A selection of lower bounds for arithmetic circuits

Neeraj Kayal and Ramprasad Saptharishi

To Somenath Biswas, on his 60th Birthday

Mathematics Subject Classification (2010). Primary 68Q25, 68W30; Secondary 12E05.

Keywords. arithmetic circuits, lower bounds, determinant, permanent.

> *It is convenient to have a measure of the amount of work involved in a computing process, even though it be a very crude one ... We might, for instance, count the number of additions, subtractions, multiplications, divisions, recordings of numbers,...*
>
> from *Rounding-off errors in matrix processes,*
>
> Alan M. Turing, 1948.

1. Introduction

Polynomials originated in classical mathematical studies concerning geometry and solutions to systems of equations. They feature in many classical results in algebra, number theory and geometry - e.g. in Galois and Abel's resolution of the solvability via radicals of a quintic, Lagrange's theorem on expressing every natural number as a sum of four squares and the impossibility of trisecting an angle (using ruler and compass). In modern times, computer scientists began to investigate as to what functions can be (efficiently) computed. Polynomials being a natural class of functions, one is naturally lead to the following question:

> *What is the optimum way to compute a given (family of) polynomial(s)?*

Now the most natural way to compute a polynomial $f(x_1, x_2, \ldots, x_n)$ over a field $\mathbb{F}$ is to start with the input variables $x_1, x_2, \ldots, x_n$ and then apply a sequence of basic operations such as additions, subtractions and multiplications[1] in order to obtain the desired polynomial f. Such a computation is called a straight line program. We often represent such a straight-line program graphically as an arithmetic circuit - wherein the overall computation corresponds to a directed acylic graph whose source nodes are labelled with the input variables $\{x_1, x_2, \ldots, x_n\}$ and the internal nodes are labelled with either $+$ or $\times$ (each internal node corresponds to one computational step in the straight-line program). We typically allow arbitrary constants from the underlying field on the incoming edges of a $+$ gate so that a $+$ gate can in fact compute an arbitrary $\mathbb{F}$-linear combination of its inputs. The complexity of the computation corresponds to the number of operations, also called the size of the corresponding arithmetic circuit. With arithmetic circuits being the relevant model, the informal question posed above can be formalized by defining the optimal way to compute a given polyomial as the smallest arithmetic circuit in terms of the size that computes it. While different aspects of polynomials have been studied extensively in various areas of mathematics, what is unique to computer science is the endeavour to prove upper and lower bounds on the size of arithmetic circuits computing a given (family of) polynomials. Here we give a biased survey of this area, focusing mostly on lower bounds. Note that there are already two excellent surveys of this area - one by Avi Wigderson [Wig02] and the other by Amir Shpilka and Amir Yehudayoff [SY10][2]. Our intention in writing the survey is the underlying hope that revisiting and assimilating the known results pertaining to circuit lower bounds will in turn help us make progress on this beautiful problem. Consequently we mostly present here those results which we for some reason felt we did not understand comprehensively enough. We conclude with ~~a~~ some recent lower bound results for homogeneous bounded depth formulas. Some notable lower bound results that we are unable to present here due to space and time constraints are as follows. A quadratic lower bound for depth three circuits by Shpilka and Wigderson [SW01], for bounded occur bounded depth formulas by Agrawal, Saha, Saptharishi and Saxena [ASSS12] and the $n^{1+\Omega(1/r)}$ lower bound for circuits of depth r by Raz [Raz10].

Overview. The state of affairs in arithmetic complexity is such that despite a lot of attention we still have only modest lower bounds for general circuits and formulas. In order to make progress, recent work has focused on restricted subclasses. We first present the best known lower bound for general circuits due to Baur and Strassen [BS83], and a lower bound for formulas due to Kalorkoti [Kal85]. The subsequent lower bounds that we present follow a common roadmap and we articulate this in Section 4, and present some simple

[1] One can also allow more arithmetic operations such as division and square roots. It turns out however that one can efficiently simulate any circuit with divisions and ~~squareroots~~ square roots by another circuit without these operations while incurring only an ~~$O(d)$-factor loss~~ polynomial factor increase in size.

[2] A more specialized survey by Chen, Kayal and Wigderson [CKW11] focuses on the applications of partial derivatives in understanding the structure and complexity of polynomials.

lower bounds to help the reader gain familiarity. We then present (a slight generalization of) an exponential lower bound for monotone circuits due to Jerrum and Snir [JS82]. Moving on to more restricted (but still nontrivial and interesting) models, we first present an exponential lower bound for depth three circuits over finite fields due to Grigoriev and Karpinski [GK98] and multilinear formulas. We conclude with some recent progress on lower bounds for homogeneous depth four circuits.

Remark. *Throughout the article, we shall use* Det_n *and* Perm_n *to refer to the determinant and permanent respectively of a symbolic* $n \times n$ *matrix* $((x_{ij}))_{1 \leq i,j \leq n}$.

2. Existential lower bounds

Before we embark on our quest to prove lower bounds for interesting families of polynomials, it is natural to ask as to what ~~are~~ bounds one can hope to achieve. For a multivariate polynomial $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$, denote by $S(f)$ the size of the smallest arithmetic circuit computing f.

Theorem 1. **[Folklore.]** *For "most" polynomials* $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ *of degree* d *on* n *variables we have*

$$S(f) \quad \geq \quad \Omega\left(\sqrt{\binom{n+d}{d}}\right).$$

Sketch of Proof. We prove this here only in the situation where the underlying field $\mathbb{F}$ is a finite field and refer the reader to another survey ([CKW11], Chapter 4) for a proof in the general case. So let $\mathbb{F} = \mathbb{F}_q$ be a finite field. Any line of a straight line program computing f can be expressed as taking the product of two $\mathbb{F}_q$-linear combinations of previously computed values. Hence the total number of straight-line programs of length s is at most $q^{O(s^2)}$. On the other hand there are $q^{\binom{n+d}{d}}$ polynomials of degree d on n variables. Hence most n-variate polynomials of degree d require straight-line programs of length at least (equivalently arithmetic circuits of size at least) $s = \Omega\left(\sqrt{\binom{n+d}{d}}\right)$. □

Hrubes and Yehudayoff [HY11] showed that in fact most n-variate polynomials of degree d *with zero-one coefficients* have complexity at least $\Omega\left(\sqrt{\binom{n+d}{d}}\right)$. Now it turns out that this is in fact a lower bound on the number of multiplications in any circuit computing a random polynomial. Lovett [Lov11] complements this nicely by giving a matching upper bound. Specifically, it was shown in [Lov11] that for any polynomial f of degree d on n variables there exists a circuit computing f having at most $\left(\sqrt{\binom{n+d}{d}}\right) \cdot (nd)^{O(1)}$ multiplications.

3. Weak lower bounds for general circuits and formulas

Despite several attempts by various researchers to prove lower bounds for arithmetic circuits or formulas, we only have very mild lower bounds for general circuits or formulas ~~insofar~~thus far. In this section, we shall look at the two modest lower bounds for general circuits and formulas.

3.1. Lower bounds for general circuits

The only super-linear lower bound we currently know for general arithmetic circuits is the following result of Baur and Strassen [BS83].

Theorem 2 ([BS83]). *Any fan-in 2 circuit that computes the polynomial* $f = x_1^{d+1} + \cdots + x_n^{d+1}$ *has size* $\Omega(n \log d)$.

3.1.1. An exploitable weakness. Each gate of the circuit Φ computes a local operation on the two children. To formalize this, define a new variable y_g for every gate $g \in \Phi$. Further, for every gate g define a quadratic equation Q_g as

$$Q_g = \begin{cases} y_g - (y_{g_1} + y_{g_2}) & \text{if } g = g_1 + g_2 \\ y_g - (y_{g_1} \cdot y_{g_2}) & \text{if } g = g_1 \cdot g_2 \end{cases}$$

Further if y_o corresponds to the output gate, then the system of equations

$$\{Q_g = 0 \ : \ g \in \Phi\} \quad \cup \quad \{y_o = 1\}$$

completely characterize the computations of Φ that results in an output of 1. The same can also be extended for *multi-output* circuits that compute several polynomials simultaneously. In such cases, the set of equations

$$\{Q_g = 0 \ : \ g \in \Phi\} \quad \cup \quad \{y_{o_i} = 1 \ : \ i = 1, \ldots, n\}$$

completely characterize computations that result in an output of all ones. The following classical theorem allows us to bound the number of common roots to a system of polynomial equations.

Theorem 3 (Bézout's theorem). *Let* $g_1, \ldots, g_r \in \mathbb{F}[X]$ *such that* $\deg(g_i) = d_i$ *such that the number of common roots of* $g_1 = \cdots = g_r = 0$ *is finite. Then, the number of common roots (counted with multiplicities) is bounded by* $\prod d_i$.

Thus in particular, if we have a circuit Φ of size s that *simultaneously* computes $\{x_1^d, \ldots, x_n^d\}$, then we have d^n inputs that evaluate to all ones (where each x_i must be a d-th root of unity). Hence, Bézout's theorem implies that

$$2^s \quad \geq \quad d^n \qquad \Longrightarrow \qquad s \quad = \quad \Omega(d \log n).$$

Observe that $\{x_1^d, \ldots, x_n^d\}$ are all first-order derivatives of $f = x_1^{d+1} + \cdots + x_n^{d+1}$ (with suitable scaling). A natural question here is the following — if f can be computed an arithmetic circuit of size s, what is the size required to compute all first-order partial derivatives of f simultaneously? The naïve approach of computing each derivative separately results in a circuit of size $O(s \cdot n)$. Baur and Strassen [BS83] show that we can save a factor of n.

Lemma 4 ([BS83]). *Let Φ be an arithmetic circuit of size s and fan-in 2 that computes a polynomial $f \in \mathbb{F}[X]$. Then, there is a multi-output circuit of size $O(s)$ computing all first order derivatives of f.*

Note that this immediately implies that any circuit computing $f = x_1^{d+1} + \cdots + x_n^{d+1}$ requires size $\Omega(d \log n)$ as claimed by Theorem 2.

3.1.2. Computing all first order derivatives simultaneously. Since we are working with fan-in 2 circuits, the number of edges is at most twice the size. Hence let s denote the number of edges in the circuit Φ, and we shall prove by induction that all first order derivatives of Φ can be computed by a circuit of size at most $5s$. Pick a non-leaf node v in the circuit Φ closest to the leaves with both its children being variables, and say x_1 and x_2 are the variables feeding into v. In other words, $v = x_1 \odot x_2$ where $\odot$ is either $+$ or $\times$.

Let Φ' be the circuit obtained by deleting the two edges feeding into v, and replacing v by a new variable. Hence, Φ' computes a polynomial $f' \in \mathbb{F}[X \cup \{v\}]$ and has at most $(s-1)$ edges. By induction on the size, we can assume that there is a circuit $\mathbb{D}(\Phi')$ consisting of at most $5(s-1)$ edges that computes all the first order derivatives of f'.

Observe that since $f' \mid_{(v = x_1 \odot x_2)} = f(\mathbf{x})$, we have that

$$\frac{\partial f}{\partial x_i} = \left(\frac{\partial f'}{\partial x_i}\right)_{v = x_1 \odot x_2} + \left(\frac{\partial f'}{\partial v}\right)_{v = x_1 \odot x_2} \left(\frac{\partial (x_1 \odot x_2)}{\partial x_i}\right).$$

Hence, if $v = x_1 + x_2$ then

$$\begin{aligned}
\frac{\partial f}{\partial x_1} &= \left(\frac{\partial f'}{\partial x_1}\right)_{v=x_1+x_2} + \left(\frac{\partial f'}{\partial v}\right)_{v=x_1+x_2} \\
\frac{\partial f}{\partial x_2} &= \left(\frac{\partial f'}{\partial x_2}\right)_{v=x_1+x_2} + \left(\frac{\partial f'}{\partial v}\right)_{v=x_1+x_2} \\
\frac{\partial f}{\partial x_i} &= \left(\frac{\partial f'}{\partial x_i}\right)_{v=x_1+x_2} \qquad \text{for } i > 2.
\end{aligned}$$

If $v = x_1 \cdot x_2$, then

$$\begin{aligned}
\frac{\partial f}{\partial x_1} &= \left(\frac{\partial f'}{\partial x_1}\right)_{v=x_1\cdot x_2} + \left(\frac{\partial f'}{\partial v}\right)_{v=x_1\cdot x_2} \cdot x_2 \\
\frac{\partial f}{\partial x_2} &= \left(\frac{\partial f'}{\partial x_2}\right)_{v=x_1\cdot x_2} + \left(\frac{\partial f'}{\partial v}\right)_{v=x_1\cdot x_2} \cdot x_1 \\
\frac{\partial f}{\partial x_i} &= \left(\frac{\partial f'}{\partial x_i}\right)_{v=x_1\cdot x_2} \qquad \text{for } i > 2.
\end{aligned}$$

Hence, by adding at most 5 additional edges to $\mathbb{D}(\Phi')$, we can construct $\mathbb{D}(\Phi)$ and hence size of $\mathbb{D}(\Phi)$ is at most $5s$. □(Lemma 4)

3.2. Lower bounds for formulas

This section would be devoted to the proof of Kalorkoti's lower bound [Kal85] for formulas computing Det_n, Perm_n.

Theorem 5 **([Kal85]).** *Any arithmetic formula computing* Perm_n *(or* Det_n*) requires* $\Omega(n^3)$ *size.*

The exploitable weakness in this setting is again to use the fact that the polynomials computed at intermediate gates share many polynomial dependencies.

Definition 6 (Algebraic independence). *A set of polynomials* $\{f_1, \ldots, f_m\}$ *is said to be* algebraically independent *if there is no non-trivial polynomial* $H(z_1, \ldots, z_m)$ *such that* $H(f_1, \ldots, f_m) = 0$.
The size of the largest algebraically independent subset of $\mathbf{f} = \{f_1, \ldots, f_m\}$ *is called the* transcendence degree *(denoted by* $\mathrm{trdeg}(f)$*).*

The proof of Kalorkoti's theorem proceeds by defining a *complexity measure* using the above notion of algebraic independence.

