number of Q_{ij}'s in each summand). Since several of the Q_{ij}'s could have degree smaller than t, it is possible that d is much larger than $\frac{n}{t}$. However, since Theorem 17 gives a bound on the top fanin, we can multiply the Q_{ij}'s of low degree to ensure that each of them (except perhaps one) has degree at least $t/2$. By this, $d \leq 2n/t + 1$ and we can then apply Theorem 17 to complete the proof of Theorem 2.

7 Discussion

The proof of Theorem 2 remains valid if we replace every occurrence of Perm_n by Det_n but there turns out to be a very interesting distinction between these two polynomials with respect to the dimension of their shifted partial derivatives. In the particular case of the determinant, Corollary 12 can be strengthened to say that the number of monomials of degree at most $n-k+\ell$ with an increasing sequence of length $(n-k)$ is not just a lower bound but is exactly equal to $\dim(\langle \boldsymbol{\partial}^{=k}(\mathsf{Det}_n)\rangle_{\leq \ell})$. This follows from the following powerful result on gröbner bases of determinantal ideals which has been proved independently by Sturmfels[Stu90], Narasimhan [Nar86] and Caniglia, Guccione and Guccione [CGG90].

Theorem 19 ([Stu90], [Nar86], [CGG90])**.** *Let $\succ$ be the lexicographic ordering on monomials defined in Section 5. Then the set of all order $r \times r$ minors of Det_n is the reduced gröbner basis for the ideal generated by them under the monomial ordering $\succ$.*

It is known that the set of 2×2 permanental minors *do not* form a gröbner basis for the ideal they generate. Thus it is presumable that $\dim(\langle \boldsymbol{\partial}^{=k}(\mathsf{Perm}_n)\rangle_{\leq \ell})$ is much larger compared to the determinant. We conclude with the following conjecture.

Conjecture 20. *There exists choices for $\ell, k \geq 0$ such that $\dim(\langle \boldsymbol{\partial}^{=k}(\mathsf{Perm}_n)\rangle_{\leq \ell})$ is super-polynomially larger (in n) than $\dim(\langle \boldsymbol{\partial}^{=k}(\mathsf{Det}_n)\rangle_{\leq \ell})$.*

Acknowledgments. We would like to thank Ravi Kannan and Satya Lokam for useful discussions and providing some relevant references.

References

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

[CGG90] L. Caniglia, J. A. Guccione, and J. J. Guccione. Ideals of generic minors. *Commutative Algebra*, 18:2633–2640, 1990.

[CLO07] D.A. Cox, J.B. Little, and D. O'Shea. *Ideals, Varieties and Algorithms.* Undergraduate texts in mathematics. Springer, 2007.

[GK98] Dima Grigoriev and Marek Karpinski. An exponential lower bound for depth 3 arithmetic circuits. In *STOC*, pages 577–582, 1998.

[GKQ12] Ankit Gupta, Neeraj Kayal, and Youming Qiao. Random arithmetic formulas can be reconstructed efficiently. Technical report, Electronic Colloquium on Computational Complexity (ECCC), 2012.

[Kay12a] Neeraj Kayal. Affine projections of polynomials. In *STOC*, pages 643–662, 2012.

[Kay12b] Neeraj Kayal. An exponential lower bound for the sum of powers of bounded degree polynomials. Technical report, Electronic Colloquium on Computational Complexity (ECCC), 2012.

[Koi12] Pascal Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theoretical Computer Science*, 448:56–65, 2012.

[Nar86] H. Narasimhan. The irreducibility of ladder determinantal varieties. *Journal of Algebra*, 102:162–185, 1986.

[NW97] N. Nisan and A. Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1997.

[Raz09] R. Raz. Multi-linear formulas for permanent and determinant are of super-polynomial size. *Journal of the Association for Computing Machinery*, 56(2), 2009.

[Rom] Dan Romik. Stirlings approximation for n!: The ultimate short proof? *The American Mathematical Monthly*, 107(6):556557.

[RY08] R. Raz and A. Yehudayoff. Lower bounds and separations for constant depth mutilinear circuits. In *Proceedings of the 23rd IEEE Annual Conference on Computational Complexity*, pages 128–139, 2008.

[Stu90] Bernd Sturmfels. Gröbner bases and stanley decompositions of determinantal rings. *Mathematische Zeitschrift*, 209:137–144, 1990.

[SW01] A. Shpilka and A. Wigderson. Depth-3 arithmetic circuits over fields of characteristic zero. *Computational Complexity*, 10(1):1–27, 2001.

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

[Val79] Leslie G. Valiant. Completeness Classes in Algebra. In *STOC*, pages 249–261, 1979.

A Proofs of binomial estimates

Lemma 21. *Let $a(n)$, $f(n)$, $g(n) : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be integer valued function such that $(f+g) = o(a)$. Then,*

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

Proof.

$$\frac{(a+f)!}{(a-g)!} = (a+f)(a+f-1)\dots(a-g)$$

$$\implies \quad a^{f+g}\left(1-\frac{g}{a}\right)^{f+g} \quad \leq \quad \frac{(a+f)!}{(a-g)!} \quad \leq \quad a^{f+g}\left(1+\frac{f}{a}\right)^{f+g}$$

$$\implies \quad (f+g)\ln\left(1-\frac{g}{a}\right) \quad \leq \quad \ln\frac{(a+f)!}{(a-g)!} - (f+g)\ln a \quad \leq \quad (f+g)\ln\left(1+\frac{f}{a}\right)$$

Using the fact that $\frac{x}{1+x} \leq \ln(1+x) \leq x$ for $x > -1$, it is easy to see that both the LHS and RHS are bounded by $O\left(\frac{(f+g)^2}{a}\right)$. □

