A survey of lower bonds on arithmetic circuit

Ramprasad Saptharishi
Microsoft Research India
ramprasad@cmi.ac.in

September 30, 2014

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus.

Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Part I

Survey 1

1 Introduction

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

1.1 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 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 [HY11a] 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.

2 Classical lower bounds for 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 thus far. In this section, we shall look at the two modest lower bounds for general circuits and formulas.

2.1 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 $\Omega(n \log d)$.*

2.1.1 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.

2.1.2 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

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

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)

2.2 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 $f_i \in F[X \setminus Y]$. 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*.

2.2.1 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 $f_v \in \mathbb{F}[X \setminus Y]$, 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 $f_v^{(0)}, f_v^{(1)} \in \mathbb{F}[X \setminus Y]$. 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 $f_1, f_2 \in \mathbb{F}[X \setminus Y]$.*

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 $\times$ 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 $f_u^{(0)}, f_u^{(1)} \in \mathbb{F}[X \setminus Y]$ 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 $f_i \in \mathbb{F}[X \setminus Y]$ 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 $2\,|\textsc{Leaf}_Y(\Phi)|$ augmented nodes, we have that $\mathrm{td}_Y(f) \leq |S_Y| \leq 4\,|\textsc{Leaf}_Y(\Phi)|$. Therefore,

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

Hence,

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

□

2.2.2 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 $\mathsf{td}_{X_i}(\mathsf{Det}_n)$ is the same for all i. Hence, it suffices to show that $\mathsf{td}_Y(\mathsf{Det}_n) = \Omega(n^2)$ 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, $\mathsf{td}_Y(\mathsf{Det}_n) \geq \mathrm{trdeg}\,\{x_{ij} x_{ji} \;:\; i \neq j\} = \Omega(n^2)$. Therefore, $\Gamma^{[\mathrm{Kal}]}(\mathsf{Det}_n) = \Omega(n^3)$. □

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

3
“Natural” proof strategies

The lower bounds presented in Chapter 2 proceeded by first identifying a *weakness* of the model, and exploiting it in an explicit manner. More concretely, Section 2.2 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 $\Gamma : \mathbb{F}[x_1, \ldots, x_n] \to \mathbb{Z}_{\geq 0}$ 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.

3.1 Some simple lower bounds

Let us start with the simplest complete[1] class of arithmetic circuits – depth-2 circuits or $\Sigma\Pi$ circuits.

3.1.1 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.
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 $\Gamma(f) = n^{\Omega(n)}$ giving the lower bound.

3.1.2 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.[2]

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

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

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 $2^{\Omega(n)}$.

3.1.3 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 $\dim\{\ell_{i1},\ldots,\ell_{id}\} \leq r$.

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 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*

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

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.

4 Lower bounds for monotone circuits

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

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 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,

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

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

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

Since $\Gamma(\mathsf{Det}_n) = n!$, this forces $s \geq \binom{n}{n/3} = 2^{\Omega(n)}$. □

We only need to prove Lemma 17 now.

4.1 Proof of Lemma 17

Without loss of generality, assume that the circuit Φ is homogeneous[1], 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.*

[1] 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 [SY10a].

Let $\textsc{ProofTrees}(\Phi)$ 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,

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

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 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. □

5 Lower bounds for depth-3 circuits over finite fields

This chapter 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 $\mathrm{rank}(T_i)$ as in Section 3.1.3 to be the dimension of the set of linear polynomials that T_i is a product of.
In Section 3.1.3, 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 3.1.3 with evaluations to show that Det_n cannot be approximated by a low-rank $\Sigma\Pi\Sigma$ circuit.

5.1 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 *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 $\mathrm{rank}(M_k(f;\mathcal{A}))$.

5.2 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 *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 $\mathcal{A} = \mathbb{F}^{n^2}\setminus\mathcal{E}$ where $\mathcal{E} = \bigcup_{T \text{ of large rank}} \mathcal{E}_T$, 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

$$\Gamma^{[\mathrm{GK}]}_{k,\mathcal{A}}(C) \quad \leq \quad s\cdot q^{\beta n}. \tag{5.1}$$

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 $0<\alpha<1$.

5.3 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 [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,

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

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,

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

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,

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

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

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

□

5.4 Putting it all together

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

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

where $H(\gamma)$ is the binary entropy function[1]. By choosing $\gamma < q^{-q/2}$, we can find a β such that $q\gamma < \beta$ (which was required in Section 5.2) and $2H(\gamma) > \beta \log q$, yielding the lower bound

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

□(Theorem 20)

[1] 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 $\binom{n}{k} \approx 2^{nH(k/n)}$.

6

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.

6.1 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 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 straight-forward 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* $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f)$ *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* $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f+g) \quad \leq \quad \Gamma^{[\mathrm{Raz}]}_{Y,Z}(f) \quad + \quad \Gamma^{[\mathrm{Raz}]}_{Y,Z}(g)$.

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*

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

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

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 $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f)$ defined above. Unfortunately, there are examples of simple polynomials like $f = (y_1 + z_1)\dots(y_n + z_n)$ with $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f) = 2^n$. Raz's idea here was to look at $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f)$ 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*

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

It is to be noted that we should expect a random polynomial to be full-rank with respect to any partition, so the measure $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f)$ 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 $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f)$ with respect to any partition. Let $|X| = n$.
Using Observation 26, $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f) \leq 2^d$ 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.

6.2 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 [SY10a].

6.2.1 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. □

6.2.2 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

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

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 $i \leq t' \overset{\text{def}}{=} \frac{\log |X|}{100}$. 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:

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

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 $t \leq g$,*

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

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.

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

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 $|Y| = |Z| = |X|/2$.

6.2.3 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 $Y = \{x_{1,1}, x_{3,3}, \dots, x_{2m-1,2m-1}\}$ and $Z = \{x_{2,2}, x_{4,4}, \dots, x_{2m,2m}\}$. For ease of notation, we shall refer to $x_{2i-1,2i-1}$ as y_i and $x_{2i,2i}$ as z_i for $i = 1, \dots, m$.
Consider the following restriction:

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

It is easy to check that $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f) = 2^m$. Although this is a single restriction with large rank, the Schwartz-Zippel-DeMillo-Lipton lemma immediately gives that random restriction would also have rank 2^m with high probability[1]. We shall record this as a lemma.

Lemma 34. *With probability at least $1/100$, we have that $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(\sigma(\mathsf{Det}_n)) = 2^m$ 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 $n^{\Omega(\log n)}$.* □

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

6.3 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 $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f)$ 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 $\Gamma^{[\mathrm{Raz}]}_{Y,Z}(f) = 2^{|X|/2}$ 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 $g_1(X_1)\dots g_\ell(X_\ell)$ 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 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 $f = g_1 \dots g_t$ 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 $\left(\frac{n}{100}\right)^{1/\Delta} = t^2$. 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*

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

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, $\Gamma^{[\text{Raz}]}_{Y,Z}(g_0) \leq 2^{n/10}$.

For the other g_i's, we can bound the probability that $\Gamma^{[\text{Raz}]}_{Y,Z}(g_i)$ 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 $(1/\text{poly}(n))^t$. Hence, with high probability

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

□

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* $\exp(n^{\Omega(1/\Delta)})$. □

7 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*

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

7.1 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 [GKKS13a, KSS13, FLMS13] that comes very close to the required threshold.

7.2 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 $\Gamma_{k,\ell}^{[\mathrm{Kay}]}$ 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,*

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

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

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

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.

7.3 Lower bounding shifted partials of explicit polynomials

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

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

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 $\left\langle \partial^{=k}(f)\right\rangle_{\leq \ell}$, 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.

7.3.1 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 [GKKS13a] 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 $\left\langle \partial^{=k}(\mathsf{Det}_n)\right\rangle_{\leq \ell} = \operatorname{span}\left\{\mathbf{x}^{\leq \ell}\partial^{=k}\mathsf{Det}_n\right\}$. [GKKS13a] 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 $x_{i_1 j_1} \ldots x_{i_{(n-k)}, j_{(n-k)}}$ 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*

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

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, [GKKS13a] assign a *canonical increasing sequence* to every monomial that contains an increasing sequence and multiply by monomials of degree ℓ that do not change the canonical increasing sequence.

Definition 49. *Let $D_2 = \{x_{1,1}, \ldots, x_{n,n}, x_{1,2}, x_{2,3}, \ldots, x_{n-1,n}\}$, 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 [GKKS13a]

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, [GKKS13a] 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.

7.3.2 Shifted partials of the Nisan-Wigderson polynomial

Very shortly after [GKKS13a]'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:*

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

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_i and the variables in 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.

7.3.3 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 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 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 $\mathrm{IMM}_{n,d}$.

Part II

Survey 2

7.4 Introduction

"What is the best way to compute a given polynomial $f(x_1, \ldots, x_n)$ from basic operations such as $+$ and $\times$?" This is the main motivating problem in the field of arithmetic circuit complexity. The notion of *complexity* of a polynomial is measured via the size of the smallest arithmetic circuit computing it. Arithmetic circuits provide a robust model of computation for polynomials. Formally, these are directed acyclic graphs with a unique sink vertex, where internal nodes are labelled by $+$ and $\times$ and each source node labelled with either a variable or a field constant. Each $+$ gate computes the sum of the polynomials computed at its children, and $\times$ gates the product. The unique sink vertex is called the root or the output gate, and the polynomial computed by that gate is the polynomial computed by the circuit.
There are several interesting questions that can be asked about arithmetic circuits, and polynomials that they compute. One category of problems are of the form, "Is there an explicit polynomial $f(x_1, \ldots, x_n)$ that require (perhaps restricted) arithmetic circuits of size $2^{\Omega(n)}$ to compute them?", or questions about proving lower bounds. Another category of problems are of the form, "Is the given circuit computing the 0 polynomial?", which is also called 'Polynomial Identity Testing (PIT)'. Yet another class of questions are of the form "Given oracle access to a circuit, can you write down the polynomial computed by this circuit?", which are also called 'polynomial reconstruction'. Several of these problems have very strong connections between each other despite being of very different flavours. Formal connections between PIT and lower bounds have been shown by [KI04, Agr05]. Further, strong lower bounds for restricted models have often been succeeded by reconstruction algorithms (at least on average). In this article we shall mainly be looking at lower bounds. For more on reconstruction and PIT questions, the author is invited to read other excellent surveys such as [SY10b, CKW11].

7.4.1 Arithmetic complexity classes

In the seminal paper of [Val79], Valiant defined two classes of polynomials which we now call VP and VNP.

Definition 65. *The class* VP *is defined as the set of all polynomial* $f(x_1, \ldots, x_n)$ *with* $\deg(f) = n^{O(1)}$ *that can be computed by an arithmetic circuit of size* $s = n^{O(1)}$.
The class VNP *is defined as the set of all polynomial* $f(x_1, \ldots, x_n)$ *such that there exists a* $g(x_1, \ldots, x_n, y_1, \ldots, y_m)$ *with* $m = n^{O(1)}$ *such that*

$$f(x_1, \ldots, x_n) \quad = \quad \sum_{y_1=0}^{1} \cdots \sum_{y_m=0}^{1} g(x_1, \ldots, x_n, y_1, \ldots, y_m)$$

The class VP is synonymous to what we understand as *efficiently computable* polynomials. The class VNP, whose definition is similar to the boolean class NP, is in some sense a notion of what deem as *explicit*.

Fact 1. *Let $f(x_1, \ldots, x_n)$ be a polynomial such that $\deg(f) = n^{O(1)}$ and given any exponent vector $e_1, \ldots, e_n$, the coefficient of the monomial $x_1^{e_1} \ldots x_n^{e_n}$ in f can be computed in polynomial time. Then, $f \in \mathsf{VNP}$.*

For example, consider the permanent of a symbolic $n \times n$ matrix. In fact, [Val79] showed that the symbolic $n \times n$ permanent is in some sense complete for the class VNP. Further, he also showed that the determinant of a symbolic $n \times n$ matrix is (almost) complete for the class VP. Separating the determinant and the permanent is the Holy Grail in the field of arithmetic circuit complexity.

Remark. Note that the above fact merely gives a sufficient condition for a polynomial to be in VNP. There are examples of polynomials f where computing the coefficient of a given monomial is believed to be very hard but $f \in \mathsf{VNP}$.[1] In this article however, all the polynomials we shall be dealing with would have this property that the coefficient of a given monomial can be efficiently computed. For more about completeness classes in arithmetic complexity, [BCS97] is a wonderful text.

7.4.2 Prior lower bounds

Proving lower bounds is generally considered challenging, in most models of computation. For general circuits, the best lower bound we have for an explicit polynomial is by [BS83] who prove an $\Omega(n \log n)$ lower bound. For the subclass of arithmetic formulas, [Kal85] has shown a $\Omega(n^{3/2})$ lower bound. On the other hand, we know by standard counting methods that most n-variate degree d polynomials require circuits of size $\Omega\left(\sqrt{\binom{n+d}{d}}\right)$.

To gain better understanding of computation by arithmetic circuits, researchers focused on proving lower bounds for restricted models of computation. One very natural restriction is the depth of the circuit. Proving lower bounds for depth two circuits are trivial. For general depth three circuits, the best lower bound we have is by [SW01] who present an $\Omega(n^2)$ lower bound. Exponential lower bounds are known with additional restrictions like *homogeneity* [NW97], *multilinearity* [Raz09, RY09], over finite fields [GR00, GK98], *monotonicity* [JS82] etc.

For multilinear models, more is known for even larger depth. [Raz09] showed an $n^{\Omega(\log n)}$ lower bound for the class of multilinear formulas. [RY09] extended those techniques to show an $2^{n^{\Omega(1/\Delta)}}$ lower bound for multilinear formulas of depth Δ.

[1] For example, consider the n^2 variate multilinear polynomial f such that the coefficient $x_{11}^{e_{11}} \ldots x_{nn}^{e_{nn}}$ is the permanent of the $n \times n$ matrix $((e_{ij}))_{i,j}$. Turns out $f \in \mathsf{VNP}$. In fact, a necessary and sufficient condition is that the coefficient of a given monomial can be computed in $\#\mathsf{P}/\mathsf{poly}$.