The Measure: For any subset of variables $Y \subseteq X$, we can write a polynomial $f \in \mathbb{F}[X]$ of the form $f = \sum_{i=1}^{s} f_i \cdot m_i$ where m_i's are distinct monomials in the variables in Y, and ~~$f_i \in F[Y]$~~ $f_i \in F[X \setminus Y]$. We shall denote by $\mathrm{td}_Y(f)$ the transcendence degree of $\{f_1, \ldots, f_s\}$
Fix a partition of variables $X = X_1 \sqcup \cdots \sqcup X_r$. For any polynomial $f \in \mathbb{F}[X]$, define the map $\Gamma^{[\mathrm{Kal}]} : \mathbb{F}[X] \to \mathbb{Z}_{\geq 0}$ as

$$\Gamma^{[\mathrm{Kal}]}(f) \quad = \quad \sum_{i=1}^{r} \mathrm{td}_{X_i}(f)$$

The lower bound proceeds in two natural steps:

1. Show that $\Gamma^{[\mathrm{Kal}]}(f)$ is *small* whenever f is computable by a *small* formula.
2. Show that $\Gamma^{[\mathrm{Kal}]}(\mathsf{Det}_n)$ is *large*.

3.2.1. Upper bounding $\Gamma^{[\mathrm{Kal}]}$ for a formula.

Lemma 7. *Let f be computed by a fan-in two formula Φ of size s. Then for any partition of variables $X = X_1 \sqcup \cdots \sqcup X_r$, we have $\Gamma^{[\mathrm{Kal}]}(f) = O(s)$.*

Proof. For any node $v \in \Phi$, let $\textsc{Leaf}(v)$ denote the leaves of the subtree rooted at v and let $\textsc{Leaf}_{X_i}(v)$ denote the leaves of the subtree rooted at v that are in the part X_i. Since the underlying graph of Φ is a tree, it follows that the size of Φ is bounded by a twice the number of leaves. For each part X_i, we shall show that $\mathrm{td}_{X_i}(f) = O(|\textsc{Leaf}_{X_i}(\Phi)|)$, which would prove the required bound.

Fix an arbitrary part $Y = X_i$. Define the following three sets of nodes

$$\begin{aligned} V_0 &= \{v \in \Phi \ : \ |\textsc{Leaf}_Y(v)| = 0 \quad \text{and} \quad |\textsc{Leaf}_Y(\textsc{Parent}(v))| \geq 2\} \\ V_1 &= \{v \in \Phi \ : \ |\textsc{Leaf}_Y(v)| = 1 \quad \text{and} \quad |\textsc{Leaf}_Y(\textsc{Parent}(v))| \geq 2\} \\ V_2 &= \{v \in \Phi \ : \ |\textsc{Leaf}_Y(v)| \geq 2\} \end{aligned}$$

Each node $v \in V_0$ computes a polynomial in ~~$f_v \in \mathbb{F}[\overline{Y}]$~~ $f_v \in \mathbb{F}[X \setminus Y]$, and we shall replace the subtree at v by a node computing the polynomial f_v. Similarly, any node $v \in V_1$ computes a polynomial of the form $f_v^{(0)} + f_v^{(1)} y_v$ for some $y_v \in Y$ and ~~$f_v^{(0)}, f_v^{(1)} \in \mathbb{F}[\overline{Y}]$~~ $f_v^{(0)}, f_v^{(1)} \in \mathbb{F}[X \setminus Y]$. We shall again replace the subtree rooted at v by a node computing $f_v^{(0)} + f_v^{(1)} y_v$.
Hence, the formula Φ now reduces to a smaller formula Φ_Y with leaves being the nodes in V_0 and V_1 (and nodes in V_2 are unaffected). We would like to show that the size of the reduced formula, which is at most twice the number of its leaves, is $O(|\textsc{Leaf}_Y(\Phi)|)$.

Observation 8. $|V_1| \leq |\textsc{Leaf}_Y(\Phi)|$.

Proof. Each node in V_1 has a distinct leaf labelled with a variable in Y. Hence, $|V_1|$ is bounded by the number of leaves labelled with a variable in Y.

□ (Obs)

This shows that the V_1 leaves are not too many. Unfortunately, we cannot immediately bound the number of V_0 leaves, since we could have a long chain of V_2 nodes each with one sibling being a V_0 leaf. The following observation would show how we can eliminate such long chains.

Observation 9. *Let u be an arbitrary node, and v be another node in the subtree rooted at u with $\textsc{Leaf}_Y(u) = \textsc{Leaf}_Y(v)$. Then the polynomial g_u computed at u and the polynomial g_v computed at v are related as $g_u = f_1 g_v + f_2$ for some* ~~$f_1, f_2 \in \mathbb{F}[\overline{Y}]$~~ $f_1, f_2 \in \mathbb{F}[X \setminus Y]$.

Proof. If $\textsc{Leaf}_Y(u) = \textsc{Leaf}_Y(v)$, then every node on the path from u to v must have a V_0 leaf as the other child. The observation follows as all these nodes are + or × gates. □ (Obs)

Using the above observation, we shall remove the need for V_0 nodes completely by augmenting each node u (computing a polynomial g_u) in Φ_Y with polynomials ~~$f_u^{(0)}, f_u^{(1)} \in \mathbb{F}[\overline{Y}]$~~ $f_u^{(0)}, f_u^{(1)} \in \mathbb{F}[X \setminus Y]$ to enable them to compute $f_u^{(1)} g_u + f_u^{(0)}$. Let this augmented formula be called $\hat{\Phi}_Y$. Using Observation 9, we can now contract any two nodes u and v with $\textsc{Leaf}_Y(u) = \textsc{Leaf}_Y(v)$, and eliminate all V_0 nodes completely. Since all V_2 nodes are internal nodes, the only leaves of the augmented formula $\hat{\Phi}_Y$ are in V_1. Hence, the size of the augmented formula $\hat{\Phi}_Y$ is bounded by $2|V_1|$, which is bounded by $2|\textsc{Leaf}_Y(\Phi)|$ by Observation 8.

Suppose Φ computes a polynomial f, which can be written as $f = \sum_{i=1}^{t} f_i \cdot m_i$ with ~~$f_i \in \mathbb{F}[\overline{Y}]$~~ $f_i \in \mathbb{F}[X \setminus Y]$ and m_i's being distinct monomials in Y. Since

$\hat{\Phi}_Y$ also computes f, each f_i is a polynomial combination of the polynomials $S_Y = \left\{ f_u^{(0)}, f_u^{(1)} \;:\; u \in \hat{\Phi}_Y \right\}$. Since $\hat{\Phi}_Y$ consists of at most $2\left|\textsc{Leaf}_Y(\Phi)\right|$ augmented nodes, we have that $\mathrm{td}_Y(f) \leq |S_Y| \leq 4\left|\textsc{Leaf}_Y(\Phi)\right|$. Therefore,

$$\mathrm{td}_Y(f) \quad = \quad \mathrm{trdeg}\left\{f_i \;:\; i \in [t]\right\} \quad \leq \quad 4\left|\textsc{Leaf}_Y(\Phi)\right|$$

Hence,

$$\Gamma^{[\mathrm{Kal}]}(\Phi) = \sum_{i=1}^{r} \mathrm{td}_{X_i}(f_i) \leq 4\left(\sum_{i=1}^{r} \left|\textsc{Leaf}_{X_i}\right|\right) = O(s)$$

□

3.2.2. Lower bounding $\Gamma^{[\mathrm{Kal}]}(\mathsf{Det}_n)$.

Lemma 10. *Let $X = X_1 \sqcup \cdots \sqcup X_n$ be the partition as defined by $X_t = \{x_{ij} \;:\; i - j \equiv t \bmod n\}$. Then, $\Gamma^{[\mathrm{Kal}]}(\mathsf{Det}_n) = \Omega(n^3)$.*

Proof. By symmetry, it is easy to see that $\mathrm{td}_{X_i}(\mathsf{Det}_n)$ is the same for all i. Hence, it suffices to show that $\mathrm{td}_Y(\mathsf{Det}_n) = \Omega(n^2)$ for $Y = X_n = \{x_{11}, \ldots, x_{nn}\}$. To see this, observe that the determinant consists of the monomials $\left(\frac{x_{11}\cdots x_{nn}}{x_{ii}x_{jj}}\right) \cdot x_{ij}x_{ji}$ for every $i \neq j$. Hence, $\mathrm{td}_Y(\mathsf{Det}_n) \geq \mathrm{trdeg}\left\{x_{ij}x_{ji} \;:\; i \neq j\right\} = \Omega(n^2)$. Therefore, $\Gamma^{[\mathrm{Kal}]}(\mathsf{Det}_n) = \Omega(n^3)$. □

The proof of Theorem 5 follows from Lemma 7 and Lemma 10.

4. "Natural" proof strategies

The lower bounds presented in Section 3 proceeded by first identifying a *weakness* of the model, and exploiting it in an explicit manner. More concretely, Section 3.2 presents a promising strategy that could be adopted to prove lower bounds for various models of arithmetic circuits. The crux of the lower bound was the construction of a good map Γ that assigned a number to every polynomial. The map $\Gamma^{[\mathrm{Kal}]}$ was useful to show a lower bound in the sense that any f computable by a *small* formula had *small* $\Gamma^{[\mathrm{Kal}]}(f)$. In fact, all subsequent lower bounds in arithmetic circuit complexity have more or less followed a similar template of a "natural proof". More concretely, all the subsequent lower bounds we shall see would essentially follow the outlined plan.

> **Step 1 (normal forms)** For every circuit in the circuit class $\mathcal{C}$ of interest, express the polynomial computed as a *small sum of simple building blocks*.

For example, every $\Sigma\Pi\Sigma$ circuit is a *small* sum of *products of linear polynomials* which are the building blocks here. In this case, the circuit model naturally admits such a representation but we shall see other examples with very different representations as sum of building blocks.

> **Step 2 (complexity measure)** Construct a map $\Gamma : \mathbb{F}[x_1, \ldots, x_n] \to \mathbb{Z}_{\geq 0}$ ~~with~~ that is *sub-additive* i.e. $\Gamma(f_1 + f_2) \leq \Gamma(f_1) + \Gamma(f_2)$

In most cases, $\Gamma(f)$ is the rank of a large matrix whose entries are linear functions in the coefficients of f. In such cases, we immediately get that Γ is sub-additive.
The strength of the choice of Γ is determined by the next step.

> **Step 3 (potential usefulness)** Show that if B is a *simple building block*, then $\Gamma(B)$ is *small.* Further, check if $\Gamma(f)$ for a *random polynomial* f is large (potentially).

This would suggest that if any f with large $\Gamma(f)$ is to be written as a sum of $B_1 + \cdots + B_s$, then sub-additivity and the fact that $\Gamma(B_i)$ is small for each i and $\Gamma(f)$ is large immediately imply that s must be large. This implies that the complexity measure Γ does indeed have a potential to prove a lower bound for the class. The next step is just to replace the *random polynomial* by an explicit polynomial.

> **Step 4 (explicit lower bound)** Find an explicit polynomial f for which $\Gamma(f)$ is large.

These are usually the steps taken in almost all the known arithmetic circuit lower bound proofs. The main ingenuity lies in constructing a useful complexity measure, which is really to design Γ so that it is small on the *building blocks.*

Of course, there could potentially be lower bound proofs that do not follow the road-map outlined. For instance, it could be possible that Γ is not small for a random polynomial, but specifically tailored in a way to make Γ large for the Perm_n. Or perhaps Γ need not even be sub-additive and maybe there is a very different way to argue that all polynomial in the circuit class have small Γ. However, this has been the road-map for almost all lower bounds so far (barring very few exceptions). As a warmup, we first present some very simple applications of the above plan to prove lower bounds for some very simple subclasses of arithmetic circuits in the next section. We then move on to more sophisticated proofs of lower bounds for less restricted subclasses of circuits.

5. Some simple lower bounds

Let us start with the simplest complete[3] class of arithmetic circuits – depth-2 circuits or $\Sigma\Pi$ circuits.

5.1. Lower bounds for $\Sigma\Pi$ circuits

Any $\Sigma\Pi$ circuit of size s computes a polynomial $f = m_1 + \cdots + m_s$ where each m_i is a monomial multiplied by a field constant. Therefore, any polynomial computed by a *small* $\Sigma\Pi$ circuit must have a *small* number of monomials. Hence, it is obvious that any polynomial that has many monomials require large $\Sigma\Pi$ circuits.

[3] in the sense that any polynomial can be computed in this model albeit of large size

This can be readily rephrased in the language of the outline described last section by defining $\Gamma(f)$ to simply be the number of monomials present in f. Hence, $\Gamma(f) \leq s$ for any f computed by a $\Sigma\Pi$ circuit of size s. Of course, even a polynomial like $f = (x_1+x_2+\cdots+x_n)^n$ have ~~$\Gamma(f) = n^n$~~ $\Gamma(f) = n^{\Omega(n)}$ giving the lower bound.

5.2. Lower bounds for $\Sigma\wedge\Sigma$ circuits

A $\Sigma\wedge\Sigma$ circuit of size s computes a polynomial of the form $f = \ell_1^{d_1} + \cdots + \ell_s^{d_s}$ where each ℓ_i is a linear polynomial over the n variables.[4]
Clearly as even a single ℓ^d could have exponentially many monomials, the Γ defined above cannot work in this setting. Nevertheless, we shall try to design a similar map to ensure that $\Gamma(f)$ is *small* whenever f is computable by a *small* $\Sigma\wedge\Sigma$ circuit.

In this setting, the *building blocks* are terms of the form ℓ^d. The goal would be to construct a *sub-additive* measure Γ such that $\Gamma(\ell^d)$ is *small*. Here is the key observation to guide us towards a good choice of Γ.

