DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Measures in Algebraic Complexity

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:38811494

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Contents

Chapter 1

Algebraic P and NP

1.1 Motivation

On the long march towards resolving P vs. NP, theoreticians either lower bound the size of Boolean circuits computing SAT or attempt to find efficient algorithms that show otherwise. With super polynomial SAT lower bounds far in the horizon, it is tempting to define an algebraic model of computation, the hope being that the vast array of algebraic tools could resolve fundamental questions in a model that is so rich in structure. The standard model for studying the complexity of computing polynomials is that of arithmetic circuits. Arithmetic circuits are directed graphs with addition and multiplication gates, that take input variables and output a computed polynomial. Over this model, we can define algebraic analogues of complexity classes P and NP, allowing us to pose the VP vs. VNP problem. Although resolving VP vs. VNP does not resolve P vs. NP, progress in the former can inform techniques in the latter and with the hindsight of decades of work it turns out that VP vs. VNP poses immense technical challenges in its own right. Consequently, arithmetic circuit complexity is fertile ground for beautiful results in polynomial identity testing and multilinear formulas, and it offers a large corpus of problems for powerful lower bounding techniques in Geometric Complexity Theory and Sum-of-Squares.

A survey of these results is beyond the scope of this thesis. So instead, we are focusing on a particular line of work revolving around a family of lower bounding techniques that study the space of the linear span of the partial derivatives of certain families of polynomials. We refer to this approach as "the method of lower bounding by partial derivative measure". We present three variants of this technique on various restricted classes of circuits. We present lower bounding by the partial derivative measure, the

shifted partial derivative measure, and the projected shifted derivative measure detailed in Chapters 2,3, and 4 respectively. Our theoretical contribution is in Chapter 3 , which is our main technical result in depth three circuits computing a family in VP.
If one chooses to skim, Chapter 3 is where the original content is.

With this landscape in view, we will begin Chapter 1 by giving the basic definitions of arithmetic circuit complexity, define VP and VNP, and prove basic completeness and hardness results for this model of computation. In addition, we will lay the foundations for depth reduction and homogenization which justify the study of restricted circuit models.

1.1.1 Introduction

An arithmetic circuit computes a polynomial by taking in input variables $x_1, x_2, .., x_n$ and computing a polynomial with the arithmetic operations $+, \times$. The complexity measure of concern is the size (edges) of the circuit representing the number of operations required to compute a polynomial.

Arithmetic circuits are imbued with a great deal of structure. In particular, when studying arithmetic circuits we want to compute a specific representation of a polynomial (a syntactic focus) whereas Boolean circuits aim to compute any representation of a function (a semantic focus). Even after decades of work, we do not know how to deterministically and efficiently determine whether a given arithmetic circuit computes the zero polynomial, nor how to efficiently reconstruct a circuit using only queries to the polynomial it computes.

1.1.2 Definitions and Preliminary Ideas

For a thorough survey on preliminary ideas in arithmetic circuit complexity see [SY10]

Definition 1.1. An Arithmetic circuit Φ over the field $\mathbb{F}$ and the set of variables $X = \{x_1, x_2, ..., x_n\}$ is a directed acyclic graph. The vertices of Φ are gates. Every gate in Φ of in-degree 0 is labeled by either a variable from X or a field element in $\mathbb{F}$. Every other gate in Φ is either a $+$ (sum gate) or $\times$ (product gate) with in-degree 2. An arithmetic formula is an arithmetic circuit that is also a directed tree from leaves to root.

Gates of in-degree 0 are input gates, and gates of out-degree 0 are output gates. The size of Φ is the number of edges in Φ. However, in the case of bounded depth circuits there is no restriction on the fan-in of gates unless explicitly stated.

Arithmetic circuits compute polynomials by taking the input variables and summing and multiplying monomials and constants until an output gate produces the polynomial in question. In this manner every polynomial $f \in \mathbb{F}[X]$ can be computed by an arithmetic circuit and by arithmetic formula. Then the natural complexity measure that arises is the number of edges or gates required for computation. At least in the bounded depth case, edges is popular because the unbounded fan-in is reasonably powerful.

Finally, it is worth remarking that arithmetic circuits compute formal polynomials in $\mathbb{F}[X]$ but not functions from $\mathbb{F}^{|X|} \to \mathbb{F}$. In this way, arithmetic circuits differ from Boolean circuits. For instance, $2x$ is the zero function over the field of two elements. The arithmetic circuit must produce the polynomial $2x$ as the output of its computation, which is necessarily the zero function, but it is not sufficient for the circuit to compute the zero function. Again, the focus is on the syntactic rather than semantic content.

This distinction can be especially confusing when discussing the fields over which certain lower bounds hold. The choice of field is for ease of analysis, but does not change the fact that arithmetic circuits compute formal polynomials.

1.1.3 VP vs. VNP

Arithmetic complexity begins with defining algebraic analogues of P, NP, and completeness results in the seminal work [Val79]. A natural starting point is then the algebraic analogue of P, the class VP of "polynomially bounded" circuits.

Definition 1.2. (VP) A family of polynomials $\{f_n\}$ over $\mathbb{F}$ is in $VP_{\mathbb{F}}$ if there exists some polynomial $t : \mathbb{N} \to \mathbb{N}$ such that for every n, both the number of variables in f_n and the degree of f_n are at most $t(n)$, and there is an arithmetic circuit of size at most $t(n)$ computing f_n.

Note that we always work with a family of polynomials, but sometimes we will work with a "polynomial" and understand that the index n defines the entire family. Note that we did not require the circuit computing f_n to have polynomial degree, but this holds without loss of generality [SY10]. Also note that even a size $O(n)$ circuit can compute a polynomial with exponential degree. The polynomial degree bound is motivated by computations over rational numbers. An exponential degree polynomial is not necessarily representable with an efficient number of bits.