7.4.3 Relevance of shallow circuits for "VP vs VNP"

The study of lower bounds for shallow circuits is not just an attempt to simplify the problem and gain insight on the larger goal. The class of shallow arithmetic circuits are surprisingly powerful, unlike the boolean case. Shallow circuits in the arithmetic world almost capture the entire computational power of unrestricted circuits!
There has been a long series of results that simulate a general arithmetic circuit C by a *shallow* circuit of size comparable to the size of C. This task simulating a circuit but another not-too-large circuit of small depth is called *depth reduction*. The first result in this regard is by [VSBR83] who proved the following.

Theorem 66 ([VSBR83]). *Let f be an n-variate degree d polynomial computed by an arithmetic circuit C of size s. Then, f can be equivalently computed by a homogeneous circuit C' of depth $O(\log d)$ with unbounded fan-in $+$ and $\times$ gates and size $s' = (nds)^{O(1)}$.*

In fact, the resulting circuit C' has the following useful structure.

- The circuit is made up of alternating layers consisting of $+$ and $\times$ gates.
- All multiplication gates have fan-in at most 5.
- If g is the polynomial computed at a multiplication gate, and g' is the polynomial computed at one of its children, then $\deg(g') \leq \deg(g)/2$.

The above theorem allows us to focus on just homogeneous circuits of $O(\log d)$ depth and attempt lower bounds for this model. Any super-polynomial lower bound for the class of $O(\log d)$ depth circuits automatically yields a super-polynomial lower bound for general circuits.

However, if we really hope to prove much stronger lower bounds for Perm_n like say $2^{\Omega(n)}$, maybe we can afford to incur a slightly larger blow-up in size to obtain an even shallower circuit. This line was first pursued by [AV08], and subsequently strengthened by [Koi12] and [Tav13] to yield the following result.

Theorem 67 ([AV08, Koi12, Tav13]). *Let f be an n-variate degree d polynomial computed by an arithmetic circuit of size s. Then f can be computed by a homogeneous $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit of size $s' \leq s^{O(\sqrt{d})}$*
More generally, for any $0 \leq r \leq d$, there is a homogeneous $\Sigma\Pi^{[O(d/r)]}\Sigma\Pi^{[r]}$ circuit of top fan-in at most $s^{O(d/t)}$ computing f.

Recall that a $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit computes a polynomial of the form

$$f \quad = \quad \sum_{i=1}^{s} Q_{i1} \dots Q_{ia} \quad , \quad \text{where } a = O(\sqrt{d}) \text{ and } \deg Q_{ij} \leq \sqrt{d}$$

In other words, if we can prove a lower bound of $n^{\omega(\sqrt{d})}$ for the class of $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuits, we would have a super-polynomial lower bound for the class of general arithmetic circuits! In fact, the model of depth 4 circuits seem so central in that almost all known lower bounds for other restricted models proceed by proving a suitable lower bound for a depth 4 analogue. Several examples of this may be seen in [KS14b].

The first breakthrough was obtained by [GKKS13a] who showed an $2^{\Omega(\sqrt{d})}$ lower bound for such circuits computing the symbolic $d \times d$ determinant or permanent. Subsequently, there was a flurry of activity[2] towards achieving the goal of proving $n^{\omega(\sqrt{d})}$ lower bounds [KSS13, FLMS13, KS14c], and this is where we currently stand.

Theorem 68. *There is an explicit homogeneous n-variate degree d polynomial f that can be computed by a homogeneous depth 4 circuit of size $n^{O(1)}$ but any $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ computing it requires top fanin $s = n^{\Omega(\sqrt{d})}$.*

If we could change the $n^{\Omega(\sqrt{d})}$ to $n^{\omega(\sqrt{d})}$ in the above theorem (of course, the polynomial f cannot then have a small arithmetic circuit computing it), we would have proved a super-polynomial lower bound for general arithmetic circuits! The following is the simplest formulation of a lower bound of shallow circuit that would imply lower bounds for general circuits.

Open Problem 1. *Find an explicit n-variate degree d polynomial f such that any expression of the form*

$$f \quad = \quad (Q_1)^{\sqrt{d}} + \cdots + (Q_s)^{\sqrt{d}} \quad , \quad \deg(Q_i) \leq \sqrt{d} \text{ for all } i$$

must have $s = n^{\omega(\sqrt{d})}$.

Subsequent to this line of work, several researchers addressed the task of proving lower bounds for homogeneous depth 4 circuits without any restriction on the fan-ins. It is worth noting that a lower bound for homogeneous depth 4 circuits must be on the total size and not the top fan-in, as otherwise one could just compute the polynomial f in a single gate of the bottom two layers.

[2] so much this is the second survey on arithmetic circuit lower bounds that the author is involved in within a year!

Why another survey?

> So why are super polynomial lower bounds still not proved? Maybe it's because not enough people are working on it.
>
> – Ran Raz (in [Raz10])

We strongly believe that the above statement really hits the nail on the head. Fortunately, over the last few years we have seen such a phenomenal activity in arithmetic circuit lower bounds and an increased optimism that we can indeed soon separate VP and VNP. The open problem stated above is simple enough (to state!) that any one can start thinking about it. Further, we already have an $n^{\Omega(\sqrt{d})}$ lower bound, and we only need to make that $n^{\omega(\sqrt{d})}$. We believe that separating VP and VNP would be solved in the not-so-distant future and the hope is that the recent surveys would assist people familiarize with the known lower bounds and develop the necessary tools. As a student, the surveys of [SY10b, CKW11] were immensely helpful and this is an attempt to give back to the community.

Recently, with Neeraj Kayal [KS14b], we presented a comprehensive exposition of almost all known lower bounds known until then with nearly complete proofs. We tried to present all of them from a single perspective of constructing *complexity measures* for appropriate depth 4 analogues. Subsequently, there has been fresh lower bounds which, although are modifications of the earlier measures, are much more delicate to analyze and employ several new ideas to assist in the calculations. The goal of this survey is to complement [KS14b] and present the key intuitions in the newer lower bounds for restricted arithmetic circuits. This article would not have complete proofs of the newer lower bounds but would hopefully present the main subtleties involved to help the interested reader to work through the full proofs themselves.

Organization

We first begin with some preliminaries and notation that would be required in Section 7.5. We then move on to present the depth reduction to depth 4 circuits to put the lower bounds in context. We then outline the general road map followed by almost all lower bound proofs in Section 7.7 and then proceed to the lower bounds of [KLSS14] and [KS14d] in Section 7.8. In Section 7.9, we focus on non-homogeneous depth 3 circuits and present the depth reduction of [GKKS13b] and the recent lower bound of [KS14a] for depth three circuits with small bottom fan-in. We look at some speculative approaches towards proving superpolynomial lower bounds for homogeneous formulas in Section 7.10 before concluding in Section 7.11.

7.5 Notation and preliminaries

7.5.1 Subclasses of circuits

We shall be considering various subclass of constant depth circuits in this article and it would be useful to fix some notation for the parameters involved.

- A $\Sigma\Pi^{[a]}\Sigma\Pi^{[b]}$ circuit computes a polynomial of the form

$$f \quad = \quad \sum_i Q_{i1} \dots Q_{ia} \quad \text{where} \quad \deg Q_{ij} \leq b$$

- A $\wedge$ refers to a layer of exponentiation gates. For example, a $\Sigma\wedge\Sigma$ circuit computes a polynomial of the form

$$f \quad = \quad \ell_1^{d_1} + \cdots + \ell_s^{d_s}$$

  where each ℓ_i is a linear polynomial.

- In general, we shall add super script such as $\Sigma^{[a]}$ or $\Pi^{[a]}$ to denote a bound on the fan-in of gates in that layer. For example, $\Sigma\Pi\Sigma^{[a]}$ would refer to depth three circuits where every linear polynomial depends on at most a variables.

Throughout the article, we would be dealing mainly with multilinear polynomials with zero-one coefficients. Thus, it would be useful to identify such any monomial of such a polynomial $P(x_1, \dots, x_n)$ by the set of variables that divide it. This shall allow us to say "$m_1 \cap m_2$" instead of $\gcd(m_1, m_2)$. Further, we shall abuse notation and say "$m \in P(\mathbf{x})$" to mean that the monomial m has a non-zero coefficient in $P(\mathbf{x})$.

7.5.2 Some useful estimates

We shall be seeing a lot of binomial coefficients and the following lemma would be useful to have a handle on how large they are.

Lemma 69. *Let n and ℓ be parameters such that $\ell = \frac{n}{2}(1-\varepsilon)$ for some $\varepsilon = o(1)$. For any a, b such that $a, b = O(\sqrt{n})$,*

$$\binom{n-a}{\ell-b} \quad = \quad \binom{n}{\ell} \cdot 2^{-a} \cdot (1+\varepsilon)^{a-2b} \cdot \mathsf{poly}(n)$$

Proof. The proof of the above lemma would repeated use [GKKS13a, Lemma 6] that

$$(n+a)! \quad = \quad n! \cdot n^a \cdot \mathsf{poly}(n)$$

for any $a = O(\sqrt{n})$.

Hence,

$$\begin{aligned}
\frac{\binom{n-a}{\ell-b}}{\binom{n}{\ell}} &= \frac{(n-a)!}{n!}\cdot\frac{\ell!}{(\ell-b)!}\cdot\frac{(n-\ell)!}{(n-\ell-a+b)!}\\
&\overset{\text{poly}}{\approx} \frac{1}{n^a}\cdot\ell^b\cdot\frac{(n-\ell)^a}{(n-\ell)^b}\\
&= \frac{\left(\frac{n}{2}\right)^a(1+\varepsilon)^a}{n^a}\cdot\frac{(1-\varepsilon)^b}{(1+\varepsilon)^b}\\
&\overset{\text{poly}}{\approx} 2^{-a}\cdot(1+\varepsilon)^{a-2b}
\end{aligned}$$

□

7.5.3 Polynomials of interest

There are a few polynomials that are the usual suspects while proving lower bounds. The polynomials that we would be dealing with in this article are defined below.

The determinant and the permanent families

The determinant of an $n \times n$ symbolic matrix shall be denoted by Det_n and is defined as

$$\mathsf{Det}_n = \sum_{\sigma\in S_n} \text{sign}(\sigma)x_{1,\sigma(1)}\dots x_{n,\sigma(n)}$$

The permanent of an $n \times n$ symbolic matrix shall be denoted by Perm_n and is defined as

$$\mathsf{Perm}_n = \sum_{\sigma\in S_n} x_{1,\sigma(1)}\dots x_{n,\sigma(n)}$$

Both of these polynomials are of degree n and over n^2 variables. We know that Det_n can be computed by a polynomial sized arithmetic circuit and it is widely believed that the permanent requires circuits of size $2^{\Omega(n)}$.

The Nisan-Wigderson polynomial families

Let n, m, d be arbitrary parameters with m being a power of a prime, and $n, d \leq m$. Since m is a power of a prime, let us identify the set $[m]$ with the field $\mathbb{F}_m$ of m elements. Note that since $n \leq m$, we have that $[n] \subseteq \mathbb{F}_m$. The Nisan-Wigderson polynomial with parameters n, m, d, denoted by $\mathrm{NW}_{n,m,d}$ is defined as

$$\mathrm{NW}_{n,m,d}(\mathbf{x}) = \sum_{\substack{p(t)\in\mathbb{F}_m[t]\\ \deg(p)\leq d}} x_{1,p(1)}\dots x_{n,p(n)}$$

That is, for every univariate polynomial $p(t) \in \mathbb{F}_m[t]$ of degree at most d, we add one monomials that encodes the 'graph' of p on the points $[n]$. This is a polynomial of degree n over mn variables.

This monomials of this polynomial satisfy a very useful low-pairwise-intersection property.

Lemma 70. *Let m_1 and m_2 be any two distinct monomials in $\mathrm{NW}_{n,m,d}(\mathbf{x})$. Then, there are at most d variables that divide both m_1 and m_2.*

Proof. Let m_1 and m_2 correspond to univariates $p_1(t), p_2(t) \in \mathbb{F}_m[t]$ of degree at most d. Then if x_{ij} divides m_1, then $p_1(i) = j$, similarly for m_2. But since p_1 and p_2 are two distinct polynomials of degree at most d, they can agree in at most d evaluations. Thus, there can be at most d variables that divide both m_1 and m_2. □

For most generic choices of the parameters, the polynomial $\mathrm{NW}_{n,m,d}$ is believed to require circuits of exponential size to compute them.

The Iterated-Matrix-Multiplication polynomial

For parameters n and d, the Iterated-Matrix-Multiplication polynomial, denoted by $\mathrm{IMM}_{n,d}$, is defined as follows

$$\mathrm{IMM}_{n,d} \quad = \quad \sum_{1 \leq i_1, \ldots, i_d \leq n} x^{(1)}_{1,i_1} x^{(2)}_{i_1,i_2} \ldots x^{(d-1)}_{i_{d-2},i_{d-1}} x^{(d)}_{i_{d-1},1}.$$

An equivalent way of defining the polynomial as the $(1,1)$-th entry of the product of d generic $n \times n$ matrices:

$$\mathrm{IMM}_{n,d} \quad = \quad \left(\begin{bmatrix} x^{(1)}_{11} & \ldots & x^{(1)}_{1n} \\ \vdots & \ddots & \vdots \\ x^{(1)}_{n1} & \ldots & x^{(1)}_{nn} \end{bmatrix} \cdots \begin{bmatrix} x^{(d)}_{11} & \ldots & x^{(d)}_{1n} \\ \vdots & \ddots & \vdots \\ x^{(d)}_{n1} & \ldots & x^{(d)}_{nn} \end{bmatrix} \right)_{(1,1)}.$$

It is often useful to think of this as the polynomial computed by a *generic algebraic branching program* of width n and depth n (where the edge connecting vertex i of layer ℓ to vertex j of layer $\ell + 1$ has weight $x^{(\ell)}_{ij}$).

This is a polynomial of degree d and over $n^2(d-2) + 2n$ variables. Further, since the polynomial corresponds to a generic algebraic branching program, $\mathrm{IMM}_{n,d}$ can be computed by an arithmetic circuit of size $\mathsf{poly}(n, d)$.

7.6 A primer on depth reduction to depth 4 circuits

In this section, we shall look at depth reduction for arithmetic circuits. As mentioned earlier, the starting point of all known depth reductions is the result of [VSBR83].