Observation 11. *Any k-th order partial derivative of ℓ^d is a constant multiple of ℓ^{d-k}.*

Hence, if $\partial^{=k}(f)$ denotes the set of k-th order partial derivatives of f, then the space spanned by $\partial^{=k}(\ell^d)$ has dimension 1. This naturally leads us to define Γ exploiting this weakness.

$$\Gamma_k(f) \quad \stackrel{\text{def}}{=} \quad \dim\left(\partial^{=k}(f)\right)$$

It is straightforward to check that Γ_k is indeed sub-additive and hence $\Gamma_k(f) \leq s$ whenever f is computable by a $\Sigma\wedge\Sigma$ circuit of size s. For a random polynomial f, we should be expecting $\Gamma_k(f)$ to be $\binom{n+k}{k}$ as there is unlikely to be any linear dependencies among the partial derivatives. Hence, all that needs to be done is to find an explicit polynomial with large Γ_k.
If we consider Det_n or Perm_n, then any partial derivative of order k is just an $(n-k)\times(n-k)$ minor. Also, these minors consist of disjoint sets of monomials and hence are linearly independent. Hence, $\Gamma_k(\mathsf{Det}_n) = \binom{n}{k}^2$. Choosing $k = n/2$, we immediately get that any $\Sigma\wedge\Sigma$ circuit computing Det_n or Perm_n must be of size $2^{\Omega(n)}$.

5.3. Low-rank $\Sigma\Pi\Sigma$

A slight generalization of $\Sigma\wedge\Sigma$ circuits is a *rank-r $\Sigma\Pi\Sigma$ circuit* that computes a polynomial of the form

$$f \quad = \quad T_1 \,+\, \ldots \,+\, T_s$$

where each $T_i = \ell_{i1}\ldots\ell_{id}$ is a product of linear polynomials such that $\dim\{\ell_{i1},\ldots,\ell_{id}\} \leq r$.

[4] such circuits are also called *diagonal depth-3 circuits* in the literature

Thus, $\Sigma\wedge\Sigma$ is a rank-1 $\Sigma\Pi\Sigma$ circuit, and a similar partial-derivative technique for lower bounds works here as well.
In the setting where r is much smaller than the number of variables n, each T_i is essentially an r-variate polynomial masquerading as an n-variate ~~polynomials~~ polynomial using an affine transformation. In particular, the set of n first order derivatives of T have rank at most r. This yields the following observation.

Observation 12. *Let* $T = \ell_1 \ldots \ell_d$ *with* $\dim\{\ell_1, \ldots, \ell_d\} \leq r$. *Then for any* k, *we have*

$$\Gamma_k(T) \quad \overset{def}{=} \quad \dim\left(\partial^{=k}(T)\right) \quad \leq \quad \binom{r+k}{k}$$

Thus once again by sub-additivity, for any polynomial f computable by a rank-r $\Sigma\Pi\Sigma$ circuit of size s, we have $\Gamma_k(f) \leq s \cdot \binom{r+k}{r}$. Note that a random polynomial is expected to have $\Gamma_k(f)$ close to $\binom{n+k}{k}$, which could be much larger for $r \ll n$. We already saw that $\Gamma_k(\mathsf{Det}_n) = \binom{n}{k}^2$. This immediately gives the following lower bound, the proof of which we leave as an exercise to the interested reader.

Theorem 13. *Let* $r \leq n^{2-\delta}$ *for some constant* $\delta > 0$. *For* $k = \varepsilon n^{\delta}$, *where* $\varepsilon > 0$ *is sufficiently small, we have*

$$\frac{\binom{n}{k}^2}{\binom{r+k}{k}} \quad = \quad \exp\left(\Omega(n^{\delta})\right)$$

Hence, any rank-r $\Sigma\Pi\Sigma$ *circuit computing* Det_n *or* Perm_n *must have size* $\exp\left(\Omega(n^{\delta})\right)$. □

This technique of using the rank of partial derivatives was introduced by Nisan and Wigderson [NW97] to prove lower bounds for *homogeneous depth-3 circuits* (which also follows as a corollary of Theorem 13). The survey of Chen, Kayal and Wigderson [CKW11] give a comprehensive exposition of the power of the *partial derivative method*.

With these simple examples, we can move on to other lower bounds for various other more interesting models.

6. Lower bounds for monotone circuits

This section would present a slight generalization of a lower bound by Jerrum and Snir [JS82]. To motivate our presentation here, let us first assume that the underlying field is $\mathbb{R}$, the field of real numbers. A monotone circuit over $\mathbb{R}$ is a circuit having $+, \times$ gates in which all the field constants are *non-negative* real numbers. Such a circuit can compute any polynomial f over $\mathbb{R}$ all of whose coefficients are nonnegative real numbers, such as for example the permanent. It is then natural to ask whether there are small monotone circuits over $\mathbb{R}$ computing the permanent. Jerrum and Snir [JS82] obtained an exponential lower bound on the size of monotone circuits over $\mathbb{R}$ computing the permanent. Note that this definition of monotone circuits is valid only over $\mathbb{R}$ (actually more generally over ordered fields but not over say finite fields) and such circuits can only compute polynomials with non-negative coefficients. Here we will present Jerrum and Snir's argument in a slightly more generalized form such that the circuit model makes sense over any field $\mathbb{F}$ and is complete, i.e. can compute any polynomial over $\mathbb{F}$. Let us first explain the motivation behind the generalized circuit model that we present here. Observe that in any monotone circuit over $\mathbb{R}$, there is no cancellation as there are no negative coefficients. Formally, for a node v in our circuits let us denote by f_v the polynomial computed at that node~~for~~. For a polynomial f let us denote by $\mathrm{Mon}(f)$ the set of monomials having a nonzero coefficient in the polynomial f.

1. If $w = u + v$ then
$$\mathrm{Mon}(f_w) = \mathrm{Mon}(f_u) \cup \mathrm{Mon}(f_v).$$
2. If $w = u \times v$ then

$$\mathrm{Mon}(f_w) = \mathrm{Mon}(f_u)\cdot\mathrm{Mon}(f_v) \overset{\text{def}}{=} \{m_1 \cdot m_2 \;:\; m_1 \in \mathrm{Mon}(f_u), m_2 \in \mathrm{Mon}(f_v)\}\,.$$

This means that for any node v in a monote circuit over $\mathbb{R}$ one can determine $\mathrm{Mon}(f_v)$ in a very syntactic manner starting from the leaf nodes. Let us make precise this syntactic ~~coomputation~~ computation that we have in mind.

Definition 14 (Formal Monomials.). *Let Φ be an arithmetic circuit. The* formal monomials *at any node $v \in \Phi$, which shall be denoted by* $\mathrm{FM}(v)$, *shall be inductively defined as follows:*

> *If v is a leaf labelled by a variable x_i, then $\mathrm{FM}(v) = \{x_i\}$. If it is labelled by a constant, then $\mathrm{FM}(v) = \{1\}$.*
>
> *If $v = v_1 + v_2$, then $\mathrm{FM}(v) = \mathrm{FM}(v_1) \cup \mathrm{FM}(v_2)$.*
> *If $v = v_1 \times v_2$, then*
>
> $$\begin{aligned} \mathrm{FM}(v) &= \mathrm{FM}(v_1) \cdot \mathrm{FM}(v_2) \\ &\overset{def}{=} \{m_1 \cdot m_2 \;:\; m_1 \in \mathrm{FM}(v_1), m_2 \in \mathrm{FM}(v_2)\} \end{aligned}$$

Note that for any node v in any circuit we have $\mathrm{Mon}(f_v) \subseteq \mathrm{FM}(v)$ but in a monotone circuit over $\mathbb{R}$ this containment is in fact an equality at every node. This motivates our definition of a slightly more general notion of a monotone circuit as follows.

Definition 15 (Monotone circuits). *A circuit C is said to be* syntactically monotone *(simply monotone for short) if* $\mathrm{Mon}(f_v) = \mathrm{FM}(v)$ *for every node v in C.*

The main theorem of this section is the following:

Theorem 16 ([JS82]). *Over any field $\mathbb{F}$, any syntactically monotone circuit C computing Det_n or Perm_n must have size at least $2^{\Omega(n)}$.*

The proof of this theorem is relatively short assuming the following structural result (which is present in standard depth-reduction proofs [VSBR83, AJMV98]).

Lemma 17. *Let f be a degree d polynomial computed by a monotone circuit of size s. Then, f can be written of the form $f = \sum_{i=1}^{s} f_i \cdot g_i$ where the f_i's and g_i's satisfy the following properties.*

1. *For each $i \in [s]$, we have $\frac{d}{3} < \deg g_i \leq \frac{2d}{3}$.*
2. *For each i, we have $\mathrm{FM}(f_i) \cdot \mathrm{FM}(g_i) \subseteq \mathrm{FM}(f)$*

We shall defer this lemma to the end of the section and first see how this would imply Theorem 16. The complexity measure $\Gamma(f)$ in this case is just the number of monomials in f, but it is the above *normal form* that is crucial in the lower bound.

Proof of Theorem 16. Suppose Φ is a circuit of size s that computes Det_n. Then by Lemma 17,

$$\mathsf{Det}_n \quad = \quad \sum_{i=1}^{s} f_i \cdot g_i$$

with $\mathrm{FM}(f_i) \cdot \mathrm{FM}(g_i) \subseteq \mathrm{FM}(\mathsf{Det}_n)$. The building blocks are terms of the form $T = f \cdot g$, where $\mathrm{FM}(f) \cdot \mathrm{FM}(g) \subseteq \mathrm{FM}(\mathsf{Det}_n)$.

Since all the monomials in Det_n are products of variables from distinct columns and rows, the rows (and columns) containing the variables f depends on is disjoint from the rows (and columns) containing variables that g depends on. Hence, there exists sets of indices $A, B \subseteq [n]$ such that f depends only on $\{x_{jk} \ : \ j \in A, k \in B\}$ and g depends only on $\left\{x_{jk} \ : \ j \in \overline{A}, k \in \overline{B}\right\}$. Further, since Det_n is a homogeneous polynomial of degree n, we also have that both f and g must be homogeneous as well. Also as all monomials of g using distinct row and column indices from $\overline{A}$ and $\overline{B}$ respectively, we see that $\deg g = |\overline{A}| = |\overline{B}|$ and $\deg f = |A| = |B|$. Using Lemma 17, let $|A| = \alpha n$ for some $\frac{1}{3} \leq \alpha \leq \frac{2}{3}$. This implies that $\Gamma(f) \leq (\alpha n)!$, and $\Gamma(g) \leq ((1-\alpha)n)!$ and hence

$$\Gamma(f \cdot g) \quad \leq \quad (\alpha n)!((1-\alpha)n)! \quad \leq \quad \frac{n!}{\binom{n}{n/3}}$$

as $\frac{1}{3} \leq \alpha \leq \frac{2}{3}$. Also, Γ is clearly sub-additive and we have

$$\Gamma(f_1 g_1 + \cdots + f_s g_s) \quad \leq \quad s \cdot \underline{(\alpha n)! \cdot ((1-\alpha)n)!}\frac{n!}{\binom{n}{n/3}}$$

Since $\Gamma(\mathsf{Det}_n) = n!$, this forces ~~$s \geq \binom{n}{\alpha n} = 2^{\Omega(n)}$ as $\frac{1}{3} \leq \alpha \leq \frac{2}{3}$~~ $s \geq \binom{n}{n/3} = 2^{\Omega(n)}$.

□

We only need to prove Lemma 17 now.

6.1. Proof of Lemma 17

Without loss of generality, assume that the circuit Φ is homogeneous[5], and consists of alternating layers of $+$ and $\times$ gates. Also, assume that all $\times$ gates have fan-in two, and orient the two children such that the formal degree of the left child is at least as large as the formal degree of the right child. Such circuits are also called *left-heavy* circuits.

Definition 18 (Proof tree). *A* proof tree *of an arithmetic circuit Φ is a sub-circuit Φ' such that*

- *The root of Φ is in Φ'*
- *If a multiplication gate with $v = v_1 \times v_2 \in \Phi'$, then v_1 and v_2 are in Φ' as well.*
- *If an addition gate $v = v_1 + \cdots + v_s \in \Phi'$, then exactly one v_i is in Φ'.*

Such a sub-circuit Φ', represented as a tree (duplicating nodes if required), shall be called a proof tree *of Φ.*

Let ProofTrees(Φ) denote the set of all proof trees of Φ. It is easy to see that any proof tree of Φ computes a monomial over the variables. Further, if Φ was a monotone circuit computing a polynomial f, then every proof tree computes a monomial in f. Therefore,

$$f = \sum_{\Phi' \in \textsc{ProofTrees}(\Phi)} [\Phi']$$

where $[\Phi']$ denotes the monomial computed by Φ'. Of course, the number of proof trees is exponential and the above expression is huge. However, we could use a divide-and-conquer approach to the above equation using the following lemma.

Lemma 19. *Let Φ' be a left-heavy formula of formal degree d. Then there is a node v on the left-most path of Φ' such that $\frac{d}{3} \leq \deg(v) < \frac{2d}{3}$.*

Proof. Pick the lowest node on the left-most path that has degree at least $\frac{2d}{3}$. Then, its left child must be a node of degree less than $\frac{2d}{3}$, and also at least $\frac{d}{3}$ (because the formula is left-heavy). □

For any proof tree Φ' and a node v on its left-most path, define $[\Phi' : v]$ to be the output polynomial of the proof tree obtained by replacing the node v on the left-most path by 1. If v does not occur on the left-most path of Φ', define $[\Phi' : v]$ to be 0. We will denote the polynomial computed at a node v

[5] It is a forklore result that any circuit can be *homogenized* with just a polynomial blow-up in size. Further, this process also preserves monotonicity of the circuit. A proof of this may be seen in [SY10].

by f_v. Then, the above equation can now be re-written as:

$$\begin{aligned} f &= \sum_{\Phi' \in \textsc{ProofTrees}(\Phi)} [\Phi'] \\ &= \sum_{\substack{v \in \Phi \\ \frac{d}{3} \leq \deg v < \frac{2d}{3}}} f_v \cdot \left(\sum_{\Phi' \in \textsc{ProofTrees}(\Phi)} [\Phi' : v] \right) \\ &= \sum_{\substack{v \in \Phi \\ \frac{d}{3} \leq \deg v < \frac{2d}{3}}} f_v \cdot g_v \qquad \text{where } g_v = \sum_{\Phi' \in \textsc{ProofTrees}(\Phi)} [\Phi' : v] \end{aligned}$$

Since $\frac{d}{3} \leq \deg v < \frac{2d}{3}$, we also have that $\frac{d}{3} < \deg g_v \leq \frac{2d}{3}$ and the last equation is what was required by Lemma 17. □

7. Lower bounds for depth-3 circuits over finite fields

This section shall present the lower bound of Grigoriev and Karpinski [GK98] for Det_n. A follow-up paper of Grigoriev and Razborov [GR00] extended the result over function fields, also including a weaker lower bound for the permanent, but we shall present a slightly different proof that works for the permanent as well.