Lemma 22. *For any constants $\alpha \geq \beta > 0$,*

$$\ln \binom{\alpha n}{\beta n} \quad = \quad aH_e\left(\frac{\beta}{\alpha}\right) n + O(\ln n)$$

Proof. By Stirling's approximation (Proposition 4),

$$\begin{aligned}\ln \frac{(\alpha n)!}{(\beta n)!((\alpha-\beta)n)!} &= (\alpha n)\ln(\alpha n) - \alpha n - (\beta n)\ln(\beta n) + \beta n \\ &\qquad - (\alpha-\beta)n\ln((\alpha-\beta)n) + (\alpha-\beta)n + O(\ln n) \\ &= n(\alpha\ln\alpha - \beta\ln\beta - (\alpha-\beta)\ln(\alpha-\beta)) + O(\ln n) \\ &= \alpha n \cdot H_e\left(\frac{\beta}{\alpha}\right) + O(\ln n)\end{aligned}$$

□

Claim 23. *Suppose $\ell = n^2 t$, $m = 2n - \frac{n}{t}$ and $k = \varepsilon\left(\frac{n}{t}\right)$ where t is an increasing function of n such that $t = o(n)$, and $\varepsilon > 0$ is a constant. Then,*

(a) $\ln \binom{n+k}{2k} = 2\varepsilon\left(\frac{n}{t}\right)\left(\ln\left(\frac{t}{2\varepsilon}\right)+1\right) \pm O\left(\frac{n}{t^2}\right)$

(b) $\ln \binom{n-k-1}{m-n+k} = (1-2\varepsilon)\left(\frac{n}{t}\right)\left(\ln\left(\frac{t}{1-2\varepsilon}\right)+1\right) \pm O\left(\frac{n}{t^2}\right)$

(c) $\ln \binom{cn/t+k-1}{k} = (c+\varepsilon)H_e\left(\frac{\varepsilon}{c+\varepsilon}\right)\cdot\left(\frac{n}{t}\right) \pm O\left(\ln n\right)$

(d) $\ln \frac{\binom{n^2-n+\ell-k+1}{n^2-2n+m+1}}{\binom{n^2+\ell+(\sqrt{n}-1)k}{n^2}} = -\left(\frac{n}{t}\right)\ln t - (1+\varepsilon)\left(\frac{n}{t}\right) \pm O\left(\frac{n}{t^2}\right)$

Proof.

(a) $\binom{n+k}{2k} = \frac{(n+k)!}{(n-k)!}\cdot\frac{1}{(2k)!}$. Since $k = o(n)$, using Lemma 21 and Lemma 22 gives

$$\begin{aligned}\ln\binom{n+k}{2k} &= 2k\ln n - (2k)\ln(2k) + 2k \ \pm\ O\left(\frac{k^2}{n}\right) \\ &= 2\varepsilon\left(\frac{n}{t}\right)\left(\ln\left(\frac{t}{2\varepsilon}\right)+1\right) \ \pm\ O\left(\frac{n}{t^2}\right)\end{aligned}$$

(b) $\binom{n-k-1}{m-n+k} = \frac{(n-k-1)!}{(n-(n/t)+k)!}\cdot\frac{1}{((n/t)-2k-1)!}$. Since $(n/t)+2k = o(n)$, Lemma 21 and Lemma 22 asserts that

$$\begin{aligned}\ln\binom{n-k-1}{m-n+k} &= \left(\frac{n}{t}-2k\right)\left(\ln n - \ln\left(\frac{n}{t}-2k\right)+1\right) \ \pm\ O\left(\frac{\left(\frac{n}{t}-2k\right)^2}{n}\right) \\ &= (1-2\varepsilon)\left(\frac{n}{t}\right)\left(\ln\left(\frac{t}{1-2\varepsilon}\right)+1\right) \ \pm\ O\left(\frac{n}{t^2}\right)\end{aligned}$$

(c) Follows directly from Lemma 22.

(d)

$$\frac{\binom{n^2+\ell-n-k+1}{n^2-2n+m+1}}{\binom{n^2+\ell+(t-1)k}{n^2}} = \frac{(n^2+\ell-n-k+1)!}{(n^2+\ell+(t-1)k)!} \cdot \frac{(n^2)!}{(n^2-(n/t)+1)!} \cdot \frac{(\ell+(t-1)k)!}{(\ell+(n/t)-n-k)!}$$

Using the fact that $tk+n = o(n^2+\ell)$, Lemma 21 can be applied on each of these ratios to give

$$\begin{aligned}
\frac{\binom{n^2+\ell-n-k+1}{n^2-2n+m+1}}{\binom{n^2+\ell+(t-1)k}{n^2}} &= \frac{1}{(n^2+\ell)^{tk+n-1}} \cdot (n^2)^{(n/t)-1} \cdot \ell^{tk+n-(n/t)} \cdot \exp\left(\frac{n}{t^2}\right) \\
&= \frac{1}{\left(1+\frac{n^2}{\ell}\right)^{tk+n+1}} \cdot \left(\frac{n^2}{\ell}\right)^{(n/t)-1} \cdot \exp\left(\frac{n}{t^2}\right) \\
\implies \ln \frac{\binom{n^2+\ell-n-k+1}{n^2-2n+m+1}}{\binom{n^2+\ell+(t-1)k}{n^2}} &= -(tk+n)\ln\left(1+\frac{1}{t}\right) - \left(\frac{n}{t}\right)\ln t \ \pm\ O\left(\frac{n}{t^2}\right) \\
&= -(1+\varepsilon)\left(\frac{n}{t}\right) - \left(\frac{n}{t}\right)\ln t \ \pm\ O\left(\frac{n}{t^2}\right)
\end{aligned}$$

□