Theorem 66 (restated). *Let f be an n-variate degree d polynomial computed by an arithmetic circuit C of size s. Then, f can be equivalently computed by a homogeneous circuit C' of depth $O(\log d)$ with unbounded fan-in $+$ and $\times$ gates and size $s' = (nds)^{O(1)}$.*
Further, the circuit C' has the following structure:

- *The circuit is made up of alternating layers consisting of $+$ and $\times$ gates.*
- *All multiplication gates have fan-in at most 5.*
- *If g is the polynomial computed at a multiplication gate, and g' is the polynomial computed at one of its children, then $\deg(g') \leq \deg(g)/2$.*

We shall not be proving this theorem here but with this as the starting point, we shall given an alternate proof of the depth reduction by [AV08, Koi12, Tav13]. This alternate proof was obtained jointly with V Vinay.

Theorem 67 (restated). *Let f be an n-variate degree d polynomial computed by an arithmetic circuit of size s. Then f can be computed by a homogeneous $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit of size $s' \leq s^{O(\sqrt{d})}$*

Proof. Start with the circuit C' obtained from Theorem 66 computing f of size $s' = s^{O(1)}$. Let g be the polynomial computed at any arbitrary gate in the circuit. From the structure of C', we have that

$$g = \sum_{i=1}^{s'} g_{i1} \cdot g_{i2} \cdot g_{i3} \cdot g_{i4} \cdot g_{i5} \tag{7.1}$$

with $\deg(g_{i1}) + \cdots + \deg(g_{i5}) \leq \deg(g)$ and $\deg(g_{ij}) \leq \deg(g)/2$ for all i, j.
Key Observation. For each i, there must be at least two j's such that $\deg(g_{ij}) \geq \deg(g)/8$.

Since the above decomposition is true for any gate in the circuit, f, polynomial computed at the root, can be written as

$$f = \sum_{i=1}^{s'} f_{i1} \cdot f_{i2} \cdot f_{i3} \cdot f_{i4} \cdot f_{i5}. \tag{7.2}$$

The RHS is a $\Sigma\Pi^{[5]}\Sigma\Pi^{[d/2]}$ circuit of top fan-in s'. The goal would be to progressively reduce the bottom fan-in at the cost of increasing the top fan-in slightly. Eventually, we want an expression where all f_{ij}'s involved have degree at most $\sqrt{d}$.
We shall follow an extremely natural strategy:

Start with (7.2) for f.

For each summand $f_{i1} \ldots f_{ir}$ in the RHS, if the largest degree f_{ij} has degree more than $\sqrt{d}$, expand that f_{ij} with the its corresponding representation using (7.1).

Repeat this process until all f_{ij}'s on the RHS have degree at most $\sqrt{d}$.

In every round of the above routine, the initial equation (7.2) for f slowly evolves. At the end of each round, the top fan-in increases by a factor of s' but there is some drop in the degree of terms involved. We now need to show that by $O(\sqrt{d})$ rounds, all of the f_{ij}'s involved would have degree bounded by $\sqrt{d}$. This would imply that the top fan-in of that equation is bounded by $s^{O(\sqrt{d})}$ as claimed.
If we take any term $f_{i1} \dots f_{ir}$ with $\deg(f_{i1}) \geq \sqrt{d}$ and expand f_{ij} via (7.1), each term in the expansion of f_{i1} must have at least two factors of degree more than $\sqrt{d}/8$ (by the key observation). Thus, in each term obtained by expanding f_{i1} in $f_{i1} \dots f_{ir}$ must have the number of factors of degree more than $\sqrt{d}/8$ increased by at least one. Since we know that no term can have more than $8\sqrt{d}$ such factors, this must imply that the number of rounds is bounded by $8\sqrt{d}$.
Thus we eventually have an equation of the form

$$f \quad = \quad \sum_{i=1}^{s'^{8\sqrt{d}}} f_{i1} \dots f_{ir} \quad \text{where for each } i, j, \quad \deg(f_{ij}) \leq \sqrt{d}$$

To ensure that $r \leq O(\sqrt{d})$, the standard trick is to take any ensure that $\deg(f_{ij}) \geq \sqrt{d}/2$ by multiplying out factors of degree smaller than $\sqrt{d}/2$. Thus, we have a $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit of top fan-in $s^{O(\sqrt{d})}$ computing f. □

The original proof of Tavenas was not much involved either, but the above proof would be able to offer more insights towards proving homogeneous formula lower bounds. We shall however defer this discussion to Section 7.10.

Surprisingly, it was shown by [GKKS13b] that over characteristic zero fields, any n-variate degree d polynomial f can be computed by a depth 3 circuit of size $n^{O(\sqrt{d})}$. We shall present its proof in Section 7.9 where it would better placed alongside the recent lower bound for depth 3 circuits by [KS14a].

7.7 'Natural' proof strategies

Most lower bounds follow the plan outlined below. There are a few notable exceptions but by and large this is the general strategy followed by almost all known lower bound proofs.

> **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. There are cases when obtaining this representation might itself be non-trivial.

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. $\Gamma(f_1 + f_2) \leq \Gamma(f_1) + \Gamma(f_2)$

This is really the most crucial part of the lower bounds. In most cases, $\Gamma(f)$ is the rank of a large matrix whose entries are linear functions in the coefficients of f. This would be the case in the lower bounds considered in this article as well. 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.

The bulk of all lower bound proofs goes into this step. In several cases, there are natural candidate polynomials for which one can show $\Gamma(f)$ is large. In some cases, it might be easier to *engineer* a polynomial for which it is easier to show that $\Gamma(f)$ is large.

With this general strategy in mind, we can go ahead to see the lower bounds of [KLSS14, KS14d].

7.8 Lower bounds for homogeneous depth four circuits

The model for which we shall be interested in proving lower bounds are homogenous depth four circuits. These circuits compute polynomials of the form

$$f \quad = \quad \sum_i Q_{i1} \ldots Q_{ia_i}$$

where each Q_{ij} is a homogeneous polynomial. This immediately forces that $\sum_{j=1}^{a_i} \deg(Q_{ij}) = \deg(f)$ for all i.

Goal. Find an explicit polynomial f (of degree d, and over n variables) such that any homogeneous depth four circuit requires size $n^{\Omega(\sqrt{d})}$. That is, if

$$f \quad = \quad \sum_i Q_{i1} \ldots Q_{ia_i}$$

for homogeneous polynomials Q_{ij}'s, then the total number of monomials present among the Q_{ij}'s must be $n^{\Omega(\sqrt{d})}$.

Intuition towards the measure - (1)

Consider an expression of the form

$$C = \sum_{i=1}^{s} Q_{i1} \ldots Q_{ia_i}$$

We shall call a summand $Q_{i1} \ldots Q_{ia_i}$ *good* if the degree of each $Q_{ij} \leq \sqrt{d}$. Let us split the above sum into *good* terms and the rest.

$$C_1 = \sum_{i=1}^{s_1} Q_{i1} \ldots Q_{ia_i} \quad \text{where} \quad \deg(Q_{ij}) \leq \sqrt{d} \quad \text{for all } i, j \tag{7.3}$$

$$C_2 = \sum_{i=s_1+1}^{s} Q_{i1} \ldots Q_{ia_i} \quad \text{where} \quad \deg(Q_{i1}) > \sqrt{d} \quad \text{for all } i > s_1 \tag{7.4}$$

If one were to just prove a lower bound for (7.3), then using the dimension of shifted partial derivatives we can obtain a lower bound of $n^{\Omega(\sqrt{d})}$. Hence let us focus on an expression of the form (7.4) and see if we can come up with a measure that gives a $n^{\Omega(\sqrt{d})}$ lower bound there as well.
Starting with (2), let us expand each Q_{i1} as a sum of monomials to obtain an expression of the form

$$C_2 = \sum_{i=1}^{s'} m_i \cdot Q_i'$$

where each m_i is a monomial of degree greater than $\sqrt{d}$, and Q_i' some polynomial of degree $d - \deg(m_i)$. The number of summands s' would be at most the size of the circuit we started out with.

Key Idea: Suppose the polynomial C_2 was multilinear, i.e. the degree in each variable is bounded by 1. Further, say $s' \leq n^{\sqrt{d}/10}$. Apply a random restriction ρ on the variables by setting each variable independently to zero with probability $p < \frac{1}{n^{1/20}}$.
If m was any monomial that was divisible by $\sqrt{d}$ disjoint variables, then $\rho(m) \neq 0$ with probability at most $\frac{1}{n^{\sqrt{d}/20}}$. Hence, the probability that $\rho(m_i) \neq 0$ for some $i \leq s'$ that is divisible by $\sqrt{d}$ variables is at most $\frac{1}{n^{\sqrt{d}/10}}$. Hence, the only terms that would survive on the RHS are terms of the form $\rho(m_i \cdot Q_i')$ where m_i is divisible by at most $\sqrt{d}$ distinct variables. But recall that $\deg(m_i) > \sqrt{d}$ and this implies that m_i is non-multilinear. If that is the case, then every monomial on the RHS is non-multilinear! Thus as long as $\rho(C_2) \neq 0$, there would be at least one multilinear monomial that survives. This would contradict our original assumption that $s' \leq n^{\sqrt{d}/10}$, giving us the lower bound we were after.

Thus, the measure for the sum of *good* terms is the dimension of shifted partial derivatives. The measure for the sum of non-*good* terms was *the number of non-zero multilinear* *monomials after a random restriction*. Hopefully some combination of these measures would give us a measure for their sum.[3]

Intuition towards the measure - (2)

The idea of using random restrictions as defined above essentially kills all monomials that are divisible by 'too many' variables. Let us consider an extreme case where every monomials in each Q_{ij} is just a power of a single variable. We shall first try to prove a lower bound for expression of the form

$$C \quad = \quad \sum_i Q_{i1} \cdots Q_{ia_i}$$

where every monomial in any Q_{ij} is a power of a single variable, i.e. each Q_{ij} is a sum of univariate polynomials.
Define the operator $\mathrm{MultiQuad}$ that acts on any polynomial Q such that $\mathrm{MultiQuad}(Q)$ is just the sum of monomials of Q of degree at most 2 in every variable. Then,

$$\begin{aligned} C &= \sum_i \mathrm{MultiQuad}(Q_{i1}) \cdots \mathrm{MultiQuad}(Q_{ia_i}) \quad + \quad \text{other terms} \\ &= \qquad C_1 \quad + \quad C_2 \end{aligned}$$

Notice that C_1 corresponds to a $\Sigma\Pi^{[d/2]}\Sigma\Pi^{[2]}$ circuit since we assume that each Q_{ij} is a sum of univariates. The dimension of shifted partial derivatives would yield a lower bound for such $\Sigma\Pi^{[d/2]}\Sigma\Pi^{[2]}$ circuits. But what really happens to C_2 as we take some partial derivative?

Key Observation. For any multilinear monomial m, the partial derivative $\partial_m(C_2)$ only consists of non-multilinear monomials.

Thus, this points towards the following modification of the traditional dimension of shifted partial derivatives:

> For any polynomial P, look at the set of polynomials of obtained as $m_1 \cdot \partial_{m_2}(P)$ where m_1 and m_2 are *multilinear monomials* of a certain degree, and compute the dimension of the *multilinear component* of these polynomials i.e. erase all monomials that are non-multilinear and then compute the dimension of the residual polynomials.

This basically allows us to completely ignore the contribution of C_2 as we have that multilinear component of $m_1\partial_{m_2}(C_2))$ is zero for every m_1 and m_2 that are multilinear.

Both these point us to a modification of the shifted partials, which [KLSS14, KS14d] refer to as *projected shifted partial derivatives*.

[3] There are some instances when this strategy can fail spectacularly. See [KS14c]

Definition 71 (Projected Shifted Partial Derivatives). *Fix parameters $k, \ell > 0$. For any polynomial P, the set of projected shifted partials of P, denoted by $\mathrm{PSD}_{k,\ell}(P)$ is defined as follows*

$$\mathrm{PSD}_{k,\ell}(P) \quad = \quad \left\{ \mathrm{mult}(m_1 \cdot \partial_{m_2}(P)) \ : \ \begin{array}{c} \deg(m_1) = \ell \ , \ \deg(m_2) = k, \\ m_1 \textit{ and } m_2 \textit{ are multilinear} \end{array} \right\}$$

where $\mathrm{mult}(f)$ refers to the polynomial f projected to only the multilinear monomials of f. The measure $\Gamma^{\mathrm{PSD}}_{k,\ell}(P)$ is defined as the dimension of the above set of polynomials, i.e.

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(P) \quad = \quad \dim\left(\mathrm{span}(\mathrm{PSD}_{k,\ell})\right)$$

The works of [KLSS14, KS14d] use this measure to prove a lower bound for "*low-support* depth 4 circuits". As sketched earlier, the task of proving lower bounds for general homogeneous depth 4 circuits can be reduced to the *low-support* depth 4 circuits via random restrictions.

7.8.1 Reducing to 'low-support' depth 4 circuits

We have already seen a sketch of how this can be done via a random restriction but let us formalize this as a lemma.

Lemma 72. *Let P be an n-variate degree d polynomial computed by a homogeneous depth 4 circuit C of size $s \leq n^{c\sqrt{d}}$, for some $c > 0$. Let ρ be a random restriction that sets each variable to zero independently with probability $1 - 1/n^{2c}$. Then with probability at least $(1 - 1/s)$, the polynomial $\rho(P)$ is computed by a homogeneous depth 4 circuit C' with bottom support at most $\sqrt{d}$ and size at most s.*

Proof. Let $\{m_1, \ldots, m_r\}$ be the set of all monomials computed at the lowest layer of the depth 4 circuit C that are divisible by more than $\sqrt{d}$ distinct variables. Since the size of C is at most s, we also have that $r \leq s$. Then,

$$\begin{array}{rrcl} & \forall i \in [r] \qquad \Pr[\rho(m_i) \neq 0] & \leq & \dfrac{1}{n^{2c\sqrt{d}}} \\ \Longrightarrow & \Pr[\exists i \ : \ \rho(m_i) \neq 0] & \leq & \dfrac{r}{n^{2c\sqrt{d}}} \leq \dfrac{1}{n^{c\sqrt{d}}} \leq \dfrac{1}{s} \end{array}$$

Thus, with probability at least $(1 - 1/s)$, all the large support monomials are killed and C reduces to a homogeneous depth 4 circuit of bottom support at most $\sqrt{d}$. □

Handling random restrictions

The previous section outlined how in essence, it would suffice to try and find an explicit polynomial for which we can prove a good enough lower bound for bounded bottom-support depth 4 circuits. Let us say that we have found an explicit polynomial g that requires depth 4 circuits of size at least $n^{\sqrt{d}/100}$. Are we done? Let us write things down formally to see exactly what we need.