Theorem 20. [GK98] *Any depth-3 circuit computing* Det_n *(or* Perm_n*) over a finite field* $\mathbb{F}_q$ *(*$q \neq 2$*) requires size* $2^{\Omega(n)}$*.*

Main idea: Let $q = |\mathbb{F}|$. Suppose $C = T_1 + \cdots + T_s$, where each T_i is a product of linear polynomials. Define $\operatorname{rank}(T_i)$ as in Section 5.3 to be the dimension of the set of linear polynomials that T_i is a product of.

In Section 5.3, we saw that the dimension of partial derivatives would handle *low rank* T_i's. As for the high rank T_i's, since T_i is a product of at least r linearly independent linear polynomials, a random evaluation keeps T_i non-zero with probability at most $\left(1 - \frac{1}{q}\right)^r$. Since q is a constant, we have that a random evaluation of a high rank T_i is almost always zero. Hence, in a sense, C can be "approximated" by just the low-rank components.

Grigoriev and Karpinski [GK98] formalize the above idea as a measure by combining the partial derivative technique seen in Section 5.3 with evaluations to show that Det_n cannot be approximated by a low-rank $\Sigma\Pi\Sigma$ circuit.

7.1. The complexity measure

For any polynomial $f \in \mathbb{F}[x_{11}, \ldots, x_{nn}]$, define the matrix $M_k(f)$ as follows — the columns of $M_k(f)$ are indexed by k-th order partial derivatives of f, and rows by elements of $\mathbb{F}^{n^2}$, with the entry being the evaluation of the partial derivative (column index) at the point (row index).

The rank of $M_k(f)$ could be a possible choice of a complexity measure but Grigoriev and Karpinski make a small modification to handle the high rank T_is. Instead, they look at the matrix $M_k(f)$ and remove a few ~~*erraneous*~~*erroneous* evaluation points and use the rank of the resulting matrix. For any $\mathcal{A} \subseteq \mathbb{F}^{n^2}$, let us define $M_k(f; \mathcal{A})$ to be the matrix obtained from $M_k(f)$ by only taking the rows whose indices are in $\mathcal{A}$. Also, let $\Gamma^{[\mathrm{GK}]}_{k,\mathcal{A}}(f)$ denote $\mathrm{rank}(M_k(f; \mathcal{A}))$.

7.2. Upper-bounding $\Gamma^{[\mathrm{GK}]}_{k,\mathcal{A}}$ for a depth-3 circuit

Our task here is to give an upper bound on the complexity measure for a $\Sigma\Pi\Sigma$-circuit of size s. We first see that the task reduces to upper bounding the measure for a single term via subadditivity. It follows from the linearity of the entries of the matrix.

Observation 21 (Sub-additivity). $\Gamma^{[\mathrm{GK}]}_{k,\mathcal{A}}(f+g) \quad \leq \quad \Gamma^{[\mathrm{GK}]}_{k,\mathcal{A}}(f) + \Gamma^{[\mathrm{GK}]}_{k,\mathcal{A}}(g).$

Now fix a threshold $r_0 = \beta n$ for some constant $\beta > 0$ (to be chosen shortly), and let $k = \gamma n$ for some $\gamma > 0$ (to be chosen shortly). We shall call a term $T = \ell_1 \cdots \ell_d$ to be of *low rank* if $\mathrm{rank}(T) \leq r_0$, and *large rank* otherwise. By the above observation, we need to upper-bound the measure $\Gamma^{[\mathrm{GK}]}_{k,\mathcal{A}}$ for each term T, for a suitable choice of $\mathcal{A}$.

Low rank terms ($\mathrm{rank}(T) \leq r_0$)**:**
Suppose $T = \ell_1 \cdots \ell_d$ with $\{\ell_1, \ldots, \ell_r\}$ being a maximal independent set of linear polynomials (with $r \leq r_0$). Then, T can be expressed as a linear combination of terms from the set $\{\ell_1^{e_1} \ldots \ell_r^{e_r} \ : \ e_i \leq d \quad \forall i \in [r]\}$. And since the matrix $M_k(f)$ depends only on evaluations in $\mathbb{F}^{n^2}$, we can use the relation that $x^q = x$ in $\mathbb{F}$ to express the function $T : \mathbb{F}^{n^2} \to \mathbb{F}$ as a linear combination of $\{\ell_1^{e_1} \ldots \ell_r^{e_r} \ : \ e_i < q \quad \forall i \in [r]\}$. Therefore, for any set $\mathcal{A} \subseteq \mathbb{F}^{n^2}$, we have that

$$\Gamma^{[\mathrm{GK}]}_{k;\mathcal{A}}(T) \quad \leq \quad \mathrm{rank}(M_k(f)) \quad \leq \quad q^r \quad \leq \quad q^{\beta n}.$$

High rank terms ($\mathrm{rank}(T) > r_0$)**:**
Suppose $T = \ell_1 \ldots \ell_d$ whose rank is greater than $r_0 = \beta n$, and let $\{\ell_1, \ldots, \ell_r\}$ be a maximal independent set. We want to use the fact that since T is a product of at least r independent linear polynomials, most evaluations would be zero. We shall be choosing our $\mathcal{A}$ to be the set where all k-th order partial derivatives evaluate to zero.
On applying the product rule of differentiation, any k-th order derivative of T can be written as a sum of terms each of which is a product of at least $r - k$ independent linear polynomials. Let us count the ~~*erraneous*~~ *erroneous points* $\mathcal{E}_T \subseteq \mathbb{F}^{n^2}$ that keep at least $r - k$ of $\{\ell_1, \ldots, \ell_r\}$ non-zero, or in other words makes at most k of $\{\ell_1, \ldots, \ell_r\}$ zero.

$$\Pr_{\mathbf{a} \in \mathbb{F}^{n^2}}[\text{at most } k \text{ of } \ell_1, \ldots, \ell_r \text{ evaluate to zero}] \leq \sum_{i=0}^{k} \binom{r}{i} \left(\frac{1}{q}\right)^i \left(1 - \frac{1}{q}\right)^{r-i}$$

Hence, we can upper-bound $|\mathcal{E}_T|$ as

$$\begin{aligned}|\mathcal{E}_T| &\leq \sum_{i=0}^{k} \binom{r}{i} (q-1)^{r-i} q^{n^2-r}\\ &= O\left(k \cdot \binom{r}{k} \left(1 - \frac{1}{q}\right)^{r-k} q^{n^2}\right) \quad \text{if } r > qk\\ &= q^{n^2} \cdot \alpha^n \quad \text{for some } 0 < \alpha < 1.\end{aligned}$$

By choosing ~~$\mathcal{A} = \mathbb{F}_{n^2} \setminus \mathcal{E}$~~ $\mathcal{A} = \mathbb{F}^{n^2} \setminus \mathcal{E}$ where $\mathcal{E} = \bigcup_{T \text{ of large rank}}$ ~~$\mathcal{E}_t$~~$\mathcal{E}_T$, we have that $M_k(T; \mathcal{A})$ is just the zero matrix and hence $\Gamma^{[\mathrm{GK}]}_{k,\mathcal{A}}(T) = 0$.

Putting it together, if $C = T_1 + \cdots + T_s$, then

$$\Gamma^{[\mathrm{GK}]}_{k,\mathcal{A}}(C) \quad \leq \quad s \cdot q^{\beta n}. \tag{1}$$

where $\mathcal{A} = \mathbb{F}^{n^2} \setminus \mathcal{E}$ for some set $\mathcal{E}$ of size at most $s \cdot \alpha^n \cdot q^{n^2}$ for some $0 < \alpha < 1$.

7.3. Lower-bounding $\Gamma^{[\mathrm{GK}]}_{k,\mathcal{A}}$ for Det_n and Perm_n

We now wish to show that $M_k(\mathsf{Det}_n; \mathcal{A})$ has large rank. The original proof of Grigoriev and Karpinski is tailored specifically for the determinant, and does not extend directly to the permanent. The following argument is a proof communicated by Srikanth Srinivasan [Sri13] that involves an elegant trick that he attributes to ~~Koutis~~[Kou08] . The following proof is presented for the determinant, but immediately extends to the permanent as well.

Note that if we were to just consider $M_k(\mathsf{Det}_n)$, it would have been easy to show that the rank is full by looking at just those evaluation points that keep exactly one $(n-k) \times (n-k)$ minor non-zero (set the main diagonal of the minor to ones, and every other entry to zero). Hence, $M_k(\mathsf{Det}_n)$ has the identity matrix *embedded inside* and hence must be full rank. However, we are missing a few of the evaluations (since a small set $\mathcal{E}$ of evaluations is removed) and we would still like to show that the matrix continues to have full column-rank.

Lemma 22. *Let $p(X)$ be a non-zero linear combination of $r \times r$ minors of the matrix $X = ((x_{ij}))$. Then,*

$$\Pr_{A \in \mathbb{F}_q^{n^2}} [p(A) \neq 0] \quad \geq \quad \Omega(1)$$

This immediately implies that for every linear combinations of the columns of $M_k(\mathsf{Det}_n)$, a constant fraction of the coordinates have non-zero values. Since we are removing merely a set $\mathcal{E}$ of size $(1 - o(1))q^{n^2}$, there must continue to exist coordinates that are non-zero. In other words, no linear combination of columns of $M_k(\mathsf{Det}_n; \mathcal{A})$ results in the zero vector.

The proof of the above lemma would be an induction on the number of minors contributing to the linear combination. As a base case, we shall use a well-known fact about Det_n and Perm_n of random matrices.

Proposition 23. *If A is a random $n \times n$ matrix with entries from a fixed finite field $\mathbb{F}_q$, then for $q \neq 2$ we have*

$$\Pr[\det(A) \neq 0] \quad \geq \quad \frac{q-2}{q-1} \quad = \quad \Omega(1)$$

We shall defer the proof of this proposition for later, and proceed with the proof of Lemma 22.

Proof of Lemma 22. If $p(X)$ is a scalar multiple of a single non-zero minor, then we already have the lemma from Proposition 23. Hence, let us assume that there are at least two distinct minors participating in the linear combination $p(X)$. Without loss of generality, assume that the first row occurs in some of the minors, and does not in others. That is,

$$\begin{aligned} p(X) &= \left(\sum_{i:\mathrm{Row}_1 \in M_i} c_i M_i\right) \quad + \quad \left(\sum_{j:\mathrm{Row}_1 \notin M_j} c_j M_j\right) \\ &= (x_{11}M_1' + \cdots + x_{1n}M_n') \quad + \quad M'' \quad \text{(expanding along the first row)} \end{aligned}$$

To understand a random evaluation of $p(X)$, let us first set rows $2, \ldots, n$ to random values, and then setting row 1 to random values.

$$\begin{aligned} \Pr_A[p(A) \neq 0] \quad \geq \quad & \Pr[x_{11}M_1' + \cdots + x_{1n}M_n' + M'' \neq 0 \mid \text{some } M_i' \neq 0] \\ & \times \Pr[\text{some } M_i' \neq 0] \end{aligned}$$

Note that once we have set rows $2, \ldots, n$ to random values, $p(X)$ reduces to a linear polynomial in $\{x_{11}, \ldots, x_{1n}\}$. Further, a random evaluation of any non-constant linear polynomial is zero with probability exactly $\left(1 - \frac{1}{q}\right)$. Hence,

$$\begin{aligned} \Pr_A[p(A) \neq 0] \quad &\geq \quad \Pr[x_{11}M_1' + \cdots + x_{1n}M_n' + M'' \neq 0 \mid \text{some } M_i' \neq 0] \\ & \qquad \times \Pr[\text{some } M_i' \neq 0] \\ &= \quad \left(1 - \frac{1}{q}\right) \cdot \Pr[\text{some } M_i' \neq 0] \end{aligned}$$

Now comes Koutis' Trick: the term $\left(1 - \frac{1}{q}\right) \cdot \Pr[\text{ some } M_i' \neq 0]$ is exactly the probability that $x_{11}M_1' + \cdots + x_{1n}M_n'$ is non-zero! Hence,

$$\begin{aligned} \Pr_A[p(A) \neq 0] \quad &= \quad \Pr[x_{11}M_1' + \cdots + x_{1n}M_n' + M'' \neq 0] \\ &\geq \quad \Pr[x_{11}M_1' + \cdots + x_{1n}M_n' \neq 0] \\ &= \quad \Pr\left[\left(\sum_{i:\mathrm{Row}_1 \in M_i} c_i M_i\right) \neq 0\right] \end{aligned}$$

which is just the linear combination obtained by only considering those minors that contain the first row. Repeating this process for other rows/columns until only one minor remains, we have

$$\Pr_A[p(A) \neq 0] \quad \geq \quad \Pr_A[\det(A) \neq 0] \quad = \quad \frac{q-2}{q-1} \quad \text{(by Proposition 23)}$$

□

We now give a proof of Proposition 23.

Proof of Proposition 23. We shall fix random values to the first row of A. Then,

$$\begin{aligned} \Pr_A[\mathsf{Det}_n(A) = 0] \quad &\leq \quad \Pr[a_{11}M_1 + \cdots + a_{1n}M_n = 0 \mid \text{some } a_{1i} \text{ non-zero}] \\ &\qquad + \quad \Pr[a_{11} = \cdots = a_{1n} = 0] \\ &= \quad \Pr[a_{11}M_1 + \cdots + a_{1n}M_n = 0 \mid \text{some } a_{1i} \text{ non-zero}] \\ &\qquad + \quad \frac{1}{q^n} \end{aligned}$$

Whenever there is some a_{1i} that is non-zero, then $a_{11}M_1 + \cdots + a_{1n}M_n$ is a non-zero linear combination of minors. By a similar argument as in the proof of Lemma 22, we have that

$$\Pr[a_{11}M_1 + \cdots + a_{1n}M_n = 0 \mid \text{not all } a_{1i} \text{ are zero}] \quad \leq \quad \Pr[\mathsf{Det}_{n-1}(A) = 0]$$

Unfolding this recursion, we have

$$\begin{aligned} \Pr[\mathsf{Det}_n(A) = 0] \quad &\leq \quad \frac{1}{q} + \frac{1}{q^2} + \cdots + \frac{1}{q^n} \quad = \quad \frac{1}{q-1} \\ \implies \Pr[\mathsf{Det}_n(A) \neq 0] \quad &\geq \quad \left(1 - \frac{1}{q-1}\right) \; = \; \frac{q-2}{q-1} \end{aligned}$$

□

7.4. Putting it all together

Hence, if Det_n is computed by a depth-3 circuit of top fan-in s over $\mathbb{F}$, then

$$\begin{aligned} s \cdot q^{\beta n} \quad &= \quad \Omega\left(\binom{n}{k}^2\right) \\ &= \quad \Omega\left(2^{2H(\gamma)\cdot n}\right) \\ \implies \log s \quad &= \quad \Omega((2H(\gamma) - \beta \log q)n) \end{aligned}$$

where $H(\gamma)$ is the binary entropy function[6]. By choosing $\gamma < q^{-q/2}$, we can find a β such that $q\gamma < \beta$ (which was required in Section 7.2) and

[6] The binary entropy function is defined as $H(\gamma) \overset{\text{def}}{=} -\gamma \log_2(\gamma) - (1-\gamma)\log_2(1-\gamma)$. It is well known that $\binom{n}{k} \approx 2^{nH(k/n)}$.

$2H(\gamma) > \beta \log q$, yielding the lower bound

$$\begin{aligned} s &= \exp\left(\Omega(q^{-q/2} \cdot q \log q \cdot n)\right) \\ &= 2^{\Omega(n)} \end{aligned}$$

□(Theorem 20)

8. Lower bounds for multilinear models

Raz [Raz09] showed that multilinear formulas computing the Det_n or Perm_n must be of size $n^{\Omega(\log n)}$. The complexity measure used by Raz also led to exponential lower bounds for constant depth multilinear circuits [RY09] and super-linear lower bounds for syntactic multilinear circuits [RSY08]. We shall first give some intuition behind the complexity measure before actually seeing the lower bounds.

8.1. The partial derivative matrix

Intuition. A natural first step is to try the simpler task of proving lower bounds for depth-3 multilinear circuits.

$$f \quad = \ell_{11} \dots \ell_{1d} + \dots + \ell_{s1} \dots \ell_{sd}$$

The task is now to construct a measure Γ such that $\Gamma(\ell_1 \dots \ell_d)$ is small whenever each ℓ_i ~~are linear polynomials~~ is a linear polynomial and different ℓ_i's are over disjoint sets of variables. Consider the simplest case of $f = (a_1 + b_1 x)(a_2 + b_2 y)$. An observation is that the coefficients of f are given by the 2×2 matrix obtained as $[a_1 \ b_1]^T [a_2 \ b_2] = \begin{bmatrix} a_1 a_2 & a_1 b_2 \\ a_2 b_1 & b_1 b_2 \end{bmatrix}$. In other words, a polynomial $f = a_0 + a_1 x + a_2 y + a_3 xy$ factorizes into two variable disjoint factors if and only if the matrix $\begin{bmatrix} a_0 & a_1 \\ a_2 & a_3 \end{bmatrix}$ has rank 1. A straightforward generalization of this to multiple variables yields the *partial derivative matrix* (which was first introduced by Nisan [Nis91] in the context of non-commutative ABPs)

Definition 24. *For any given partition of variables $X = Y \sqcup Z$, define the* partial derivative matrix *$M_{Y,Z}(f)$ to be the matrix described as follows — the rows are indexed by monomials in Y, columns indexed by monomials in Z, and the (i,j)-th entry of the matrix is the coefficient of the monomial $m_i(Y) \cdot m_j(Z)$ in f. We shall use $\Gamma^{[\text{Raz}]}_{Y,Z}(f)$ to denote* $\mathrm{rank}(M_{Y,Z}(f))$. *Further, we shall call a polynomial f to be* full-rank *if $M_{Y,Z}(f)$ is full-rank.*

Here are some basic properties of the partial derivative matrix which would be extremely useful in later calculations.

Observation 25 (Sub-additivity). *For any partition $X = Y \sqcup Z$ and any pair of multilinear polynomials f and g in $\mathbb{F}[X]$ we have $\Gamma^{[\text{Raz}]}_{Y,Z}(f+g) \leq \Gamma^{[\text{Raz}]}_{Y,Z}(f) + \Gamma^{[\text{Raz}]}_{Y,Z}(g)$*

Proof. Follows from the linearity of the matrix. □

Observation 26 (Multiplicativity). *If $f_1 \in \mathbb{F}[Y_1, Z_1]$ and $f_2 \in \mathbb{F}[Y_2, Z_2]$ with $Y = Y_1 \sqcup Y_2$ and $Z = Z_1 \sqcup Z_2$, then*

$$\Gamma^{[\text{Raz}]}_{Y,Z}(f_1 \cdot f_2) \quad = \quad \Gamma^{[\text{Raz}]}_{Y_1,Z_1}(f_1) \quad \cdot \quad \Gamma^{[\text{Raz}]}_{Y_2,Z_2}(f_2)$$

Proof. It is not hard to see that $M_{Y,Z}(f_1 \cdot f_2)$ is the tensor product $M_{Y_1,Z_1}(f_1) \otimes M_{Y_2,Z_2}(f_2)$, and the rank of a tensor product of two matrices is the product of the ranks. □

Observation 27. $\Gamma^{[\text{Raz}]}_{Y,Z}(f) \quad \leq \quad 2^{\min(|Y|,|Z|)}$

Proof. The number of rows is $2^{|Y|}$ and number of columns is $2^{|Z|}$, and hence the rank is upper-bounded by the minimum. □

Let us get back to lower bounds for multilinear models, and attempt to use $\Gamma^{[\text{Raz}]}_{Y,Z}(f)$ defined above. Unfortunately, there are examples of simple polynomials like $f = (y_1 + z_1) \dots (y_n + z_n)$ with $\Gamma^{[\text{Raz}]}_{Y,Z}(f) = 2^n$. Raz's idea here was to look at $\Gamma^{[\text{Raz}]}_{Y,Z}(f)$ for a *random partition*, and show that with high probability the rank of the partial derivative matrix is far from full. As a toy example, we shall see why this has the potential to give lower bounds for depth-3 multilinear circuits.

Lemma 28. *Let $f(X) = \ell_1 \dots \ell_d$ be an n-variate multilinear polynomial. If $X = Y \sqcup Z$ is a random partition with $|Y| = |Z| = |X|/2$, then with high probability we have*

$$\Gamma^{[\text{Raz}]}_{Y,Z}(f) \quad \leq \quad 2^{|X|/2} \cdot 2^{-|X|/16}.$$

It is to be noted that we should expect a random polynomial to be full-rank with respect to any partition, so the measure $\Gamma^{[\text{Raz}]}_{Y,Z}(f)$ is expected to be $2^{|X|/2}$ which should yield a lower bound of $2^{\Omega(|X|)}$.

Sketch of Proof. Without loss of generality we can assume that each ℓ_i depends on at least two variables as removing the ℓ_i's that depend on just one variable does not alter $\Gamma^{[\text{Raz}]}_{Y,Z}(f)$ with respect to any partition. Let $|X| = n$. Using Observation 26, $\Gamma^{[\text{Raz}]}_{Y,Z}(f) \leq 2^d$ and hence if $d < n/3$ then we are done. Hence assume that $d \geq n/3$. By a simple averaging argument, there must hence be at least $d/4$ of the ℓ_i's that depend on at most 3 variables; we shall refer to these as the *small* ℓ_i's.

Since the partition is chosen at random, on expectation a quarter of the small ℓ_i's would have all its variables mapped to either Y or Z, hence not contributing to $\Gamma^{[\text{Raz}]}_{Y,Z}(f)$. Therefore, with high probability,

$$\Gamma^{[\text{Raz}]}_{Y,Z}(f) \quad \leq \quad 2^d \cdot 2^{-d/16} \quad \leq \quad 2^{n/2} \cdot 2^{-n/16}.$$

□

More generally, if $f = g_1(X_1)\dots g_t(X_t)$ where the X_i's are mutually disjoint, then a random partition is very unlikely to partition all the X_i's into almost equal parts. This shall be formalized in the next section to prove the lower bound for multilinear formulas.

8.2. Lower bound for multilinear formulas

We now present the lower bound for multilinear formulas due to [Raz09]. The first step of our roadmap is to find a suitable normal form for multilinear formulas. The normal form that we use is from the survey by Shpilka and Yehudayoff [SY10].

8.2.1. Formulas to log-product sums. The following structural lemma shows that any multilinear formula can be converted in to a small sum of *log-product* polynomials. The techniques of the following lemma can also be used in other settings with minor modifications, and we shall encounter a different version of this lemma later as well.

Definition 29. *A multilinear polynomial $f \in \mathbb{F}[X]$ is called a* multilinear log-product *polynomial if $f = g_1 \dots g_t$ and there exists a partition of variables $X = X_1 \sqcup \dots \sqcup X_t$ such that*

- $g_i \in \mathbb{F}[X_i]$ *for all $i \in [t]$.*
- $\frac{|X|}{3^i} \leq |X_i| \leq \frac{2|X|}{3^i}$ *for all i, and* $|X_t| = 1$

Lemma 30. *Let Φ be a multilinear formula of size s computing a polynomial p. Then f can be written as a sum of $(s+1)$ log-product multivariate polynomials.*

Proof. Similar to Lemma 19, let v be a node in Φ such that set of variables X_v that it depends on satisfies $\frac{|X|}{3} \leq |X_v| \leq \frac{2|X|}{3}$. If Φ_v is the polynomial computed at this node, then f can be written as

$$f \quad = \quad \Phi_v \cdot g_1 + \Phi_{v=0} \quad \text{for some } g_1 \in \mathbb{F}[X \setminus X_v].$$

where $\Phi_{v=0}$ is the formula obtained by replacing the node v by zero. Note that the subtree at the node v is completely disjoint from $\Phi_{v=0}$. Hence the sum of the sizes of Φ_v and $\Phi_{v=0}$ is at most s. Hence, $g_1 \in \mathbb{F}[X \setminus X_v]$ and $\frac{|X|}{3} \leq |X \setminus X_v| \leq \frac{2|X|}{3}$. Inducting on the formulas Φ_v and $\Phi_{v=0}$ gives the lemma. □

8.2.2. Log-products are far from full-rank on a random partition. The main technical part of the proof is to show that log-product multivariate polynomials are far from full-rank under a random partition of variables. This would let us show that a sum of log-product multivariate polynomials cannot be full rank unless it is a very large sum.

Main idea: Suppose $f = g_1 \dots g_t$ where each $g_i \in \mathbb{F}[X_i]$. Let $X = Y \sqcup Z$ be a random partition with $|Y| = |Z| = |X|/2$, and $Y_i = Y \cap X_i$ and $Z_i = Z \cap X_i$. Let $d_i = \left|\frac{|Y_i| - |Z_i|}{2}\right|$ measure the imbalance between the sizes of Y_i and Z_i, and we shall say X_i is k-imbalanced if $d_i \geq k$. Let $b_i = \frac{|Y_i| + |Z_i|}{2} = \frac{|X_i|}{2}$.

By Observation 26, we know that

$$\begin{aligned}\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f) &= \Gamma^{[\mathrm{Raz}]}_{Y_i,Z_i}(g_1)\dots\Gamma^{[\mathrm{Raz}]}_{Y_i,Z_i}(g_t)\\ &\leq 2^{\min(|Y_1|,|Z_1|)}\cdot\dots\cdot 2^{\min(|Y_t|,|Z_t|)}\\ &= 2^{b_1-d_1}\dots 2^{b_t-d_t} = \frac{2^{|X|/2}}{2^{d_1+\dots+d_t}}.\end{aligned}$$

Hence, even if one of the X_i's is a little imbalanced, then the product is far from full-rank.

Lemma 30 shows that the size of X_i decreases slowly with i, and it is not hard to show that $|X_i| \geq \sqrt{|X|}$ for $i \leq t' \stackrel{\mathrm{def}}{=} \frac{\log|X|}{100}$. We wish to show that the probability that none of g_i (for $i \leq t'$) is k-unbalanced for $k = |X|^{1/20}$ is very small. Let $\mathcal{E}_i$ be the event that X_i is not k-unbalanced. The goal is to upper bound the probability that all the events $\mathcal{E}_i$ hold. These probability calculations would follow from this lemma about the *hypergeometric distribution.*

Hypergeometric Distribution: Fix parameters $n, g, r \geq 0$, and let $G \subseteq [n]$ with $|G| = g$. Informally, the hypergeometric distribution is the distribution obtained on the intersection sizes of a random set of size r with a fixed set of size g from a universe of size n. Formally, the random variable $\mathcal{H}(n,g,r)$ is defined as:

$$\Pr\left[\mathcal{H}(n,g,r) = k\right] \quad = \quad \Pr_{R\subseteq[n],|R|=r}\left[|R\cap G| = k\right] = \frac{\binom{g}{k}\binom{n-g}{r-k}}{\binom{n}{r}}$$

The following lemma shows that for a fairly large range of parameters, the hypergeometric distribution does not put too much mass on any value.

Lemma 31. *Let n, g, r be parameters such that $\frac{n}{4} \leq r \leq \frac{3n}{4}$ and $0 \leq g \leq \frac{2n}{3}$. Then for any $t \leq g$,*

$$\Pr\left[\mathcal{H}(n,g,r) = t\right] \quad\leq\quad O\left(\frac{1}{\sqrt{g}}\right).$$

The proof of this lemma follows from standard binomial coefficient estimates on the probability.