Say the polynomial we wish to show requires large homogeneous depth 4 circuits is f. Let us assume on the contrary that f can be computed by homogeneous depth 4 circuits of size $s < n^{\sqrt{d}/10000}$. Then, by Lemma 72, $\rho(f)$ can be computed by a homogeneous depth 4 circuits of bottom support bounded by $\sqrt{d}/1000$ of size s. We want to be able to say that this is a contradiction. We might be able to say that if $\rho(f)$ has g as *a projection*, that is, but setting more variables to zero in $\rho(f)$ we obtain g.

Both the results of [KLSS14] and [KS14d] proceed by showing that the polynomial g, for which they show a lower bound for bounded bottom support circuits, is robust enough to yield the lower bound even after random restriction. The calculations become trickier because the calculations of $\Gamma^{[\mathrm{PSD}]}_{k,\ell}(\rho(f))$. However, in this survey we shall use an easier approach to generically lift any g to a different polynomial f such that $\rho(f)$ has g as a projection. This trick came up during discussions with Mrinal Kumar.

Lemma 73. *Let ρ be a random restriction that sets each variable to zero independently with probability $1-p$. For any polynomial $f(y_1, \ldots y_n)$, define $f \circ \mathrm{Lin}_p$ as*

$$f \circ \mathrm{Lin}_p \quad = \quad f\left(\sum_{i=1}^{t} y_{1i}, \cdots, \sum_{i=1}^{t} y_{nt}\right) \quad \textit{where } t = \left(\frac{1}{p}\right) n \log n$$

Then, $\rho(f \circ \mathrm{Lin}_p)$ has f as a projection with probability $1 - 1/2^n$.

Proof. For any $i = 1, \ldots, n$

$$\begin{aligned} \Pr[\rho(y_{i1}) = \ldots \rho(y_{it}) = 0] &= (1-p)^t \\ &= \frac{1}{n \cdot 2^n} \\ \implies \Pr[\exists i \ : \ \rho(y_{i1}) = \ldots \rho(y_{it}) = 0] &\leq \frac{1}{2^n} \end{aligned}$$

Hence, with probability at least $1 - 1/2^n$, for each i there is some j such that $\rho(y_{ij}) \neq 0$. Therefore, with probability at least $1 - 1/2^n$, the polynomial f is a projection of $\rho(f \circ \mathrm{Lin}_p)$. □

In all the applications, as in Lemma 72, we would have $p = 1/n^{O(1)}$. Thus, we would only incur a polynomial blow-up in the number of variables from f to $f \circ \mathrm{Lin}_p$. Hence, we can focus on proving a lower bound a homogeneous depth 4 circuit of bottom support at most r (which would eventually be something like $\sqrt{d}/100$).

Lemma 74 ([KLSS14])**.** *Let P be an n-variate degree d polynomial computed by a homogeneous depth 4 circuit of size s and bottom-support at most r. Then for any k, ℓ such that $\ell + rk \leq n/2$,*

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(P) \quad \leq \quad s \cdot \binom{\frac{2d}{r} + k}{k} \cdot \binom{n}{\ell + rk}.$$

The proof of this lemma is exactly along the description in of Intuition - (2): split the circuit into multiquadratic and non-multiquadratic part, and show that the non-multiquadratic part contributes no multilinear monomials. But to just put things in perspective, we shall be dealing with parameters $r = \sqrt{d}/100$, $k = \sqrt{d}$ and $\ell = \frac{n}{2}(1-\varepsilon)$ for $\varepsilon = o(1)$. The above bound, by Lemma 69, can be seen to reduce to

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(P) \quad \leq \quad s \cdot \binom{n}{\ell} \cdot (1+\varepsilon)^{2rs} \cdot 2^{O(\sqrt{d})}$$

Sanity checks

Let us first check if this measure can at least in principle yield a lower bound for us. The best way to do this is to get some heuristic estimate of what we expect the measure to be for a random n-variate degree d polynomial R.

Heuristic Estimate. For a random n-variate degree d polynomial R, we expect the $\Gamma^{\mathrm{PSD}}_{k,\ell}(R)$ to be as large as it can be, i.e.

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

As a first step, one should first check that if we could indeed find a polynomial P for which the bound is as large as stated above, do we get a useful lower bound from Lemma 74? Turns out that if we were to choose our parameters carefully, we do indeed get the lower bound. Just to give a sense of how *careful* we need to be, here is some of the parameters that are chosen in [KLSS14, KS14d].

- The number of variables n is at least the cube of the degree d.
- The model we shall be working with is bottom-support r where $r = \sqrt{d}/1000$.
- The order of derivatives $k = \sqrt{d}$.
- The degree of the shift ℓ shall be chosen as $\ell = \frac{n}{2}(1-\varepsilon)$ where $\varepsilon = \frac{\log d}{c\sqrt{d}}$ for a suitable constant c.

The above choice of parameters might already seem pretty fragile but these are not the most delicate choices! While proving the lower bound on $\Gamma^{\mathrm{PSD}}_{k,\ell}$ for an explicit polynomial, the number of monomials etc. need to be tailored to perfection to make the proof work.

7.8.2 The surrogate rank approach of [KLSS14]

The goal is now to find an explicit polynomial P such that $\mathrm{PSD}_{k,\ell}(P)$ has large rank. One way to prove that a set of polynomials are linearly independent is to show that they have distinct leading monomials (as used [GKKS13a] etc.) Another method is to show that these polynomials are *almost orthogonal*. An example of this phenomenon can be seen in the following fact.

Fact 2. *Let M be a square matrix such that the absolute value of the diagonal entry is larger than sum of the absolute values of the non-diagonal entries in that row or column, i.e. $|M_{ii}| \geq \sum_{j \neq i} |M_{ij}|$ for all i. Then the matrix M is full rank.*

Such matrices are also called *diagonally dominant matrices*, and captures the notion of *almost orthogonal* vectors alluded to earlier. For symmetric matrices M, the following bound of Alon [Alo09].

Lemma 75 ([Alo09])**.** *For any real symmetric matrix M,*

$$\operatorname{rank}(M) \quad \geq \quad \frac{(\operatorname{Tr}(M))^2}{\operatorname{Tr}(M^2)}$$

We'll see the proof of this shortly but it would shed some more intuition to see what the above lemma yields for a diagonally dominant matrix. Let M be a matrix of the form

$$M \quad = \quad \begin{bmatrix} D & d & \dots & d \\ d & D & \dots & d \\ \vdots & \vdots & \ddots & \vdots \\ d & d & \dots & D \end{bmatrix}_{r \times r}$$

Then, $\operatorname{Tr}(M) = D \cdot r$, and $\operatorname{Tr}(M^2) = (D^2 + (r-1)d^2)r = O(D^2 r + r^2 d^2)$. If $D > (r-1)d^2$, then $\operatorname{Tr}(M^2) = O(D^2 r)$. Thus, the above lemma gives that $\operatorname{rank}(M) = \Omega(r)$.

Proof. By the spectral theorem, any real symmetric matrix has a basis of eigen vectors with eigenvalues $\lambda_1, \dots, \lambda_n$ where n is the dimension of the matrix. If $\lambda_1, \dots, \lambda_r$ are the non-zero eigenvalues, then

$$\begin{aligned} \operatorname{Tr}(M) &= \sum_{i=1}^{r} \lambda_i \\ &\leq \sqrt{r} \cdot \left(\sum_{i=1}^{r} \lambda_i^2 \right) = \sqrt{r} \cdot \operatorname{Tr}(M^2) \\ \implies r &\geq \frac{(\operatorname{Tr}(M))^2}{\operatorname{Tr}(M^2)} \end{aligned}$$

□

The bound of [KLSS14] for an explicit polynomial P proceeds by considering the matrix B where each row is indexed by a pair of multilinear monomials (m_1, m_2) of degree k and ℓ respectively, and the row is just the coefficients of the monomials of $\operatorname{mult}(m_2 \partial_{m_1}(P))$ in a fixed order. Note that B is not even a square matrix, and certainly not symmetric. However, the matrix $M = BB^T$ is a symmetric square matrix such that $\mathrm{rank}(M) \leq \mathrm{rank}(B)$.

Let us spend some time understand the entries of M. The (i,j)-th entry of M is precisely the inner-product of row i and row j of B. If P is a polynomial with just zero-one coefficients, then the i-th diagonal entry is precisely the number of non-zero entries in row i of B. Thus,

$$\begin{aligned}\mathrm{Tr}(M) &= \text{number of non-zero entries in } B\\ &= (\text{\# cols of } B) \cdot \mathop{\mathbb{E}}_i[\text{\# non-zero entries in } i\text{-th col of } B]\end{aligned}$$

The calculation for $\mathrm{Tr}(M^2)$ requires a little more care. Let M_i refer to the i-th row of M and B_i refer to the i-th row of B. Then,

$$\begin{aligned}\mathrm{Tr}(M^2) &= \sum_i \langle M_i, M_i\rangle\\ &= \sum_i\sum_j \langle B_i, B_j\rangle^2 \quad = \quad \sum_i\sum_j\left(\sum_m B_{im}B_{jm}\right)^2\\ &= \sum_i\sum_j\sum_m B_{im}^2B_{jm}^2 \quad + \quad \sum_i\sum_j\sum_{m\neq m'} B_{im}B_{im'}B_{jm}B_{jm'}\\ &= \sum_m\left(\sum_i\sum_j B_{im}B_{jm}\right) \quad + \quad \sum_i\sum_j\sum_{m\neq m'} B_{im}B_{im'}B_{jm}B_{jm'}\\ &= \qquad T_1 \qquad + \qquad T_2\end{aligned}$$

The first term T_1 is easy to calculate:

$$\begin{aligned}T_1 \quad &= \quad (\text{\# cols of } B)\cdot \mathop{\mathbb{E}}_i[(\text{\# non-zero entries in } i\text{-th col of } B)^2]\\ &\overset{\text{(hopefully)}}{\approx} \quad (\text{\# cols of } B)\cdot \mathop{\mathbb{E}}_i[(\text{\# non-zero entries in } i\text{-th col of } B)]^2\end{aligned}$$

The term T_2 roughly corresponds to the number of 2×2 submatrices of B that is $\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$. If we could somehow show that there are not too many such submatrices, then $\mathrm{Tr}(M^2)$ is essentially dominated by T_1. That would then yield that $\mathrm{rank}(M) \gtrapprox$ (# cols of B).

Obtaining a bound on T_2:

$$T_2 \quad = \quad \sum_i\sum_j\sum_{m\neq m'} B_{im}B_{im'}B_{jm}B_{jm'}$$

Each term $B_{im}B_{im'}B_{jm}B_{jm'}$ that is non-zero corresponds to a 2×2 submatrix of B (indexed by rows i,j and columns m,m') that is $\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$.

The columns of B are indexed by multilinear monomials of degree $\ell + d - k$, and the rows of B are indexed by a derivative and a shift. Let row i correspond to $\mathrm{mult}(\gamma_1 \cdot \partial_{\alpha_1}(P))$ and row j to $\mathrm{mult}(\gamma_1 \cdot \partial_{\alpha_1}(P))$. Thus, if the 2×2 minor indexed by rows i, j and columns m, m' equals $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, then there exists $\beta_1, \beta_2, \beta_3, \beta_4 \in P$ such that

$$\begin{aligned} m &= \frac{\beta_1}{\alpha_1} \cdot \gamma_1 = \frac{\beta_3}{\alpha_2} \cdot \gamma_2 \\ m' &= \frac{\beta_2}{\alpha_1} \cdot \gamma_1 = \frac{\beta_4}{\alpha_2} \cdot \gamma_2 \\ &\implies \frac{\beta_1}{\beta_3} = \frac{\beta_2}{\beta_4} \end{aligned}$$

Following notation used in [KLSS14], we shall call $\beta_1, \beta_2, \beta_3, \beta_4$ as the *label* of the 2×2 minor. Since $m \neq m'$, we also have that $\beta_1 \neq \beta_2$. What we'd like to say that the only way $\beta_1/\beta_3 = \beta_2/\beta_4$ is if $\beta_3 = \beta_1$ and $\beta_2 = \beta_4$. This need not be true in general of course, but this is where the choice of the polynomial comes in.

Claim 76. *If P is the $\mathrm{NW}_{d,d^3,e}$ polynomial for $e = \frac{d}{3}$ then any 2×2 minor of B (with the order of derivatives $k = o(d)$) that is $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ has label $\beta_1, \beta_2, \beta_3, \beta_4$ where $\beta_1 = \beta_3$ and $\beta_2 = \beta_4$, or $\beta_1 = \beta_2$ and $\beta_3 = \beta_4$.*

Proof. Assume that $\beta_1 \neq \beta_3$. Then by Lemma 70 we know that they differ in at least $2d/3$ places. But then, $\beta_1/\beta_3 = \beta_2/\beta_4$ forces that β_1 and β_3 must agree at least $2d/3$ places forcing $\beta_1 = \beta_2$. □

Thus, for the NW-polynomial the number of such boxes is quite small. Using this, albeit with a reasonable amount of sweat, one can estimate T_2 to show that $T_2 = O(T_1)$. Thus, [KLSS14] obtain the following bound.

Lemma 77 ([KLSS14])**.** *For the polynomial $\mathrm{NW}_{d,d^3,e}$, for $e = \frac{d}{3}$, and $k = \sqrt{d}$ and $\ell = \frac{n}{2}\left(1 - \frac{\log d}{\sqrt{d}}\right)$ we have the bound*

$$\Gamma_{k,\ell}^{\mathrm{PSD}}(\mathrm{NW}_{d,d^3,e}) \quad \geq \quad \frac{1}{\mathsf{poly}(n,d)} \cdot \min\left(\binom{n}{\ell + d - k}, \binom{d}{k}^2 \cdot d^k \cdot k! \cdot \binom{n}{\ell}\right)$$

Note that the first term of the $\min$ in the RHS is the number of columns of B, as we had heuristically estimated. Simplifying the RHS using Lemma 69, we get

$$\Gamma_{k,\ell}^{\mathrm{PSD}}(\mathrm{NW}_{d,d^3,e}) \quad \geq \quad \frac{1}{\mathsf{poly}(n,d)} \cdot \binom{n}{\ell} \cdot \exp\left(c \cdot \varepsilon(d-k)\right)$$

for some constant $c > 0$. Since $\varepsilon = \frac{\log d}{\sqrt{d}}$, we get

$$\Gamma_{k,\ell}^{\mathrm{PSD}}(\mathrm{NW}_{d,d^3,e}) \quad \geq \quad \frac{1}{\mathsf{poly}(n,d)} \cdot \binom{n}{\ell} \cdot \exp\left(c \cdot \sqrt{d} \cdot \log d\right)$$

With the above bound and Lemma 74, we get the lower bound of [KLSS14].