Let us go back to estimating the probability that all the events $\mathcal{E}_i$ hold.

$$\Pr\left[\mathcal{E}_1 \wedge \dots \wedge \mathcal{E}_{t'}\right] \quad=\quad \Pr[\mathcal{E}_1]\cdot\Pr[\mathcal{E}_2 \mid \mathcal{E}_1]\cdots\Pr[\mathcal{E}_{t'} \mid \mathcal{E}_1 \wedge\dots\wedge\mathcal{E}_{t'-1}].$$

The event $\mathcal{E}_1$ is just the probability that a random set Y of size $|X|/2$ intersects X_1 in t places where $t \in \left[\frac{|X_i|}{2} - k, \frac{|X_i|}{2} - k\right]$. This is just a particular setting of the hypergeometric distribution and Lemma 31 asserts that

$$\Pr[\mathcal{E}_1] \quad\leq\quad O\left(\frac{2k}{\sqrt{|X|}}\right).$$

To apply a similar bound for the other terms, consider the event $\mathcal{E}_i$ given that $\mathcal{E}_1, \ldots, \mathcal{E}_{i-1}$ hold. Let $X' = X \setminus (X_1 \cup \ldots \cup X_{i-1})$ and $Y' = Y \cap X'$. The fact that $\mathcal{E}_1, \ldots, \mathcal{E}_{i-1}$ hold means that the partition has been fairly balanced in the first $(i-1)$ parts and hence $|Y'| \leq |X'| + ik$. Hence, we would still be in the range of parameters in Lemma 31 to also get that

$$\begin{aligned} \forall i \leq t' \quad \Pr[\mathcal{E}_i \mid \mathcal{E}_1 \wedge \cdots \wedge \mathcal{E}_{i-1}] &\leq O\left(\frac{2k}{\sqrt{|X|}}\right) \\ \implies \Pr\left[\mathcal{E}_1 \wedge \cdots \wedge \mathcal{E}_{t'}\right] &\leq |X|^{-\varepsilon \log |X|} \quad \text{for some } \varepsilon > 0 \\ \implies \Pr\left[\Gamma^{[\mathrm{Raz}]}_{Y,Z}(g_1 \ldots g_t) \leq 2^{(|X|/2) - |X|^{1/20}}\right] &\leq |X|^{-\varepsilon \log |X|}. \end{aligned}$$

Hence, if $g_1 \ldots g_t$ is a log-product multilinear polynomial, then with probability at least $\left(1 - |X|^{-\varepsilon \log |X|}\right)$ we have that $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(g_1 \ldots g_t) \leq 2^{(|X|/2) - |X|^{1/20}}$. Further, if f is computable by a multilinear formula of size s then, by Lemma 30, f can be written as a sum of $(s+1)$ log-product multilinear polynomials. Hence, with probability at least $\left(1 - (s+1)|X|^{-\varepsilon \log |X|}\right)$ we have that

$$\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f) \quad \leq \quad (s+1) \cdot 2^{(|X|/2) - |X|^{1/20}}.$$

Hence, if $(s+1) < |X|^{(\varepsilon/2) \log |X|}$, then with high probability a random partition would ensure $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f) \ll 2^{|X|/2}$. Let us record this as a lemma.

Lemma 32. *Let $f \in \mathbb{F}[X]$ be computed by a multilinear formula of size $s < |X|^{(\varepsilon/2) \log |X|}$ for a small enough constant $\varepsilon > 0$. Then with probability at least $(1 - |X|^{-(\varepsilon/2) \log |X|})$ we have*

$$\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f) \quad \leq \quad (s+1) \cdot 2^{|X|/2} \cdot 2^{-|X|^{1/20}}$$

for a random partition $X = Y \sqcup Z$ with $|Y| = |Z| = |X|/2$.

8.2.3. Det_n and Perm_n have large rank. The last step of the proof would be to find an explicit polynomial whose partial derivative matrix under a random partition has large rank. As earlier, our candidate polynomial would be Det_n or Perm_n. Unfortunately, both these polynomials are over n^2 variables and degree n. It is not hard to verify that the rank of the partial derivative matrix of Det_n or Perm_n can never be greater than 2^{2n}. Hence directly using Lemma 32, we would have $2^{O(n)}$ competing with $2^{n^2/2 - n^{O(1)}}$ which is simply futile. A simple fix is to first randomly restrict ourselves to fewer variables and then apply Lemma 32.

Let $m = n^{1/3}$. Let σ be a random restriction that assigns random values to all but $2m$ randomly chosen variables. We shall call this set of $2m$ variables as X, and randomly partition this into two sets Y and Z of size m each. Hence, $\sigma(\mathsf{Det}_n)$ reduces to a multilinear polynomial over $2m$ variables. It is also worth noting that a multilinear formula remains a multilinear formula under this restriction. The following claim is easy to verify.

Claim 33. *With probability at least $1/2$, the variables in X belong to distinct rows and columns.*

We shall restrict ourselves to only these random restrictions, and without loss of generality let the sets be $Y = \{x_{1,1}, x_{3,3}, \ldots, x_{2m-1,2m-1}\}$ and $Z = \{x_{2,2}, x_{4,4}, \ldots, x_{2m,2m}\}$. For ease of notation, we shall refer to $x_{2i-1,2i-1}$ as y_i and $x_{2i,2i}$ as z_i for $i = 1, \ldots, m$.
Consider the following restriction:

$$\begin{aligned} f &= \mathsf{Det}\begin{bmatrix} y_1 & 1 & & & & & & \\ 1 & z_1 & & & & & & \\ & & \ddots & & & & & \\ & & & y_m & 1 & & & \\ & & & 1 & z_m & & & \\ & & & & & 1 & & \\ & & & & & & \ddots & \\ & & & & & & & 1 \end{bmatrix} \\ &= (y_1 z_1 - 1) \ldots (y_m z_m - 1) \end{aligned}$$

It is easy to check that $\Gamma_{Y,Z}^{[\mathrm{Raz}]}(f) = 2^m$. Although this is a single restriction with large rank, the ~~Schwarz-Zippel~~ Schwartz-Zippel-DeMillo-Lipton lemma immediately gives that random restriction would also have rank 2^m with high probability[7]. We shall record this as a lemma.

Lemma 34. *With probability at least* $1/100$*, we have that* $\Gamma_{Y,Z}^{[\mathrm{Raz}]}(\sigma(\mathsf{Det}_n)) = 2^m$ *where* σ *is a random restriction to* $2m$ *variables for* $m = n^{1/3}$.

Combining Lemma 34 with Lemma 32, we have the following theorem.

Theorem 35 ([Raz09])**.** *Any multilinear formula computing* Det_n *or* Perm_n *must be of size* $n^{\Omega(\log n)}$. □

8.3. Stronger lower bounds for constant depth multilinear formulas

Looking back at Lemma 32, we see that whenever $f(X)$ is computable by a size s multilinear formula $\Gamma_{Y,Z}^{[\mathrm{Raz}]}(f)$ is exponentially smaller than $2^{|X|/2}$ with probability $\left(1 - s \cdot |X|^{-\varepsilon \log |X|}\right)$. Hence we had to settle for a $n^{\Omega(\log n)}$ lower bound not because of the rank deficit but rather because of the bounds in the probability estimate. Unfortunately, this lower bound technique cannot yield a better lower bound for multilinear formulas as there are explicit examples of polynomials computable by poly-sized multilinear circuits with $\Gamma_{Y,Z}^{[\mathrm{Raz}]}(f) = 2^{|X|/2}$ under *every* partition [Raz06]. However, the probability bound can be improved in the case of constant depth multilinear circuits to give stronger lower bounds.

Note that Lemma 32 was proved by considering *multilinear log-products* (Definition 29) as the building blocks. To show that a multilinear log product $g_1(X_1) \ldots g_\ell(X_\ell)$ has small rank under a random partition, we argued that the probability that all the X_i's are partitioned in a roughly balanced fashion

[7] provided the underlying field is large, but this isn't really a concern as we can work with a large enough extension if necessary

is quite small. This was essentially done by thinking of this as $\ell = O(\log n)$ close-to-independent events, each with probability $1/\mathrm{poly}(n)$.
If ℓ was much larger than $\log n$ (with other parameters being roughly the same), it should be intuitively natural to expect a much lower probability of all the X_i's being partitioned in a roughly balanced manner. This indeed is the case for constant depth multilinear circuits, and we briefly sketch the key points where they differ from the earlier proof. The first is an analogue of Definition 29 in this setting.

Definition 36. *A multilinear polynomial f is said to be a* multilinear t-product *if f can be written as $f = g_1 \ldots g_t$ with the following properties:*

- *The variable sets of the g_i are mutually disjoint*
- *Each g_i non-trivially depends on at least t variables*

Lemma 37. *Let f be a multilinear polynomial of degree d over n variables that is computed by a depth-Δ multilinear formula Φ of size s. Then, f can be written as a sum of at most s multilinear t-products for $t = (n/100)^{1/2\Delta}$, and a multilinear polynomial of degree at most $n/100$.*

Proof. If $d < n/100$, then the lemma is vacuously true. Since Φ is a formula of depth Δ and computes a polynomial of degree $d > n/100$, there must be at least one product gate v of fan-in at least ~~$\left(\frac{n}{100}\right)^{1/\Delta}$~~ $\left(\frac{n}{100}\right)^{1/\Delta} = t^2$. Then similar to Lemma 30,

$$f \quad = \quad \Phi_v \cdot f' + \Phi_{v=0}$$

As Φ_v is a product of t^2 polynomials, by grouping the factors together we have that $\Phi_v \cdot f'$ is a multilinear t-product. Further, $\Phi_{v=0}$ is a multilinear polynomial that is computable by a depth-Δ formula of smaller size and we can induct on $\Phi_{v=0}$. □

Lemma 38. *Let $f(X)$ be an n-variate polynomial computed by a depth-Δ multilinear formula of size s. If $X = Y \sqcup Z$ is a randomly chosen partition with $|Y| = |Z| = n/2$, then with probability at least $(1 - s \cdot \exp(-n^{\Omega(1/\Delta)}))$ we have*

$$\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f) \quad \leq \quad (s+1) \cdot 2^{n/2} \cdot \exp(-n^{\Omega(1/\Delta)})$$

Sketch of Proof. By Lemma 37, we have that f can be written as $g_0 + g_1 + \cdots + g_s$ where $\deg(g_0) \leq n/100$ and $g_1, \ldots, g_s$ are multilinear t-products. Note that since g_0 is a multilinear polynomial of degree at most $(n/100)$, the number of monomials in g_0 is at most $\binom{n}{n/100} \leq 2^{n/10}$. Hence, $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(g_0) \leq 2^{n/10}$.
For the other g_i's, we can bound the probability that $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(g_i)$ is large in a very similar fashion as in Lemma 32, as the probability that all the factors of g_i are partitioned in a balanced manner is roughly the intersection of t independent events. By very similar estimates, this probability can be bounded by $(1/\mathrm{poly}(n))^t$. Hence, with high probability

$$\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f) \quad \leq \quad \Gamma^{[\mathrm{Raz}]}_{Y,Z}(g_0) + \cdots + \Gamma^{[\mathrm{Raz}]}_{Y,Z}(g_s) \quad \leq \quad (s+1) \cdot 2^{n/2} \cdot \exp(-n^{\Omega(1/\Delta)}).$$

□

Combining Lemma 38 with Lemma 34, we have the following theorem of Raz and Yehudayoff.

Theorem 39 ([RY09]). *Any multilinear formula of depth Δ computing* Det_n *or* Perm_n *must be of size* $\exp(n^{\Omega(1/\Delta)})$. □

9. Lower bounds for depth-4 circuits

This section shall address a recent technique for proving lower bounds for some depth-4 circuits.

Definition 40. *A* depth-4 circuit*, also referred to as a* $\Sigma\Pi\Sigma\Pi$ *circuit, computes a polynomial of the form*

$$f \quad = \quad Q_{11}\dots Q_{1d} \quad + \cdots + \quad Q_{s1}\dots Q_{sd}$$

The number of summands s is called the top fan-in *of the circuit.*
Further, a $\Sigma\Pi^{[a]}\Sigma\Pi^{[b]}$ *circuit is a depth-4 circuit computing a polynomial of the form*

$$f \quad = \quad Q_{11}\dots Q_{1a} \quad + \cdots + \quad Q_{s1}\dots Q_{sa} \quad \text{where } \deg Q_{ij} \le b \text{ for all } i,j.$$

9.1. Significance of the model

In a surprising series of results on depth reduction, Agrawal and Vinay [AV08] and subsequent strengthenings of Koiran [Koi12] and Tavenas [Tav13] showed that depth-4 circuits more or less capture the complexity of general circuits.

Theorem 41 ([AV08, Koi12, Tav13]). *If f is an n variate degree-d polynomial computed by a size s arithmetic circuit, then f can also be computed by a* $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ *circuit of size* $\exp\left(O(\sqrt{d}\log s)\right)$.

Conversely, if an n-variate degree-d polynomial requires $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ *circuits of size* $\exp\left(\Omega(\sqrt{d}\log s)\right)$*, then it requires arbitrary depth arithmetic circuits of size* $n^{\Omega(\log s/\log n)}$ *to compute it.*

Thus proving strong enough lower bounds for this special case of depth-4 circuits imply lower bounds for general circuits. The main results of the section is some recent lower bound [GKKS13, KSS13, FLMS13] that comes very close to the required threshold.

9.2. Building the complexity measure

As a simpler task, let us first attempt to prove lower bounds for expressions of the form

$$f \quad = \quad Q_1^d \quad + \cdots + \quad Q_s^d$$

where each of the Q_i's are quadratics. This is exactly the problem studied by Kayal [Kay12], which led to the complexity measure for proving depth-4 lower bounds.

The goal is to construct a measure Γ such that $\Gamma(f)$ is small whenever f is a power of a quadratic. As a first attempt, let us look at the space of k-th order partial derivatives of Q^d (for a suitable choice of k). Unlike the case of $\Sigma\wedge\Sigma$-circuits where the the space of k-th order partial derivatives of ℓ^d had dimension 1, the space of partial derivatives of Q^d could be as large as it can be expected. Nevertheless, the following simple observation would provide the key intuition.

Observation 42. *Any k-th order partial derivative of Q^d is of the form $Q^{d-k}p$ where p is a polynomial of degree at most k. Hence, if $k \ll d$, then all k-th order partial derivatives of Q^d share large common factors.*

This suggests that instead of looking at linear combinations of the partial derivatives of Q^d, we should instead be analysing *low-degree polynomial combinations* of them.

Definition 43. *Let $\partial^{=k}(f)$ refer to the set of all k-th order partial derivatives of f, and $\mathbf{x}^{\leq \ell}$ refer to the set of all monomials of degree at most ℓ. The* shifted partials *of f, denoted by $\left\langle \partial^{=k}\left(f\right)\right\rangle_{\leq \ell}$, is the vector space spanned by $\left\{\mathbf{x}^{\leq \ell} \cdot \partial^{=k}(f)\right\}$. The dimension of this space shall be denoted by $\Gamma^{[\mathrm{Kay}]}_{k,\ell}(f)$.*

The above observation shows that any element of $\left\langle \partial^{=k}\left(Q^d\right)\right\rangle_{\leq \ell}$ is divisible by Q^{d-k} and we thereby have the following lemma.

Lemma 44. *If $f = Q^d$ where Q is a quadratic, then $\Gamma^{[\mathrm{Kay}]}_{k,\ell}(f) \leq \binom{n+k+\ell}{n}$, the number of monomials of degree $(k+\ell)$.*

Note that if f was instead a random polynomial, we would expect the measure $\dim\left(\left\langle \partial^{=k}\left(f\right)\right\rangle_{\leq \ell}\right)$ to be about $\binom{n+k}{n} \cdot \binom{n+\ell}{n}$, which is *much* larger than $\binom{n+k+\ell}{n}$ for suitable choice of k, ℓ. Hence this measure $\Gamma^{[\mathrm{Kay}]}_{k,\ell}$ is certainly potentially useful for this model. Very similar to the above lemma, one can also show the following upper bound for the *building blocks* of $\Sigma\Pi^{[a]}\Sigma\Pi^{[b]}$ circuits.

Lemma 45. *Let $f = Q_1 \ldots Q_a$ with $\deg Q_i \leq b$ for all i. Then,*

$$\Gamma^{[\mathrm{Kay}]}_{k,\ell}(f) \quad = \quad \dim\left(\left\langle \partial^{=k}\left(f\right)\right\rangle_{\leq \ell}\right) \quad \leq \quad \binom{a}{k}\binom{n+(b-1)k+\ell}{n}.$$

It is easy to check that $\Gamma^{[\mathrm{Kay}]}_{k,\ell}$ is a sub-additive measure, and we immediately have this corollary.

Corollary 46. *Let f be an n-variate polynomial computed by a $\Sigma\Pi^{[a]}\Sigma\Pi^{[b]}$ circuit of top fan-in s. Then,*

$$\Gamma^{[\mathrm{Kay}]}_{k,\ell}(f) \quad \leq \quad s \cdot \binom{a}{k}\binom{n+(b-1)k+\ell}{n}.$$

Or in other words for any choice of k, ℓ, we have that any $\Sigma\Pi^{[a]}\Sigma\Pi^{[b]}$ circuit computing a polynomial f must have top fan-in s at least

$$\frac{\Gamma^{[\mathrm{Kay}]}_{k,\ell}(f)}{\binom{a}{k}\binom{n+(b-1)k+\ell}{n}}.$$

Intuition from algebraic geometry. Another perspective for the shifted partial derivatives comes from algebraic geometry. Any zero $a \in \mathbb{F}^n$ of Q is a zero of *multiplicity* d of Q^d. This implies that the set of common zeroes of all k-th order partial derivatives of Q^d (for $k \approx \sqrt{d}$) is *large*. On the other hand if f is a random polynomial, then with high probability there are no roots of large multiplicity.

In algebraic geometry terminology, the common zeroes of a set of polynomials is called the *variety* of the ideal generated by them. Further there is also a well-defined notion of a *dimension of a variety* which measures how large a variety is. Let $\mathbb{F}[\mathbf{x}]_{\leq r}$ refer to the set of polynomials of degree at most r, and let $\gamma_I(r) = \dim\left(I \cap \mathbb{F}[\mathbf{x}]_{\leq r}\right)$. Intuitively, if $\gamma_I(r)$ is large, then there are *many constraints* and hence the variety is *small*. In other words the growth of $\gamma_I(r)$ is inversely related to the dimension of the variety of I, and this is precisely captured by what is known as the *Affine Hilbert function of I*. More about the precise definitions of the Affine Hilbert function etc. can be found in any standard text in algebraic geometry such as [CLO07].

In our setting, the ideal we are interested in is $I = \left\langle \partial^{=k} f \right\rangle$. If f is a homogeneous polynomial, then $I \cap \mathbb{F}[\mathbf{x}]_{\leq r} = \left\langle \partial^{=k}(f) \right\rangle_{\leq \ell}$ where $\ell = r - (\deg(f) - k)$. Hence studying the dimension of shifted partial derivatives is exactly studying $\gamma_I(r)$ which holds all information about the dimension of the variety.

9.3. Lower bounding shifted partials of explicit polynomials

For a random polynomial $R(\mathbf{x})$, we would expect that

$$\Gamma^{[\mathrm{Kay}]}_{k,\ell}(R) \quad \approx \quad \min\left\{ \binom{n+\ell+d-k}{n}, \binom{n+k}{n}\binom{n+\ell}{n} \right\}$$

The terms on the RHS correspond to trivial upper bounds, where first term is the total number of monomials of degree $(\ell + d - k)$ and the second term is the total number shifted partials.

Claim 47. *For $k = \varepsilon\sqrt{d}$ for a small enough $\varepsilon > 0$, and $\ell = \frac{cn\sqrt{d}}{\log n}$ for a large enough constant c, we have*

$$\frac{\min\left\{ \binom{n+\ell+d-k}{n}, \binom{n+k}{n}\binom{n+\ell}{n} \right\}}{\binom{O(\sqrt{d})}{k}\binom{n+(\sqrt{d}-1)k+\ell}{n}} \quad = \quad 2^{\Omega(\sqrt{d}\log n)}$$

The proof of this claim is easily obtained by using standard asymptotic estimates of binomial coefficients. Note that using Corollary 46, the above claim shows that if we can find an explicit polynomial whose dimension of shifted partials are as large as above, then we would have an $\exp(\Omega(\sqrt{d}\log n))$ lower bound for the top fan-in of $\Sigma\Pi^{[\sqrt{d}]}\Sigma\Pi^{[\sqrt{d}]}$ circuits computing this polynomial.

If we have a set of polynomials with distinct leading monomials, then they are clearly linearly independent. Hence one way of lower bounding the dimension of a space of polynomials is to find a sufficiently large set of polynomials with distinct monomials in the space. The vector space of polynomials we are interested is $\left\langle \partial^{=k}\left(f\right)\right\rangle_{\leq \ell}$, and if we choose a structured polynomial f we can hope to be able to estimate the number of distinct leading monomials in this vector space.

9.3.1. Shifted partials of the determinant and permanent. The first lower bound for $\Sigma\Pi^{[\sqrt{d}]}\Sigma\Pi^{[\sqrt{d}]}$ circuits was by Gupta, Kamath, Kayal and Saptharishi [GKKS13] for the determinant and the permanent polynomial. We shall describe the lower bound for Det_n, although it would carry over immediately to Perm_n as well. As mentioned earlier, we wish to estimate the number of distinct leading monomials in $\left\langle \partial^{=k}\left(\mathsf{Det}_n\right)\right\rangle_{\leq \ell} = \operatorname{span}\left\{\mathbf{x}^{\leq \ell}\partial^{=k}\mathsf{Det}_n\right\}$. [GKKS13] made a relaxation to merely count the number of distinct leading monomials among the generators $\left\{\mathbf{x}^{\leq \ell}\partial^{=k}\mathsf{Det}_n\right\}$ instead of their span.

The first observation is that any k-th order partial derivative of Det_n is just an $(n-k)\times(n-k)$ minor. Let us fix a monomial ordering induced by the lexicographic ordering on the variables:

$$x_{11} \succ x_{12} \cdots \succ x_{1n} \succ x_{21} \succ \cdots \succ x_{nn}.$$

Under this ordering, the leading monomial of any minor is just the product of variables on the main diagonal of the sub-matrix corresponding to the minor, and hence is a term of the form $x_{i_1 j_1}\dots x_{i_{(n-k)},j_{(n-k)}}$ where $i_1 < \cdots < i_{n-k}$ and $j_1 < \cdots < j_{n-k}$; let us call such a sequence of indices as an $(n-k)$-increasing sequences in $[n]\times[n]$. Further, for any $(n-k)$-increasing sequence, there is a unique minor M whose leading monomial is precisely the product of the variables indexed by the increasing sequence. Therefore, the task of lower bounding distinct leading monomials in $\left\{\mathbf{x}^{\leq \ell}\partial^{=k}\mathsf{Det}_n\right\}$ reduces to the following combinatorial problem.

Claim 48. *For any $k, \ell > 0$, we have*

$$\Gamma^{[\mathrm{Kay}]}_{k,\ell}(\mathsf{Det}_n) \quad \geq \quad \#\left\{\begin{array}{c}\textit{monomials of degree } (\ell+n-k) \textit{ that}\\ \textit{contain an } (n-k)\textit{-increasing sequence}\end{array}\right\}.$$

We could start with an $(n-k)$-increasing sequence, and multiply by a monomial of degree ℓ to obtain a monomial containing an increasing sequence. Of course, the issue is that this process is not invertible and hence we might overcount. To fix this issue, [GKKS13] assign a *canonical increasing sequence* to every monomial that contains an increasing sequence and multiply by monomials of degree ℓ that ~~does~~ do not change the canonical increasing sequence.

Definition 49. *Let $D_2 = \{x_{1,1},\dots,x_{n,n},x_{1,2},x_{2,3},\dots,x_{n-1,n}\}$, the main diagonal and the diagonal just above it. For any monomial m define the* canonical increasing sequence of m, *denoted by* $\chi(m)$, *as* $(n-k)$*-increasing sequence of* m *that is entirely contained in* D_2 *and is ordered highest according to the ordering '*$\succ$*'. If* m *contains no* $(n-k)$*-increasing sequence entirely in* D_2*, then we shall say the canonical increasing sequence is empty.*

The reason we restrict ourselves to D_2 is because it is easier to understand which monomials change the canonical increasing sequence and which monomials do not.

Lemma 50. *Let* S *be an* $(n-k)$*-increasing sequence completely contained in* D_2*, and let* m_S *be the monomial obtained by multiplying the variables indexed by* S*. There are at least* $(2(n-k)-1)$ *variables in* D_2 *such that if* m *is any monomial over these variables, then* $\chi(m_S) = \chi(m \cdot m_S)$*.*

Proof. Note that for any $x_{i,j} \in D_2$ other than $x_{n,n}$, exactly one of $x_{i+1,j}$ or $x_{i,j+1}$ is in D_2 as well; let us refer to this element in D_2 as the *companion* of $x_{i,j}$. It is straightforward to check that for any $(n-k)$-increasing sequence S, the elements of S and their companions do not alter the canonical increasing sequence. □

It is a simple exercise to check that the number of $(n-k)$-increasing sequences contained in D_2 is $\binom{n+k}{2k}$. Further, as we are free to use the n^2-2n+1 variables outside D_2, and the $2(n-k)-1$ variables that don't alter the canonical increasing sequence, we have the following lemma.

Lemma 51. *For any* $k, \ell \geq 0$*,*

$$\dim\left(\left\langle \partial^{=k}\left(\mathsf{Det}_n\right)\right\rangle_{\leq \ell}\right) \quad \geq \quad \binom{n+k}{2k}\binom{(n^2-2n+1)+2(n-k)-1+\ell}{\ell}.$$

Although this lower bound is not as large as expected for a random polynomial, this is still sufficient to give strong lower bounds for depth-4 circuits. By choosing $k = \varepsilon\sqrt{n}$ for a small enough $\varepsilon > 0$, and $\ell = n^2\sqrt{n}$, Lemma 51 with Corollary 46 yields the lower bound of Gupta, Kamath, Kayal and Saptharishi [GKKS13]

Theorem 52. *Any* $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ *circuit computing* Det_n *or* Perm_n *has top fanin* $2^{\Omega(\sqrt{n})}$*.* □

It is worth noting that although Claim 47 suggests that we should be able to obtain a lower bound of $\exp(\Omega(\sqrt{n}\log n))$ for Det_n, [GKKS13] also showed that the above estimate for the dimension of shifted partial derivatives for the determinant is fairly tight. Hence the dimension of shifted partials cannot give a stronger lower bound for the determinant polynomial. However, it is possible that the estimate is *not* tight for the permanent and the dimension of shifted partial derivatives of the permanent is provably strictly larger than that of the determinant! It is conceivable that one should be able to prove an $\exp(\Omega(\sqrt{n}\log n))$ lower bound for the permanent using this measure.
Indeed, subsequently an $\exp(\Omega(\sqrt{d}\log n))$ was proved [KSS13, FLMS13] for other explicit polynomials which we now outline.

9.3.2. Shifted partials of the Nisan-Wigderson polynomial. Very shortly after [GKKS13]'s $2^{\Omega(\sqrt{n})}$ lower bound, Kayal, Saha and Saptharishi [KSS13] gave a stronger lower bound for a different polynomial. Their approach was to engineer an explicit polynomial F for which the dimension of shifted partial derivatives is easier to estimate. The main idea was that, if any k-th order partial derivative of the engineered polynomial is a monomial, then once again estimating $\dim\left(\left\langle\partial^{=k}(F)\right\rangle_{\leq \ell}\right)$ reduces to a monomial counting problem. If we could ensure that no two monomials of F have a gcd of degree k or more, then we would immediately get that all k-th order partial derivatives of F are just monomials (albeit possibly zero). If we were to interpret the set of non-zero monomials of F as just subsets over the variables, then the above constraint can be rephrased as a set system with *small pairwise intersection.* Such systems are well studied and are known as Nisan-Wigderson designs [NW94]. With this in mind, [KSS13] studied the following polynomial family inspired by an explicit construction of a Nisan-Wigderson design.