Theorem 78 ([KLSS14]). *Any depth 4 homogeneous circuit of bottom support $r = \sqrt{d}/1000$ computing the polynomial $\mathrm{NW}_{d,d^3,d/3}$ over a characteristic zero field must have top fan-in $s = d^{\Omega(\sqrt{d})}$.*
In fact, more generally, any homogeneous depth 4 circuit of bottom support bounded by r computing $\mathrm{NW}_{d,m,e}$ for suitably chosen parameters must have top fanin $s = d^{\Omega(d/r)}$.

Coupling with Lemma 73, we obtain (a slight reformulation of) their main theorem.

Theorem 79 ([KLSS14]). *Any depth 4 homogeneous computing the polynomial $\mathrm{NW}_{d,d^3,d/3} \circ \mathrm{Lin}$ over a characteristic zero field must have size $s = d^{\Omega(\sqrt{d})}$.*

7.8.3 The leading monomial approach of [KS14d]

Shortly after [KLSS14], a purely combinatorial proof of the result was presented by [KS14d]. More over, they were able to prove the lower bound of $n^{\Omega(\sqrt{d})}$ for the size of any homogeneous depth 4 circuit computing $\mathrm{IMM}_{n,d}$ (for some suitable choices of n and d). This was a strengthening of [KLSS14] in two ways – (1) it worked over any field, and (2) the lower bound was for a polynomial that we know can be computed small arithmetic circuit.
The calculations of [KS14d] are much more trickier than [KLSS14] but there are quite a few interesting ideas that would even have application in other areas.

The earlier lower bounds of [GKKS13a, KSS13, FLMS13] required a lower bound on the dimension of shifted partial derivatives of a polynomial P, and this was obtained by finding a *large* set of *distinct leading monomials*. In [KS14d], they take this approach but require a very careful analysis. The key difference in this setting is the following:

> If β is the leading monomial of a polynomial P, then for any monomial γ, we also have that $\beta \cdot \gamma$ is the leading monomial of γP.
>
> However, the leading monomial of $\mathrm{mult}(\gamma P)$ could be $\beta' \cdot \gamma$ for some $\beta' \neq \beta$ (as higher monomials could be made non-multilinear during the shift by γ).

The multilinear projection makes the task of counting leading monomials much harder and [KS14d] come up with a clever method to estimate this.

Leading monomials after multilinear projections

Let P the polynomial for which we are trying to lower bound $\Gamma^{\mathrm{PSD}}_{k,\ell}(P)$. For every monomial multilinear monomial α of degree k, and a monomial $\beta \in \partial_\alpha(P)$, define the set $A(\alpha, \beta)$ as

$$A(\alpha,\beta) \quad = \quad \left\{ \gamma \ : \ \begin{array}{c} \deg(\gamma) = \ell + d - k \text{ and there is a } \gamma' \text{ of degree } \ell \\ \text{such that } \gamma = \mathrm{LM}(\mathrm{mult}(\gamma' \cdot \partial_\alpha(P))) = \gamma' \cdot \beta \end{array} \right\}$$

In other words, we want the number of distinct monomials that are contributed by β, which are also distinct leading monomials obtained from $\partial_\alpha(P)$ that are divisible by β. We then have

$$\Gamma_{k,\ell}^{\mathrm{PSD}}(P) \quad \geq \quad \left| \bigcup_{\alpha,\beta} A(\alpha,\beta) \right| \tag{7.5}$$

The standard technique to obtain a lower bound on the union of sets is via the *Inclusion-Exclusion* principle.

Lemma 80 (Inclusion-Exclusion Principle)**.** *For any collection of sets $A_1, \ldots, A_r$,*

$$\left| \bigcup_i A_i \right| \quad \geq \quad \sum_i |A_i| \quad - \quad \sum_{i \neq j} |A_i \cap A_j|$$

If we were to somehow show that $\sum_{i\neq j} |A_i \cap A_j| \leq \frac{1}{2} \sum_i |A_i|$, then we obtain that $|\cup_i A_i| \geq \frac{1}{2} \cdot \sum_i |A_i|$. This is what shall be employed for the sets $A(\alpha,\beta)$, except that we quickly run into two immediate problems.

1. How do we even estimate $A(\alpha,\beta)$? The set of γ' such that $\gamma'\beta = \mathrm{LM}(\partial_\alpha(P))$ do not seem to have any nice combinatorial structure.
2. What if it so happens that $\sum |A(\alpha_1,\beta_1) \cap A(\alpha_2,\beta_2)| = 100 \sum |A(\alpha,\beta)|$? Inclusion-Exclusion does not yield anything in that case.

It so turns out that the second point actually is the case. In fact for $\mathrm{IMM}_{n,d}$, the second term turns out to be greater than the first term by a factor of $n^{\sqrt{d}/1000}$ or so! In [KS14d], they prove a wonderful strengthened version of the Inclusion-Exclusion principle which allows them to handle the second hurdle.

Lemma 81 (Stronger Inclusion-Exclusion [KS14d])**.** *Let $A_1, \ldots, A_r$ be sets such that there is some $\lambda > 1$ such that*

$$\sum_{i \neq j} |A_i \cap A_j| \quad \leq \quad \sum_i \lambda \cdot |A_i|$$

Then,

$$\left| \bigcup_i A_i \right| \quad \geq \quad \left(\frac{1}{4\lambda} \right) \cdot \left(\sum_i |A_i| \right)$$

In other words, as long as the second term of the Inclusion-Exclusion principle is *not too much larger* than the first term, we still can get non-trivial bounds on the union.

Proof. Let $p = \frac{1}{2\lambda} < 1$. Define sets $A_1', \ldots, A_r'$ such that $A_i' \subseteq A_i$ obtained by adding each element of A_i to A_i' independently with probability p. Since $A_i' \subseteq A_i$, we also have that $|\cup A_i| \geq |\cup A_i'|$. By linearity of expectation,

$$\mathbb{E}\left[\sum_i |A_i'| \right] \quad = \quad p \sum_i |A_i|$$

More importantly, by the sampling process,

$$\mathbb{E}\left[\left|A'_i \cap A'_j\right|\right] \quad = \quad p^2 \cdot |A_i \cap A_j|$$

as any common element must be added to both A'_i *and* A'_j, and either of these events happen independently with probability p each. Since $\sum_{i,j}\left|A'_i \cap A'_j\right|$ drops by a factor of p^2, we are now in a position to apply the Lemma 80 to the A'_is.

$$\begin{aligned}
\left|\bigcup A_i\right| &\geq \mathbb{E}\left[\left|\bigcup A'_i\right|\right] \\
&\geq \mathbb{E}\left[\sum_i |A'_i|\right] \quad - \quad \mathbb{E}\left[\left|A'_i \cap A'_j\right|\right] \\
&= p\left(\sum_i |A_i|\right) \quad - \quad p^2\left(\sum_{i\neq j}|A_i \cap A_j|\right) \\
&\geq p\left(\sum_i |A_i|\right) \quad - \quad p^2\lambda\left(\sum_i |A_i|\right) \\
&\geq \frac{p}{2}\left(\sum_i |A_i|\right) \quad = \quad \frac{1}{4\lambda}\left(\sum_i |A_i|\right)
\end{aligned}$$

□

We can now proceed to lower bound $|\bigcup A(\alpha,\beta)|$ via inclusion exclusion.

Estimating $|\bigcup A(\alpha,\beta)|$ via Inclusion-Exclusion

$$\left|\bigcup_{\alpha,\beta} A(\alpha,\beta)\right| \quad \geq \quad \sum_{\alpha,\beta}|A(\alpha,\beta)| \quad - \quad \sum_{(\alpha,\beta)\neq(\alpha',\beta')}|A(\alpha,\beta)\cap A(\alpha',\beta')|$$

Let us first address the term $\sum|A(\alpha,\beta)|$. As mentioned earlier, it is not an easy task to get a good handle on the set $A(\alpha,\beta)$ for polynomial such as NW or IMM, for any reasonable monomial ordering. However, [KS14d] circumvent this difficult by using an indirect approach to estimate this term.

For any derivative α and $\beta \in \partial_\alpha(P)$, define the set $S(\alpha,\beta)$ as the following set of multilinear monomials of degree ℓ that is disjoint from β.

$$S(\alpha,\beta) \quad = \quad \left\{\gamma \;:\; \begin{array}{c}\gamma \text{ is multilinear, has} \\ \text{degree } \ell \text{ and } \gcd(\beta,\gamma)=1\end{array}\right\}$$

This on the other hand is independent of any monomial ordering, and is also easy to calculate:

$$\text{For every } \alpha,\beta \qquad |S(\alpha,\beta)| \quad = \quad \binom{n-d+k}{\ell}.$$

Lemma 82 ([KS14d]). *For any α,*

$$\sum_{\beta} |A(\alpha,\beta)| \quad \geq \quad \left|\bigcup_{\beta} S(\alpha,\beta)\right|$$

Proof. Consider any $\gamma \in \bigcup_\beta S(\alpha,\beta)$. By definition, there is at least one non-multilinear monomial in $\gamma \cdot \partial_\alpha(P)$. Thus, in particular $\mathrm{LM}(\mathrm{mult}(\gamma \cdot \partial_\alpha(P))$ is non-zero and equal to some $\gamma \cdot \beta$ for some monomial $\beta \in \partial_\alpha(P)$. This also implies that $\gamma' = \gamma \cdot \beta \in A(\alpha,\beta)$. This yields an injective map ϕ

$$\phi : \bigcup_{\beta} S(\alpha,\beta) \quad \rightarrowtail \quad \{(\beta,\gamma') \ : \ \beta \in \partial_\alpha(P) \ , \ \gamma' \in A(\alpha,\beta)\}$$

Since the size of the RHS is precisely $\sum_\beta |A(\alpha,\beta)|$, the lemma follows. □

Thus, by another use of Inclusion-Exclusion on the $S(\alpha,\beta)$'s, we get

$$\begin{aligned}\left|\bigcup_{\alpha,\beta}(\alpha,\beta)\right| \quad &\geq \quad \sum_{\alpha,\beta}|A(\alpha,\beta)| \quad - \quad \sum_{(\alpha,\beta)\neq(\alpha',\beta')} |A(\alpha,\beta)\cap A(\alpha',\beta')| \\ &\geq \quad \sum_{\alpha}\left(\sum_{\beta}|S(\alpha,\beta)|\right) \quad - \quad \sum_{\alpha}\left(\sum_{\beta\neq\beta'}|S(\alpha,\beta)\cap S(\alpha,\beta')|\right) \\ &\qquad - \quad \sum_{(\alpha,\beta)\neq(\alpha',\beta')} |A(\alpha,\beta)\cap A(\alpha',\beta')|\end{aligned}$$

Let us call the three terms in the RHS of the last equation as T_1, T_2 and T_3 respectively. Since we know the size of each $S(\alpha,\beta)$ exactly, the value of T_1 is easily obtained.

Lemma 83 ([KS14d]).

$$\begin{aligned}T_1 \quad &= \quad \textit{(\# derivs)} \cdot \textit{(\# mons in a deriv)} \cdot \binom{n-d+k}{\ell} \\ &\approx \quad \textit{(\# derivs)} \cdot \textit{(\# mons in a deriv)} \cdot \binom{n}{\ell} \cdot \left(\frac{1+\varepsilon}{2}\right)^{d-k}\end{aligned}$$

Let $T_1(\alpha) = \sum_\beta |S(\alpha,\beta)|$ for any choice of α. So far we have not used any property of the polynomial P. But this becomes crucial in the calculation of T_2 and T_3. To get a sense of how these calculations proceed in [KS14d], we outline the calculation of T_2 for the case of $P = \mathrm{NW}_{d,m,e}$ for suitable choices of the parameters m, d, e.

Lemma 84 ([KS14d]). *For the polynomial* $\mathrm{NW}_{d,m,e}$*, if* $n = md$ *and* $\ell = \frac{n}{2}(1-\varepsilon)$ *for* $\varepsilon = o(1)$

$$T_2 \quad \leq \quad \textit{(\# derivs)} \cdot \textit{(\# mons per deriv)}^2 \cdot \binom{n}{\ell} \cdot \left(\frac{1+\varepsilon}{2}\right)^{2d-2k}$$

Proof. For any fixed derivative α, define

$$T_2(\alpha) = \sum_{\beta \neq \beta'} |S(\alpha,\beta) \cap S(\alpha,\beta')| .$$

For any pair of multilinear degree $(d-k)$ monomials $\beta \neq \beta' \in \partial_\alpha(P)$ such that $\deg(\gcd(\beta,\beta')) = t$, we know that

$$|S(\alpha,\beta) \cap S(\alpha,\beta')| \quad = \quad \binom{n-2d+2k+t}{\ell}$$

Thus, if we can count the number of pairs (β,β') that agree on exactly t places, we can obtain $T_2(\alpha)$. Note that for $\mathrm{NW}_{d,m,e}$, any two $\beta,\beta' \in \partial_\alpha(\mathrm{NW}_{d,m,e})$ can agree on at most $e-k$ places. Further, the number of pairs that agree in exactly $0 \leq t \leq e-k$ places is at most

$$m^{e-k} \cdot \binom{d-k}{t} \cdot (m-1)^{e-t}$$

as there are m^{e-k} choices for β, and $\binom{d-k}{t}$ choices for places where they may agree, and $(m-1)^{e-t}$ choices for β' that agree with β on those t places. Thus,

$$\begin{aligned}
T_2(\alpha) \quad &\leq \quad \sum_{t=0}^{e-k} m^{e-k} \cdot \binom{d-k}{t} \cdot (m-1)^{e-t} \cdot \binom{n-2d+2k+t}{\ell} \\
&\approx \quad \sum_{t=0}^{e-k} m^{e-k} \cdot \binom{d-k}{t} \cdot (m-1)^{e-t} \cdot \binom{n}{\ell} \frac{1}{2^{2d-2k-t}} \cdot (1+\varepsilon)^{2d-2k-t} \\
&\leq \quad m^{2e} \binom{n}{\ell} \left(\frac{1+\varepsilon}{2}\right)^{2d-2k} \cdot \sum_{t=0}^{e-k} \binom{d-k}{t} \left(\frac{2}{(1+\varepsilon)m}\right)^{t} \\
&\leq \quad m^{2e} \binom{n}{\ell} \left(\frac{1+\varepsilon}{2}\right)^{2d-2k} \cdot \left(1 + \frac{2}{(1+\varepsilon)m}\right)^{d-k} \\
&= \quad m^{2e} \cdot \binom{n}{\ell} \cdot \left(\frac{1+\varepsilon}{2}\right)^{2d-2k} \cdot O(1) \qquad \text{if } m = \Omega(d)
\end{aligned}$$

Thus,

$$T_2 \quad \leq \quad (\text{\# derivs}) \cdot (\text{\# mons per deriv})^2 \cdot \binom{n}{\ell} \cdot \left(\frac{1+\varepsilon}{2}\right)^{2d-2k}$$

□

Combining this with Lemma 83 and using Lemma 81,

$$\sum_{\alpha,\beta} |A(\alpha,\beta)| \quad \geq \quad (\text{\# derivs}) \cdot \frac{T_1(\alpha)}{\max(2, \frac{4T_2(\alpha)}{T_1(\alpha)})}$$

To maximize this, if we choose the parameters m, d, e such that $T_1(\alpha) = T_2(\alpha)$, we obtain the following corollary.

Corollary 85. *Consider the polynomial* $\mathrm{NW}_{d,m,e}$ *with* $n = md$ *and* $m = \Omega(d)$*. If* $\ell = \frac{n}{2}(1-\varepsilon)$ *for* $\varepsilon = o(1)$ *and* e *chosen so that*

$$m^{e-k} \quad = \quad \left(\frac{2}{1+\varepsilon}\right)^{d-k} \cdot 2^{\Theta(\sqrt{d})}$$

then

$$\sum_{\alpha,\beta} |A(\alpha,\beta)| \quad \geq \quad \textit{(\# derivs)} \cdot \binom{n}{\ell} \cdot 2^{\Theta(\sqrt{d})}$$

Proof. If $T_1(\alpha) = T_2(\alpha) \cdot 2^{-\Theta(\sqrt{d})}$ then

$$\begin{aligned} \sum_{\alpha,\beta} |A(\alpha,\beta)| \quad &\geq \quad \text{(\# derivs)} \cdot \frac{T_1(\alpha)}{\max(2, \frac{4T_2(\alpha)}{T_1(\alpha)})} \\ &= \quad \text{(\# derivs)} \cdot \frac{T_1(\alpha)^2}{4T_2(\alpha)} \\ &= \quad \text{(\# derivs)} \cdot \binom{n}{\ell} \cdot 2^{\Theta(\sqrt{n})} \end{aligned}$$

Note that $T_1(\alpha) = T_2(\alpha) \cdot 2^{-\Theta(\sqrt{d})}$ forces

$$\text{(\# mon per deriv)} \quad = \quad m^{e-k} \quad = \quad \left(\frac{2}{1+\varepsilon}\right)^{d-k} \cdot 2^{\Theta(\sqrt{d})}$$

□

Note that e needs to tailored very precisely to force the above condition! If e is chosen too large or small, we get nothing from this whole exercise!
In the case of IMM this calculations gets a lot messier. The calculation would similarly force that the number of monomials must be in a very narrow range. This is achieved by instead looking at a random subgraph of the generic ABP of suitable sparsity to ensure the following two properties:

- The number of monomials in any derivative is exactly as demanded.
- 'Most' pairs of monomials (β, β') agree on 'few' places.

Upper bounding $\sum |A(\alpha,\beta) \cap A(\alpha',\beta')|$

We are still left with the task of upper bounding

$$T_3 \quad = \quad \sum_{(\alpha,\beta)\neq(\alpha',\beta')} |A(\alpha,\beta) \cap A(\alpha',\beta')|$$

As mentioned earlier, we really do not have a good handle on the set $A(\alpha, \beta)$, and certainly not on the intersection of two such sets. Once again, we shall use a proxy that is easier to estimate to upper bound T_3.
The set $A(\alpha, \beta) \cap A(\alpha', \beta')$ consists of multilinear monomials γ of degree $\ell + d - k$ such that there exists multilinear monomials γ', γ'' of degree ℓ satisfying

$$\begin{aligned} \gamma &= \gamma'\beta \quad = \quad \gamma''\beta', \\ \gamma'\beta &= \mathrm{LM}(\mathrm{mult}(\gamma'\partial_\alpha(P))) \\ \text{and} \quad \gamma''\beta' &= \mathrm{LM}(\mathrm{mult}(\gamma''\partial_{\alpha'}(P))) \end{aligned}$$

This in particular implies that γ must be divisible by both β and β'.

Observation 86. *If* $\deg(\gcd(\beta, \beta')) = t$, *then*

$$|A(\alpha, \beta) \cap A(\alpha', \beta')| \quad \leq \quad \binom{n - 2d + 2k + t}{\ell - d + k + t}$$

Proof. Every monomial $\gamma \in A(\alpha, \beta) \cap A(\alpha', \beta')$ must be divisible by β and β'. Since $|\beta \cup \beta'| = 2d - 2k - t$, the number of choices of γ is precisely

$$\binom{n - (2d - 2k - t)}{(\ell + d - k) - (2d - 2k - t)} \quad = \quad \binom{n - 2d + 2k + t}{\ell - d + k + t} \qquad \square$$

One needs a similar argument as in the case of T_2 to figure out how many pairs $(\alpha, \beta) \neq (\alpha', \beta')$ are there with $\deg(\gcd(\beta, \beta')) = t$ and sum them up accordingly. We shall just state the bound of [KS14d] here without proof.

Lemma 87 ([KS14d])**.** *For the polynomial* $\mathrm{NW}_{d,m,e}$*, and* $n = md$ *and* $\ell = \frac{n}{2}(1 - \varepsilon)$ *for* $\varepsilon = o(1)$*,*

$$T_3 \quad \leq \quad (\textit{\# deriv})^2 (\textit{\# mons per deriv})^2 \cdot \binom{n}{\ell} \cdot \left(\frac{1}{2}\right)^{2d-2k}$$

Recalling that we have chosen our parameters so that

$$(\text{\# mons per deriv}) = \left(\frac{2}{1 + \varepsilon}\right)^{d-k} \cdot 2^{\Theta(\sqrt{d})}$$

the above equation reduces to

$$T_3 \quad \leq \quad (\text{\# deriv})^2 \left(\frac{1}{1 + \varepsilon}\right)^{2(d-k)} \cdot \binom{n}{\ell}.$$

Combining with Corollary 85, we obtain the required bound for $|\bigcup A(\alpha, \beta)|$.

Lemma 88. *Consider polynomial* $\mathrm{NW}_{d,m,e}$ *where* $n = md$ *and* e *chosen so that*

$$m^{e-k} \quad = \quad \left(\frac{2}{1+\varepsilon}\right)^{d-k} \cdot 2^{\Theta(\sqrt{d})}$$

If $\varepsilon = \frac{\log d}{c\sqrt{d}}$ *for a large enough constant* c*, and* $k = O(\sqrt{d})$ *and* $\ell = \frac{n}{2}(1-\varepsilon)$*, then*

$$\Gamma^{\mathrm{PSD}}_{k,\ell}(\mathrm{NW}_{d,m,e}) \quad \geq \quad \left|\bigcup_{\alpha,\beta} A(\alpha,\beta)\right| \quad \geq \quad \binom{n}{\ell} \cdot (1+\varepsilon)^{2d-2k} \cdot 2^{\Theta(\sqrt{d})}$$

With Lemma 74, we obtain the lower bound for low-bottom-support homogeneous depth 4 circuits.

Theorem 89 ([KS14d])**.** *Any homogeneous depth* 4 *circuit with bottom support bounded by* $r = \sqrt{d}/1000$ *computing, over any field* $\mathbb{F}$*, the polynomial* $\mathrm{NW}_{d,m,e}$ *with parameters as defined above must have top fan-in* $s = d^{\Omega(\sqrt{d})}$*.*
In fact, more generally, any homogeneous depth 4 *circuit of bottom support bounded by* r *computing* $\mathrm{NW}_{d,m,e}$ *for suitably chosen parameters must have top fanin* $s = d^{\Omega(d/r)}$*.*

Again, coupling with Lemma 73, we obtain (a slight reformulation of) their theorem.

Theorem 90 ([KLSS14])**.** *Any homogeneous depth* 4 *circuit computing, over any field* $\mathbb{F}$*, the polynomial* $\mathrm{NW}_{d,m,e} \circ \mathrm{Lin}$ *with parameters as defined above must have top fan-in* $s = d^{\Omega(\sqrt{d})}$*.*
A similar lower bound $d^{\Omega(\sqrt{d})}$ *holds also for the polynomial* $\mathrm{IMM}_{n,d} \circ \mathrm{Lin}$ *for suitable choices of* n *and* d*.*

7.9 Non-homogeneous depth 3 circuits

In a very recent result, [KS14a] show that similar techniques can also be used to prove lower bounds for subclasses of non-homogeneous depth three circuits, namely depth three circuits with *bounded bottom fan-in*. We shall denote the class of depth three circuits of bottom fan-in bounded by r as $\Sigma\Pi\Sigma^{[r]}$ circuits.
But before we see this lower bound, let us first understand the computational power of depth three circuits, and the depth reduction of [GKKS13b].

7.9.1 Computational power of depth three circuits

A $\Sigma\Pi\Sigma$ circuit computes a polynomial of the form

$$f \quad = \quad \sum_{i=1}^{s} \ell_{i1} \ldots \ell_{iD}$$

If the circuit is non-homogeneous, the degree of the circuit D could potentially be much larger than $\deg(f)$.

The class of depth three arithmetic circuits can compute polynomials in non-trivial ways. To illustrate a couple of examples, there is a homogeneous $\Sigma\Pi\Sigma$ circuit for Perm_n of size $2^{O(n)}$ called Ryser's Formula [Rys63]

$$\mathsf{Perm}_n \quad = \quad \sum_{S \subseteq [n]} (-1)^{n-|S|} \prod_{i=1}^{n} \left(\sum_{j \in S} x_{ij} \right) \tag{7.6}$$

On the other hand, no $\Sigma\Pi\Sigma$ circuit for the Det_n significantly better than writing it as a sum of $n!$ monomials was known (until [GKKS13b]). Further, the elementary symmetric polynomials $\mathrm{Esym}_k(x_1, \ldots, x_n)$ of degree k defined as

$$\mathrm{Esym}_k(\mathbf{x}) \quad = \quad \sum_{\substack{S \subset \mathbf{x} \\ |S|=k}} \prod_{x_i \in S} x_i$$

can be computed by a non-homogeneous depth 3 circuit of size $O(n^2)$ over any characteristic zero field. In stark contrast, [NW97] showed that any homogeneous depth 3 circuit computing Esym_k requires size $n^{\Omega(k)}$. [NW97] also showed a $2^{\Omega(n)}$ lower bound for homogeneous depth 3 circuits computing Perm_n or Det_n.
Also, the results of [GR00, GK98] showed a $2^{\Omega(n)}$ lower bound for $\Sigma\Pi\Sigma$ circuits *over finite fields* that compute Det_n or Perm_n. All these results seemed to suggest that there perhaps is an $2^{\Omega(n)}$ lower bound for $\Sigma\Pi\Sigma$ circuits computing Det_n over characteristic zero fields as well. If it was true over finite fields, and for homogeneous $\Sigma\Pi\Sigma$ circuits, how much power can characteristic zero fields and non-homogeneity add? As it turns out, quite a lot!

Theorem 91 ([GKKS13b]). *Let f be an n-variate degree d polynomial computed by an arithmetic circuit of size s over any characteristic zero field. Then there is a $\Sigma\Pi\Sigma$ circuit of size $s' \leq s^{O(\sqrt{d})}$ that computes f.*

Corollary 92 ([GKKS13b]). *There is a $\Sigma\Pi\Sigma$ circuit over $\mathbb{Q}$, the field of rational numbers, of size $n^{O(\sqrt{n})}$.*

The proof is quite short and comprises of two steps using known reductions, and going through a bizarre intermediate model of *depth 5 powering circuits*. Simply presenting the proof step-by-step would rob the readers of the intuition as to why one would study depth 5 powering circuits. This result was really a bi-product of an attempt to prove a stronger lower bound for depth 4 circuits. We believe this perspective, albeit lengthier, is more insightful than seeing the proof directly.

Towards proving better lower bounds for depth 4 circuits

From Theorem 67, it suffices to prove a better lower bound for explicit polynomials computed as

$$f \quad = \quad \sum_{i=1}^{s} Q_{i1} \dots Q_{ir} \quad \text{where} \quad \deg(Q_{ij}) \leq \sqrt{d}\ ,\ r \leq O(\sqrt{d}) \tag{7.7}$$

The goal is to show a lower bound of $s = n^{\omega(\sqrt{d})}$. Perhaps a simpler question to ask is to prove a lower bound for expressions of the form

$$f \quad = \quad \sum_{i=1}^{s} Q_i^{\sqrt{d}} \quad \text{where } \deg(Q_i) \leq \sqrt{d} \tag{7.8}$$

Fortunately, if the goal is to prove lower bounds of $n^{\omega(\sqrt{d})}$, then without loss of generality we can focus on this equation instead!

Lemma 93. *Over any characteristic zero field, given an expression of the form*

$$f \quad = \quad \sum_{i=1}^{s} Q_{i1} \dots Q_{ir} \quad \textit{where} \quad \deg(Q_{ij}) \leq \sqrt{d}\ ,\ r \leq O(\sqrt{d})$$

there is an equivalent equation

$$f \quad = \quad \sum_{i=1}^{s'} Q_i^{r} \quad \textit{where} \quad \deg(Q_i) \leq \sqrt{d}$$

with $s' \leq s \cdot 2^{O(\sqrt{r})}$.