Definition 53 (Nisan-Wigderson Polynomial). *. Let n be a power of 2 and let $\mathbb{F}_n$ be the finite field with n elements that are identified with the set $\{1,\ldots,n\}$. For any $0 \leq k \leq n$, the polynomial NW_k is a n^2-variate polynomial of degree n defined as follows:*

$$\mathrm{NW}_k(x_{1,1},\ldots,x_{n,n}) \quad = \quad \sum_{\substack{p(t)\,\in\,\mathbb{F}_n[t]\\ \deg(p)\,<\,k}} x_{1,p(1)}\cdots x_{n,p(n)}.$$

It is easy to show that the above family of polynomials is in VNP. Further, since any two distinct univariate polynomials of degree less than k intersects in less than k places, we have the following observation.

Observation 54. *Any two monomials of NW_k intersect in less than k variables. Hence, any k-th order partial derivative of $\mathrm{NW}_k(\mathbf{x})$ is a monomial (which could possibly be zero).* □

Hence, the problem of lower bounding the shifted partials of NW_k reduces to the problem of counting distinct monomials of degree $\ell + d - k$ that are divisible by one of these k-th order derivatives. [KSS13] additionally used the observation that two random k-th order partial derivatives of NW_k are monomials that are *far* from each other. Using this, they estimate the number of distinct shifts of these monomials and showed that the dimension of shifted partial derivatives of NW_k is very close to the trivial upper bound as in Claim 47. We sketch the argument by Chillara and Mukhopadhyay [CM14]. Formally, for any two multilinear monomials m_1 and m_2, let the $\Delta(m_1, m_2)$ denote $\min\{|m_1| - |m_1 \cap m_2|, m_2 - |m_1 \cap m_2|\}$ (abusing notation by identifying the multilinear monomials with the set of variables that divide it).

Lemma 55 ([CM14]). *Let $m_1, \ldots, m_s$ be monomials over N variables such that $\Delta(m_i, m_j) \geq d$ for all $i \neq j$. Then the number of distinct monomials that may be obtained by multiplying some m_i by arbitrary monomials of degree ℓ is at least $s\binom{N+\ell}{N} - \binom{s}{2}\binom{N+\ell-d}{N}$.*

Proof. For $i = 1, \ldots, s$, let A_i be the set of monomials that can be obtained by multiplying m_i with a degree ℓ monomial. By inclusion-exclusion,

$$\left|\bigcup_{i=1}^{s} A_i\right| \quad \geq \quad \sum_{i=1}^{s} |A_i| \quad - \quad \sum_{i<j} |A_i \cap A_j| \,.$$

Note that each A_i is of size exactly $\binom{N+\ell}{N}$. Further, since $\Delta(m_i, m_j) \geq d$, any monomial that is divisible by m_i and m_j must necessarily be divisible by ~~m_1~~ m_i and the variables in ~~m_2 not in m_1~~ m_j not in m_i. Hence, $|A_i \cap A_j| \leq \binom{N+\ell-d}{N}$. The lemma follows by substituting these above. □

Note that any two distinct monomials of NW_k intersect in at most k places. For each monomial m_i of NW_k, let m_i' be any non-zero k-th order partial derivative of m_i. Therefore, $\Delta(m_i', m_j') \geq n - 2k \geq \frac{n}{2}$ for $k = \varepsilon\sqrt{n}$. Since we have n^k monomials of pairwise distance at least $n/2$, the above lemma immediately yields a lower bound for the shifted partials of NW_k.

Theorem 56 ([KSS13]). *Let $k = \varepsilon\sqrt{d}$ for some constant $\varepsilon > 0$. Then for any $\ell = \Theta\left(\frac{n^2\sqrt{n}}{\log n}\right)$,*

$$\dim\left(\left\langle \partial^{=k}\left(\mathrm{NW}_k\right)\right\rangle_{\leq \ell}\right) \quad \geq \quad \frac{n^k}{2} \cdot \binom{n^2+\ell}{n^2}$$

Sketch of Proof. As mentioned earlier, we have n^k monomials $\{m_i'\}$ with pairwise distance at least $\frac{n}{2}$. Using Lemma 55, it suffices to show that

$$n^k \cdot \binom{n^2+\ell}{n^2} \quad \geq \quad 2 \cdot \binom{n^k}{2} \cdot \binom{n^2+\ell-\frac{n}{2}}{n^2}$$

and this follows easily from standard binomial coefficient estimates. □

Combining with Corollary 46, we have the lower bound of [KSS13] using standard estimates.

Theorem 57 ([KSS13]). *Any $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ computing the NW_k polynomial, where $k = \varepsilon\sqrt{n}$ for a sufficiently small $\varepsilon > 0$, must have top fan-in* $\exp(\Omega(\sqrt{n}\log n))$. □

[KSS13] used the above lower bound to give an $n^{\Omega(\log n)}$ lower bound for a subclass of formulas called *regular formulas*. The interested reader can refer to [KSS13] for more details.

9.3.3. Shifted partials of the Iterated-matrix-multiplication polynomial. Fourier, Limaye, Malod and Srinivasan [FLMS13] showed the same lower bound as [KSS13] but for the *iterated matrix multiplication* polynomial which is known to have polynomial sized circuits computing it.

Definition 58 (Iterated matrix multiplication polynomial). *Let $M_1, \ldots, M_d$ be $n \times n$ matrices with distinct variables as entries, i.e. $M_k = \left(\left(x_{ij}^{(k)}\right)\right)_{i,j \leq n}$ for $k = 1, \ldots, d$. The polynomial $\mathrm{IMM}_{n,d}$ is a $(n^2 d)$-variate degree-d polynomial defined as the $(1,1)$-th entry of the matrix product $M_1 \ldots M_d$:*

$$\mathrm{IMM}_{n,d}(\mathbf{x}) \quad = \quad (M_1 \ldots M_d)_{1,1}$$

A more useful perspective is to interpret this as a *canonical algebraic branching program.*

Definition 59 (Algebraic branching program). *An algebraic branching program (ABP) comprises of a layered directed graph G with $(d+1)$ layers of vertices, where the first and last layer consists of a single node (called source and sink respectively), all other layers consist of n vertices, and edges are only between successive layers and have linear polynomials as edge-weights. The ABP is set to compute the polynomial f defined as*

$$f(\mathbf{x}) \quad = \quad \sum_{\text{source-sink path } \rho} \mathrm{weight}(\rho)$$

where the weight *of any path is just the product of the edge weights on the path.*

The canonical ABP comprises of the graph where the i-th vertex of layer $(\ell - 1)$ is connected to the j-th vertex of layer ℓ with edge-weight $x_{ij}^{(\ell)}$ for every choice of i, j and ℓ. It is easy to see that the polynomial computed by the canonical ABP is in fact $\mathrm{IMM}_{n,d}$.

To lower bound the dimension of shifted partial derivatives of $\mathrm{IMM}_{n,d}$, firstly note that a derivative with respect to any variable (or edge) simply results in the sum of all source-sink paths that *pass* through this edge. [FLMS13] use the following simple but crucial observation to assist in bounding the dimension of shifted partials.

Observation 60. *Assume that d is ~~odd~~even. Let $e_1, e_3, \ldots, e_{d-1}$ be an arbitrary set of edges such that e_i is between layer i and $i+1$. Then, there is a unique path from source to sink that passes through all these edges.*

Proof. Since these are edges in alternate layers, their starting and ending points uniquely determine the edges that are picked up from the even-numbered layers to complete the source-sink path. □

Since we are interested in k-th order derivatives for $k \approx \varepsilon\sqrt{d}$, [FLMS13] consider the following restriction by removing some edges from the underlying graph:

- Select $(2k-1)$ layers $\ell_1, \ldots, \ell_{2k-1}$ that are roughly equally spaced between ~~first and~~ the first and the last layer. These layers, and the first and the last layers, shall be untouched and shall be called *pristine layers.*
- In all the other layers, retain only those edges connecting vertex i of this layer to vertex i of the next.

This restriction effectively makes the graph similar to an ABP with $2k+1$ layers. Let the polynomial computed by the restricted ABP be $\mathrm{IMM}'_{n,d}(\mathbf{x})$. Since $\mathrm{IMM}'_{n,d}$ was obtained by just setting some variables of $\mathrm{IMM}_{n,d}$ to zero, the dimension of shifted partial derivatives of $\mathrm{IMM}'_{n,d}$ can only be smaller than that of $\mathrm{IMM}_{n,d}$. Similar to Observation 60, we have the following observation.

Observation 61. *For every choice of k edges from odd-numbered pristine layers, there is a unique source-sink path that passes through them.*
In other words, for any choice of k variables chosen by picking one from each odd-numbered pristine layer, then the k-th order partial derivative of $\mathrm{IMM}'_{n,d}$ with respect to these k variables is a non-zero monomial.

Once again, we can lower bound the dimension of shifted partial derivatives of $\mathrm{IMM}'_{n,d}$ by a monomial counting problem. Similar to the earlier case, [FLMS13] show that the monomials thus obtained are *far* from one another. We state their main lemma below without proof.

Lemma 62 ([FLMS13])**.** *There are at least $n^{k/2}$ monomials of $\mathrm{IMM}'_{n,d}$ of pairwise distance at least $\frac{n}{4}$.*

Again, using Lemma 55 and standard binomial coefficient estimates, this implies that the shifted partial derivatives of $\mathrm{IMM}'_{n,d}$ is almost as large as the trivial upper bound.

Theorem 63 ([FLMS13])**.** *Let $k = \varepsilon\sqrt{d}$ for a sufficiently small $\varepsilon > 0$ and ℓ be an integer such that $n^{1/16} \leq \frac{N+\ell}{\ell} \leq n^{1/4}$ where N is the number of variables $\mathrm{IMM}'_{n,d}$ depends on. Then,*

$$\begin{aligned}\dim\left(\left\langle \partial^{=k}\left(\mathrm{IMM}_{n,d}\right)\right\rangle_{\leq \ell}\right) &\geq \dim\left(\left\langle \partial^{=k}\left(\mathrm{IMM}'_{n,d}\right)\right\rangle_{\leq \ell}\right) \\ &= \Omega\left(n^{k/2}\cdot\binom{N+\ell}{\ell}\right).\end{aligned}$$

□

Combining with Corollary 46, we get the lower bound of [FLMS13].

Theorem 64 ([FLMS13])**.** *Any $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit computing $\mathrm{IMM}_{n,d}$, with $d \leq n^{\delta}$ for a sufficiently small $\delta > 0$, has top fan-in $\exp(\Omega(\sqrt{d}\log n))$.* □

Similar to [KSS13], the above result also implies $n^{\Omega(\log n)}$ lower bounds for regular formulas computing $\mathrm{IMM}_{n,d}$.

10. Conclusion

The field of arithmetic complexity, like Boolean complexity, abounds with open problems and proving lower bounds for almost any natural subclass of arithmetic circuits is interesting especially if the currently known techniques/

complexity measures do not apply to that subclass[8]. The surveys [Wig02, SY10, CKW11] mark out the frontiers of this area in the form of many open problems and we invite the reader to try some of them~~!~~.

References

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

[ASSS12] Manindra Agrawal, Chandan Saha, Ramprasad Saptharishi, and Nitin Saxena. Jacobian hits circuits: hitting-sets, lower bounds for depth-d occur-k formulas & depth-3 transcendence degree-k circuits. In *Symposium on Theory of Computing (STOC)*, pages 599–614, 2012.

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

[BS83] Walter Baur and Volker Strassen. The complexity of partial derivatives. *Theoretical Computer Science*, 22:317–330, 1983.

[CKW11] Xi Chen, Neeraj Kayal, and Avi Wigderson. Partial Derivatives in Arithmetic Complexity (and beyond). *Foundation and Trends in Theoretical Computer Science*, 2011.

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

[CM14] Suryajith Chillara and Partha Mukhopadhyay. Depth-4 Lower Bounds, Determinantal Complexity : A Unified Approach. *Symposium on Theoretical Aspects of Computing (STACS)*, 2014.

[FLMS13] Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:100, 2013.

[GK98] Dima Grigoriev and Marek Karpinski. An exponential lower bound for depth 3 arithmetic circuits. In *Symposium on Theory of Computing (STOC)*, pages 577–582, 1998.

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

[GR00] Dima Grigoriev and Alexander A. Razborov. Exponential lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. *Appl. Algebra Eng. Commun. Comput.*, 10(6):465–487, 2000.

[HY11] Pavel Hrubeš and Amir Yehudayoff. Arithmetic complexity in ring extensions. *Theory of Computing*, 7(8):119–129, 2011.

[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.

[8] Some of the complexity measures that we describe here yield lower bounds for slightly more general subclasses of circuits.

[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. 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.

[Kou08] Ioannis Koutis. Faster algebraic algorithms for path and packing problems. In *ICALP*, pages 575–586, 2008.

[KSS13] Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A superpolynomial lower bound for regular arithmetic formulas. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:91, 2013.

[Lov11] Shachar Lovett. Computing polynomials with few multiplications. *Theory of Computing*, 7(13):185–188, 2011.

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

[NW94] Noam Nisan and Avi Wigderson. Hardness vs Randomness. *Journal of Computer and System Sciences*, 49(2):149–167, 1994.

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

[Raz06] Ran Raz. Separation of multilinear circuit and formula size. *Theory of Computing*, 2(1):121–135, 2006.

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

[Raz10] Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. In *Symposium on Theory of Computing (STOC)*, pages 659–666, 2010.

[RSY08] Ran Raz, Amir Shpilka, and Amir Yehudayoff. A lower bound for the size of syntactically multilinear arithmetic circuits. *SIAM Journal on Computing*, 38(4):1624–1647, 2008.

[RY09] Ran Raz and Amir Yehudayoff. Lower bounds and separations for constant depth multilinear circuits. *Computational Complexity*, 18(2):171–207, 2009.

[Sri13] Srikanth Srinivasan. personal communication, 2013.

[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.

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

[VSBR83] Leslie G. Valiant, Sven Skyum, S. Berkowitz, and Charles Rackoff. Fast Parallel Computation of Polynomials Using Few Processors. *SIAM Journal of Computing*, 12(4):641–644, 1983.

[Wig02] Avi Wigderson. Arithmetic complexity - a survey. Lecture Notes, 2002.

Neeraj Kayal
Microsoft Research
Bangalore, India
e-mail: neeraka@microsoft.com

Ramprasad Saptharishi
Microsoft Research
Bangalore, India
e-mail: ramprasad@cmi.ac.in