Proof. Consider Ryser's formula (7.6) applied for to the $r \times r$ matrix where each row is $[y_1, \dots, y_r]$.

$$\mathsf{Perm}\begin{bmatrix} y_1 & \cdots & y_r \\ \vdots & \ddots & \vdots \\ y_1 & \cdots & y_r \end{bmatrix} \quad = \quad r! \cdot y_1 \dots y_r \quad = \quad \sum_{S \subseteq [r]} \left(\sum_{j \in S} y_j \right)^n$$

This specific identity is often attributed to Fischer [Fis94]. The lemma follows by applying this identity on each term $Q_{i1} \dots Q_{ir}$. □

Note that since we need to divide by $r!$, the above lemma fails over low characteristic fields, in particular finite fields. Thus, proving an $n^{\omega(\sqrt{d})}$ lower bound for expressions such as (7.8) implies an $n^{\omega(\sqrt{d})}$ lower bound for expressions such as (7.7). We shall call expressions such as (7.8) as $\Sigma \wedge \Sigma \Pi^{[\sqrt{d}]}$ circuits.

Just as we converted the top Π layer into powering layers using Fischer's identity, the same can be done to the lower layer of Π gates as well.

Corollary 94. *If a homogeneous n-variate degree d polynomial f can be computed by a $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ of size $s = n^{O(\sqrt{d})}$, then f can also be computed by an $\Sigma\wedge^{[O(\sqrt{d})]}\Sigma\wedge^{[\sqrt{d}]}\Sigma$ circuit of size $s' = s\cdot 2^{O(\sqrt{d})}$. Conversely, if f requires $\Sigma\wedge^{[O(\sqrt{d})]}\Sigma\wedge^{[\sqrt{d}]}\Sigma$ circuits of size $s' = n^{\omega(\sqrt{d})}$ to compute it, then f cannot be computed by polynomial sized arithmetic circuits.*

We shall take a small detour to see if non-homogeneous depth 3 circuits can be converted to homogeneous shallow circuits without much blow-up in size.

Non-homogeneous depth 3 to homogeneous depth 5 circuits

Let f be a homogeneous degree d polynomial computed by a possibly non-homogeneous depth 3 circuit C of the form

$$f \quad = \quad \sum_{i=1}^{s} \ell_{i1} \ldots \ell_{iD}$$

As a first step, let us extract the degree d homogeneous component of each summand on the RHS. Since f is a homogeneous degree d polynomial, f has to be sum of the degree d homogeneous components of each summand on the RHS. Consider a single term of the form

$$T \quad = \quad (\ell_1 + \alpha_1) \cdots (\ell_D + \alpha_D)$$

where each ℓ_i is a homogeneous linear polynomial, and α are elements from the field. Assuming that the first r of the α_i's are zero, we can write T in the form (with some reuse of symbols)

$$\begin{aligned} T &= \alpha \cdot \ell_1 \ldots \ell_r \cdot (\ell_{r+1} + 1) \ldots (\ell_D + 1) \\ \implies \quad \mathrm{Hom}_d(T) &= \ell_1 \ldots \ell_r \cdot \mathrm{Esym}_{d-r}(\ell_{r+1}, \ldots, \ell_D) \end{aligned}$$

where $\mathrm{Esym}_k(\mathbf{x})$, the elementary symmetric polynomial of degree k defined as

$$\mathrm{Esym}_k(\mathbf{x}) \quad = \quad \sum_{\substack{S \subset \mathbf{x} \\ |S| = k}} \prod_{x_i \in S} x_i$$

Hence, if we can show that $\mathrm{Esym}_{d-r}(\mathbf{x})$ has a not-too-large homogeneous depth 4 circuit, then we can immediately infer that f can be computed by a not-too-large homogeneous depth 5 circuit. The following identities, attributed to Newton (cf. [Lit50]), is exactly what we need. Define the *power symmetric polynomials*, denoted by $\mathrm{Pow}_k(\mathbf{x})$ as

$$\mathrm{Pow}_k(\mathbf{x}) \quad = \quad \sum_{x_i \in \mathbf{x}} x_i^k$$

Lemma 95 (Newton Identities). *Let* $\mathrm{Esym}_k(x_1, \ldots, x_m)$ *and* $\mathrm{Pow}_k(x_1, \ldots, x_m)$ *denote the* elementary symmetric *and* power symmetric *polynomials of degree k respectively, as defined above. Then,*

$$\mathrm{Esym}_k \quad = \quad \frac{1}{k!}\cdot \begin{vmatrix} \mathrm{Pow}_1 & 1 & 0 & \cdots & \\ \mathrm{Pow}_2 & \mathrm{Pow}_1 & 2 & 0 & \cdots \\ \vdots & & \ddots & \ddots & \\ \mathrm{Pow}_{k-1} & \mathrm{Pow}_{k-2} & \cdots & \mathrm{Pow}_1 & k-1 \\ \mathrm{Pow}_k & \mathrm{Pow}_{k-1} & \cdots & \mathrm{Pow}_2 & \mathrm{Pow}_1 \end{vmatrix}.$$

Expanding the determinant on the RHS, we obtain a homogeneous expression

$$\mathrm{Esym}_k(\mathbf{x}) \quad = \quad \sum_{\mathbf{a}\,:\,\sum_i ia_i = k} \alpha_{\mathbf{a}} \cdot (\mathrm{Pow}_1)^{a_1} \ldots (\mathrm{Pow}_k)^{a_k} \tag{7.9}$$

The number of summands bounded by the number of non-negative solutions to $\sum ia_i = k$, which is precisely the number of partitions of k. By the estimates of [HR18], we know that the number of partitions of k is bounded by $2^{\Theta(\sqrt{k})}$. Thus, (7.9) yields a homogeneous depth 4 circuit for $\mathrm{Esym}_k(x_1, \ldots, x_m)$ of size $2^{\Theta(\sqrt{k})} \cdot m$. In fact, the circuit is a homogeneous $\Sigma\Pi\Sigma\wedge$ circuit, i.e. a $\Sigma\Pi\Sigma\Pi$ circuit where the bottom layer of multiplication in fact just raises a single variable to a higher power.

Corollary 96. *Let T be a product of D linear polynomials over n variables, not necessarily homogeneous. Then, the degree d homogeneous component of T, denoted by $\mathrm{Hom}_d(T)$ can be computed by a homogeneous $\Sigma\Pi\Sigma\wedge$ circuit of size $nD \cdot 2^{O(\sqrt{d})}$.*
Hence, if f is a homogeneous degree d polynomial over n variables that is computed by a non-homogeneous depth 3 circuit C of size s, then f can be computed by a homogeneous $\Sigma\Pi\Sigma\wedge\Sigma$ circuit of size $\mathsf{poly}(ns) \cdot 2^{O(\sqrt{d})}$.

To convert the $\Sigma\Pi\Sigma\wedge\Sigma$ circuit to a $\Sigma\wedge\Sigma\wedge\Sigma$ circuit, we could use Fischer's identity again. At first sight, it appears as though this would yield a blow up of 2^d as some of the product gates could have fan-in d. However, notice that the sum is over a_i's satisfying $\sum i \cdot a_i = d$. Hence, there can be at most $O(\sqrt{d})$ of the a_i's that are non-zero. By looking at Fischer's identity applied on $y_1^{a_1} \ldots y_d^{a_d}$ more carefully, we see that it uses at most $(1 + a_1) \ldots (1 + a_d) \leq d^{O(\sqrt{d})}$ distinct linear powers instead of the naïve bound of 2^d. This fact of expressing any degree d monomial over m variables as a $\Sigma\wedge\Sigma$ circuit of size $d^{O(m)}$ was also observed by Ellison [Ell69].
Thus, if f admits a poly-sized depth three circuit, then f also admits a homogeneous $\Sigma\wedge\Sigma\wedge\Sigma$ circuit of size $d^{O(\sqrt{d})} \cdot \mathsf{poly}(n)$. The following lemma summarizes this discussion.

Lemma 97. *Let f be an n-variate degree d polynomial that is computable by depth three circuit of size s over $\mathbb{Q}$. Then, f is equivalently computable by a homogeneous $\Sigma\wedge\Sigma\wedge\Sigma$ circuit of size $d^{O(\sqrt{d})} \cdot \mathsf{poly}(s)$.*
Conversely, if f requires $\Sigma\wedge\Sigma\wedge\Sigma$ circuits of size $n^{\omega(\sqrt{d})}$ over $\mathbb{Q}$ to compute it, then f requires depth three circuits of size $n^{\omega(\sqrt{d})}$.

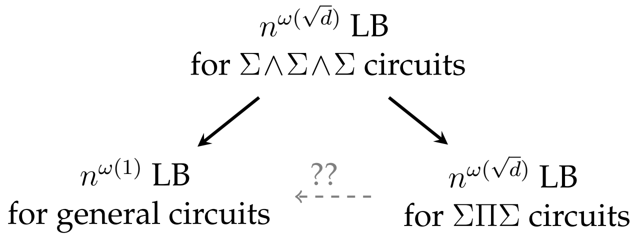
$n^{\omega(\sqrt{d})}$ LB
for $\Sigma\wedge\Sigma\wedge\Sigma$ circuits
$n^{\omega(1)}$ LB
for general circuits
??
$n^{\omega(\sqrt{d})}$ LB
for $\Sigma\Pi\Sigma$ circuits

Figure 7.1: Power of $\Sigma\wedge\Sigma\wedge\Sigma$ ckts.

Completing the picture

We now have an interesting situation (Figure 7.1). On one hand, Corollary 94 states that a lower bound of $n^{\omega(\sqrt{d})}$ for $\Sigma\wedge\Sigma\wedge\Sigma$ circuits would yield a super-polynomial lower bound for general arithmetic circuits. On the other, Lemma 97 states that an $n^{\omega(\sqrt{d})}$ lower bound for $\Sigma\wedge\Sigma\wedge\Sigma$ circuits would yield a lower bound of $n^{\omega(\sqrt{d})}$ for depth three circuits.
Could this just be a coincidence? Or, is it the case that any poly-sized arithmetic circuit can be equivalently expressed as a depth three circuit of size $n^{O(\sqrt{d})}$ over $\mathbb{Q}$? As it turns out, there is indeed a depth reduction to convert any arithmetic circuit to a not-too-large depth three circuit over $\mathbb{Q}$.

To complete the picture, it suffices to show that a $\wedge\Sigma\wedge$ circuit can be expressed as a $\Sigma\Pi\Sigma$ circuit. This would automatically imply a reduction from $\Sigma\wedge\Sigma\wedge\Sigma$ circuits to $\Sigma\Pi\Sigma$ circuits. The last step of the puzzle is the *duality trick* of [Sax08].

Lemma 98 (The Duality Trick [Sax08])**.** *There exists univariate polynomials f_{ij}'s of degree at most b such that*

$$(z_1 + \cdots + z_s)^b \quad = \quad \sum_{i=1}^{sb+1} f_{i1}(z_1)\dots f_{is}(z_s).$$

It is worth noting that the degree of each term on the RHS is sb, whereas the LHS just has degree b. This is the place where non-homogeneity is introduced. Applying the above lemma for a $\wedge\Sigma\wedge$ circuit such as $(y_1^a + \cdots + y_s^a)^b$ gives

$$\begin{aligned}(y_1^a + \cdots + y_s^a)^b &= \sum_{i=1}^{sb+1}\prod_{j=1}^{s} f_{ij}(y_j^a)\\ &= \sum_{i=1}^{sb+1}\prod_{j=1}^{s} \tilde{f}_{ij}(y_j)\end{aligned}$$

where $\tilde{f}_{ij}(y) = f_{ij}(y^a)$. Since each $\tilde{f}_{ij}(y)$ is a univariate polynomial, it can be factorized completely over the $\mathbb{C}$, the field of complex numbers. Hence, if $f_{ij}(y) = \prod_k(y - \zeta_{ijk})$, then we get

$$\begin{aligned}(y_1^a + \cdots + y_s^a)^b &= \sum_{i=1}^{sb+1} \prod_{j=1}^{s} \tilde{f}_{ij}(y_j) \\ &= \sum_{i=1}^{sb+1} \prod_{j=1}^{s} \prod_{k=1}^{b} (y_j - \zeta_{ijk})\end{aligned}$$

which is a depth three circuit! Thus, $(y_1^a + \cdots + y_s^a)$ can be expressed as a depth three circuit of size $\mathsf{poly}(s, a, b)$ over $\mathbb{C}$. With a little more effort, one can construct a depth three circuit over $\mathbb{Q}$ as well. Summarizing this is a lemma, we have the following.

Lemma 99. *Any n-variate degree d polynomial f computed by a homogeneous $\Sigma\wedge\Sigma\wedge\Sigma$ of size s over a characteristic zero field $\mathbb{F}$ can also be computed by a depth three circuit of size $\mathsf{poly}(s, n, d)$ over $\mathbb{F}$.*

Combining with Corollary 94 and Theorem 67, we obtain the main result of [GKKS13b].

Theorem 91 (restated). *Let f be an n-variate degree d polynomial computed by an arithmetic circuit of size s over any characteristic zero field. Then there is a $\Sigma\Pi\Sigma$ circuit of size $s' \leq s^{O(\sqrt{d})}$ that computes f.*

Remark. Note that if we were to start with a degree d polynomial f and apply the above depth reduction, all the linear polynomials that we obtain at bottom are essentially from the application of Fischer's identity on the bottom Π layer of fanin $\sqrt{d}$ of the $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit. Hence, the each of the linear polynomials that appear in the final $\Sigma\Pi\Sigma$ circuit depend on at most $\sqrt{d}$ variables. In other words, the above Theorem yields a reduction to $\Sigma\Pi\Sigma^{[\sqrt{d}]}$ circuits.

7.9.2 Lower bounds for $\Sigma\Pi\Sigma$ circuits with small bottom fan-in

Now let us focus on $\Sigma\Pi\Sigma^{[r]}$ circuits, where all linear polynomials in the circuit depend on at most r variables. The following is the key observation of [KS14a] and can be verified easily.

Observation 100 ([KS14a])**.** *Starting with a $\Sigma\Pi\Sigma^{[r]}$ circuit C of size s computing a homogeneous n-variate polynomial of degree d, the resulting $\Sigma\Pi\Sigma\wedge\Sigma$ circuit C' obtained from Corollary 96 is in fact a $\Sigma\Pi\Sigma\wedge\Sigma^{[r]}$ circuit of size $s' = \mathsf{poly}(ns) \cdot 2^{O(\sqrt{d})}$.*
Thus, by expanding the all powers of linear polynomials computed in the bottom two layers of the $\Sigma\Pi\Sigma\wedge\Sigma$ circuit C', the circuit C' can be rewritten as a homogeneous depth 4 circuit of bottom support bounded by r and size $s'' = s' \cdot d^r$

This observation in combination with Theorem 78 immediately yields the main theorem of [KS14a].

Theorem 101 ([KS14a]). *Over any characteristic zero field $\mathbb{F}$, any $\Sigma\Pi\Sigma^{[r]}$ circuit C computing the polynomial $\mathrm{IMM}_{n,d}$, for suitably chosen parameters n and d with $n = d^{O(1)}$, must have size $s = n^{\Omega(d/r)}$.*

7.9.3 Extensions to low-bottom-fanin depth 5 circuits

[KS14a] also prove lower bounds for depth 5 circuits where the bottom fan-in is bounded. The result proceeds by analysing the random restriction process carefully to decompose any $\Sigma\Pi\Sigma\Pi\Sigma^{[\tau]}$ circuit into a $\Sigma\Pi\Sigma\Pi^{[\sqrt{d}]}$ circuit and another circuit C' such that $\Gamma_{k,\ell}^{[\mathrm{PSD}]}(C') = 0$. We just state their theorem here without proof.

Theorem 102. *Let $\mathbb{F}$ be a field of characteristic zero, and let $0 \leq \mu < 1$. If $\alpha = \frac{2\mu+1}{1-\mu}$ and $\tau = O(N^{\mu})$, then there is a family of n-variate degree d polynomials $\{f_n\}$ in VNP with $n \in [d^{2+\alpha}, 2d^{2+\alpha}]$ such that any homogeneous $\Sigma\Pi\Sigma\Pi\Sigma^{[\tau]}$ circuit computing this polynomial requires size $n^{\Omega(\sqrt{d})}$.*

7.10 Speculation about lower bounds for homogeneous formulas

In this section, we shall look at a possible approach to proving an $n^{\Omega(\log n)}$ lower bound for homogeneous formulas. It is conceivable that variants of the dimension of shifted partial derivatives would be able to give such a lower bound. We will not be presenting any candidate measures, but would instead present a normal form that could perhaps be useful.

7.10.1 A stronger(?) depth reduction for homogeneous formulas

If we were to prove a lower bound for homogeneous formulas via a depth reduction to $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuits, the first step would be to answer the following question:

> Suppose we apply Theorem 67 to a polynomial sized circuit C to obtain a $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit C', and also apply Theorem 67 to a polynomial sized homogeneous formulas $\tilde{C}$ to obtain a $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuit $\tilde{C}'$, how is $\tilde{C}'$ structurally different from C'?

Unless we are able to find a non-trivial structural difference between $\tilde{C}'$ and C', it does not make sense to attempt proving lower bounds for homogeneous formulas via $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuits. The original proof of [Tav13] of Theorem 67 does not seem to suggest any structural difference between the two. However, the alternate proof described in Section 7.6 allows one to understand this difference better.

Recall how the alternate proof proceeded. If g is the polynomial computed by any gate in the circuit C of size s, then g can be written as

$$g \quad = \quad \sum_{i=1}^{\mathsf{poly}(s)} g_{i1} \cdot g_{i2} \cdot g_{i3} \cdot g_{i4} \cdot g_{i5}$$

where $\sum_j \deg(g_{ij}) = \deg(g)$ for all i and $\deg(g_{ij}) \leq \deg(g)/2$. Further g_{ij}s are polynomials computed by the children/grandchildren of g. With an equation as above, we could recursively keep expanding the larger degree g_{ij}'s to eventually get all degrees to be less than $\sqrt{d}$.

For homogeneous formulas, we can start with a slightly more structured equation instead of the one above. This more structured decomposition was first observed by [HY11b].

Lemma 103 ([HY11b])**.** *Let Φ be a homogeneous formula of size s. If f is a polynomial computed at an arbitrary gate of f, then f can be written as*

$$f \quad = \quad \sum_{i=1}^{s} f_{i1} \cdot f_{i2} \cdot \cdots \cdot f_{i\ell} \tag{7.10}$$

where

- *Each f_{ij} is computable by a homogeneous formula of size at most s*
- $\sum_j \deg(f_{ij}) = \deg(f)$ *for all i*
- $\left(\frac{1}{3}\right)^i \deg(f) \leq \deg(f_{ij}) \leq \left(\frac{2}{3}\right)^i \deg(f)$ *for all i, j.*
- $\deg(f_{i\ell}) = 1$ *for all i.*

Proof. Assume that the Φ is a formula of fan-in 2 at each gate. This would only increase the depth by a polynomial factor. Starting from the root, walk down the tree by always picking the child of largest degree until we hit a node v of degree at most $\frac{2\deg(f)}{3}$ for the first time. Since the path always picked the child of largest degree, we must have that

$$\frac{\deg(f)}{3} \quad \leq \quad \deg(v) \quad \leq \quad \frac{2\deg(f)}{3}$$

Let Φ_v denote the sub-formula rooted at v, and let $\Phi_{v=0}$ refer to the formula obtained from Φ by replacing the sub-tree rooted at v by 0. Let s_1 and s_2 be the size of Φ_v and $\Phi_{v=0}$ respectively. (We shall abuse notation and also use Φ_v and $\Phi_{v=0}$ to refer to the polynomial computed by these formulas.) Then,

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

for some polynomial A. Note that homogeneity implies that $\deg(A) + \deg(\Phi_v) = \deg(f)$ and hence $\frac{\deg(f)}{3} \leq \deg(A) \leq \frac{2\deg(f)}{3}$. ($A$ is going to play the role of f_{i1} for some of the i's.)

Observe that the formulas Φ_v and $\Phi_{v=0}$ 'partition' the formula Φ and hence $s_1 + s_2 \leq s$. By induction on these smaller formulas, we can write

$$\begin{aligned}\Phi_v &= \sum_{i=1}^{s_1} g_{i1} \cdots g_{i\ell} \\ \Phi_{v=0} &= \sum_{i=1}^{s_2} h_{i1} \cdots h_{i\ell}\end{aligned}$$

satisfying the necessary conditions. Since $\frac{\deg(f)}{3} \leq \deg(\Phi_v) \leq \frac{2\deg(f)}{3}$, we have that

$$f \quad = \quad \sum_{i=1}^{s_1} A \cdot g_{i1} \cdots g_{i\ell} \quad + \quad \sum_{i=1}^{s_2} h_{i1} \cdots h_{i\ell}$$

satisfies all the degree conditions with A as claimed. To complete the proof, it suffices to show that A can be computed by a homogeneous formula of size s. Indeed, the polynomial A is just the product of all siblings of multiplication gates encountered in the path from v to the root. Since each of the siblings are disjoint sub-formulas of Φ, the polynomial A is computable by a homogeneous formula of size at most s. □

With equation (7.10) instead, we can repeat the strategy we used to prove Theorem 67.

> Start with (7.10) for the root of the homogeneous formula.
>
> For each summand $f_{i1} \dots f_{ir}$ in the RHS, if the largest degree f_{ij} has degree more than $\sqrt{d}$, expand that f_{ij} with the its corresponding representation using Lemma 103.
>
> Repeat this process until all f_{ij}'s on the RHS have degree at most $\sqrt{d}$.

Again, in the expansion of f of degree d via Lemma 103, every term on the LHS has at least two factors of degree more than $d/9$. The same proof would then yield a $\Sigma\Pi\Sigma\Pi^{[\sqrt{d}]}$ circuit of top fan-in at most $s^{O(\sqrt{d})}$. Did we gain anything with this? We certainly did – observe that every expansion via Lemma 103 yields $O(\log d)$ more factors in each term. In the case of Theorem 67, we gained only constantly many factors at each term. Thus, in the resulting depth 4 circuit has the form

$$f \quad = \quad \sum_{i=1}^{s^{O(\sqrt{d})}} Q_{i1} \dots Q_{ir} \quad , \quad \text{where } 1 \leq \deg(Q_{ij}) \leq \sqrt{d}$$

and, most importantly, $r = O(\sqrt{d} \log d)$ as opposed to $O(\sqrt{d})$ in the case of Theorem 67. This seems to be a key structural difference between depth 4 circuits obtained from homogeneous formulas as opposed to depth 4 circuits obtained from general arithmetic circuits! We summarize this below.

Theorem 104. *If f is an n-variate degree d polynomial computed by a homogeneous formula of size s, then there is a homogeneous $\Sigma\Pi^{[O(\sqrt{d}\log d)]}\Sigma\Pi^{[\sqrt{d}]}$ circuit computing f with top fanin at most $s^{O(\sqrt{d})}$.*

Is this useful?

It is not clear if the above structural difference can be exploited to give a complexity measure. But it is very much possible that the some small modification of measure of dimension of shifted partials derivatives might be a measure that works. The reason we believe that is because the results of [GKKS13a, KSS13, KS14c] give explicit n-variate degree d polynomials that admit a top fan-in lower bound of $n^{\Omega(d/t)}$ for depth 4 circuits with *maximum bottom degree* bounded by t.

Question. Could that be changed to give an $n^{\Omega(d/t)}$ lower bound for depth 4 circuits with *average bottom degree* bounded by t?

If this were true, then we obtain an $n^{\Omega(\log d)}$ lower bound for the size of homogeneous formulas computing an explicit n-variate degree d polynomial. Also, we need to keep in mind that any circuit of polynomial size has an equivalent homogeneous formula of size $n^{O(\log d)}$. Hence, if we are hoping to come up with a method that might prove a lower bound for homogeneous formulas but not for general circuits, then method should not be able to yield a lower bound better than $n^{\Omega(\log d)}$. This indeed seems to be the case in this approach. Maybe this is the right depth reduction to work with to prove lower bounds for homogeneous formulas, maybe not. Either way, we shall probably find out soon enough!

7.11 Conclusion

Quite a lot seems to be happening lately in arithmetic circuits. The last few results were on $n^{\Omega(\sqrt{d})}$ lower bounds for homogeneous depth 4, non-homogeneous depth 3 circuits with small bottom fanin, and homogeneous depth 5 with small bottom fanin. Perhaps in the near future, we would be able to obtain $n^{\Omega(\sqrt{d})}$ lower bounds for non-homogeneous depth 3 or homogeneous depth 5 circuits without any bottom fanin restrictions. These are all interesting problems to work on, and should be very much within reach of current techniques. However, it is to be noted that if we wish to separate VP and VNP, we need to break past $n^{\Omega(\sqrt{d})}$. It appears (at least to us) this task would require very different techniques and seems unlikely that small variants of shifted partial derivatives might just get us past $n^{\Omega(\sqrt{d})}$. Nevertheless, Open Problem 1 presents a concrete and extremely simple looking model to work with, for which we need to prove an $n^{\omega(\sqrt{d})}$ lower bound to separate VP and VNP. We conclude by stating the problem again to emphasize the point.

Open Problem. *Find an explicit n-variate degree d polynomial f such that any expression of the form*

$$f \quad = \quad (Q_1)^{\sqrt{d}} + \cdots + (Q_s)^{\sqrt{d}} \quad , \quad \deg(Q_i) \leq \sqrt{d} \text{ for all } i$$

must have $s = n^{\omega(\sqrt{d})}$.

Bibliography

[Agr05] Manindra Agrawal. Proving Lower Bounds Via Pseudo-random Generators. In *Foundations of Software Technology and Theoretical Computer Science Science (FSTTCS)*, pages 92–105, 2005.

[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.

[Alo09] Noga Alon. Perturbed identity matrices have high rank: Proof and applications. *Combinatorics, Probability and Computing*, 18(1-2):3–15, March 2009.

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

[BCS97] Peter Bürgisser, Michael Clausen, and Mohammad A. Shokrollahi. *Algebraic Complexity Theory*, volume 315 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, 1997.

[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.

[Ell69] W.J. Ellison. A 'waring's problem' for homogeneous forms. *Proceedings of the Cambridge Philosophical Society*, 65:663–672, 1969.

[Fis94] I. Fischer. Sums of like powers of multivariate linear forms. *Mathematics Magazine*, 67(1):59–61, 1994.

[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.

[GKKS13a] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Approaching the chasm at depth four. In *Conference on Computational Complexity (CCC)*, 2013.

[GKKS13b] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic Circuits: A Chasm at Depth Three. In *Foundations of Computer Science (FOCS)*, pages 578–587, 2013.

[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.

[HR18] G. H. Hardy and S. Ramanujan. Asymptotic formula in combinatory analysis. *Proceedings of the London Mathematical Society*, s2-17(1):75–115, 1918.

[HY11a] Pavel Hrubeš and Amir Yehudayoff. Arithmetic complexity in ring extensions. *Theory of Computing*, 7(8):119–129, 2011.

[HY11b] Pavel Hrubes and Amir Yehudayoff. Homogeneous formulas and symmetric polynomials. *Computational Complexity*, 20(3):559–578, 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.

[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.

[KI04] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity*, 13(1-2):1–46, 2004.

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

[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.

[KS14a] Neeraj Kayal and Chandan Saha. Lower Bounds for Depth Three Arithmetic Circuits with small bottom fanin. *Electronic Colloquium on Computational Complexity (ECCC)*, 2014.

[KS14b] Neeraj Kayal and Ramprasad Saptharishi. A selection of lower bounds for arithmetic circuits. In Manindra Agrawal and V Arvind, editors, *Perspectives in Computational Complexity*. "Birkhäuser", Basel, 2014.

[KS14c] Mrinal Kumar and Shubhangi Saraf. The limits of depth reduction for arithmetic formulas: it's all about the top fan-in. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 136–145, 2014.

[KS14d] Mrinal Kumar and Shubhangi Saraf. On the power of homogeneous depth 4 arithmetic circuits. In *Foundations of Computer Science (FOCS)*, 2014.

[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.

[Lit50] D.E. Littlewood. *The Theory of Group Characters and Matrix Representations of Groups*. Ams Chelsea Publishing. AMS Chelsea Pub., 2nd edition, 1950.

[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. How to fool people to work on circuit lower bounds. Invited talk at Microsoft Research, 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.

[Rys63] H. J. Ryser. Combinatorial mathematics. *Math. Assoc. of America*, 14, 1963.

[Sax08] Nitin Saxena. Diagonal Circuit Identity Testing and Lower Bounds. In *ICALP (1)*, pages 60–71, 2008.

[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.

[SY10a] 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.

[SY10b] Amir Shpilka and Amir Yehudayoff. Arithmetic Circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5(3-4):207–388, 2010.

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

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

[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.