In particular, the determinant polynomial is in VP. This can be demonstrated by Gaussian elimination which needs division gates. Removing the division gates will be discussed later in the chapter. We define the determinant as follows.

$$DET_n(X) = \sum_{\sigma \in S_n} sgn(\sigma) \prod_{i=1}^{n} x_{i,\sigma(i)} \tag{1.1}$$

for $X = (x_{i,j})$ an $n \times n$ matrix, S_n the set of permutations of n elements and $sgn(\sigma)$ is the signatures of the permutation σ.

VNP is then the algebraic analogue of NP.

Definition 1.3. (VNP) A family of polynomials f_n over $\mathbb{F}$ is p-definable is there exists two polynomially bounded functions $t, k : \mathbb{N} \to \mathbb{N}$ and a family $\{g_n\}$ in $VP_{\mathbb{F}}$ such that for every n

$$f_n(x_1, ..., x_{k(n)}) = \sum_{w \in \{0,1\}^{t(n)}} g_{t(n)}(x_1, ..., x_{k(n)}, w_1, ..., w_{t(n)})$$

One way to think of VNP is to regard the w as witnesses and the summing as an analog of searching for witnesses in NP. We've also modified the existential quantifier in NP to the addition operation, which makes VNP an analogue of P as well. It is also apparent that the sum of a single polynomially bounded circuit gives us $VP \subseteq VNP$. A famous example of a polynomial in VNP is the permanent.

$$PERM_n(X) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} x_{i,\sigma(i)} \tag{1.2}$$

It is quite remarkable that removing the signature of permutations can make a polynomial so much harder to compute, but this is consistent with our experience of permanent and determinant computation outside of the setting of arithmetic circuits. The smallest known circuit computing the permanent is that given by Ryser's formula.

Lemma 1.4. *[Rys63] For every $n \in \mathbb{N}$, $PERM_n(X) = \sum_{T \in [n]} (-1)^{n-|T|} \prod_{i=1}^{n} \sum_{j \in T} x_{i,j}$*

We now pose the algebraic analogue of the P vs. NP question.

Valiant's Hypothesis I: VP $\neq$ VNP

Critically, we also need notions of reduction and examples of completeness.

Definition 1.5. A polynomial $f(x_1, x_2, ..., x_n)$ over $\mathbb{F}$ is called a projection of a polynomial $g(y_1, ..., y_m)$ over $\mathbb{F}$ if there exists an assignment $p \in (\{x_1, ..., x_n\} \cup \mathbb{F})^m$ such that $f(x_1, ..., x_n) \equiv g(p_1, p_2, ..., p_m)$. In other words, f can be derived from g by simple

substitutions. This can be extended to projections between families of polynomials. The family $\{f_n\}$ is a p-projection of the family $\{g_n\}$ if there exists a polynomially bounded $t : \mathbb{N} \to \mathbb{N}$ such that for every n, f_n is a projection of $g_{t(n)}$

Valiant showed that VP and VNP are closed under projections, and that the permanent is complete for the class VNP.

Theorem 1.6. *[Val79] For any field $\mathbb{F}$ such that $char(\mathbb{F}) \neq 2$, the family $\{PERM_n\}$ is VNP-complete.*

Henceforth, when we say "lower bound" we mean for a family of polynomials in VNP unless specified otherwise. Similarly the determinant is VQP-complete where VQP is the set of families of polynomials with variables, degree, and circuits bounded quasi-polynomially ($2^{\text{polylog}(n)}$)

Theorem 1.7. *[Val79] The family $\{DET_n\}$ is VQP-complete. It is also VP-complete with respect to quasi-polynomial projections.*

A quasi-polynomial projection is a generalization of polynomial projection where we replace $t(n) : \mathbb{N} \to \mathbb{N}$ to be a quasi-polynomially bounded function.

We present a few more results that motivate Valiant's extended hypothesis.

Theorem 1.8. *[VSBR83] Let f be a degree r polynomial computed by a size s circuit. Then f can be computed by a circuit of size $poly(r, s)$ and depth $O(\log r(\log r + \log s))$.*

If we define VNC^k to be polynomials with polynomially bounded degree and variables of depth $O(\log^k n)$, then the above theorem shows $VP = VNC^2$, which stands in stark contrast to the Boolean world where it is conjectured that $P \neq NC$. Furthermore, this has implications for arithmetic formulas. In particular, every polynomial in VP has a formula of quasi-polynomial size.

Theorem 1.9. *[Val79] For any polynomial f in $\mathbb{F}[X]$ that can be computed by a formula of size s over F, there is a matrix A of dimensions $(s+1) \times (s+1)$ whose entries are in $X \cup \mathbb{F}$ such that $DET(A) = f$*

We are now ready to state Valiant's extended hypothesis that the permanent does not belong to VQP.

Valiant's Extended Hypothesis: VNP $\not\subset$ VQP

To prove the extended hypothesis it suffices to show that one cannot represent $PERM_n$ as the determinant of a matrix of dimension quasi-polynomial in n (i.e DET is not

a quasi-polynomial projection of PERM). Rather incredibly, this statement makes no mention of circuits. But alas, quasi-polynomial lower bounds are far away and our best lower bound is quadratic.

Theorem 1.10. *[MR04] [CCL08] Let F be a field of characteristic different than two and let $X = (x_{i,j})$ be a matrix of variables. Then, any matrix A whose entries are linear functions in $\{x_{i,j}\}$ over $\mathbb{F}$ such that $DET(A) = PERM_n(X)$ must be of dimension at least $n^2/2$*

The proof idea is to compute the rank of the Hessian matrix of $PERM_n(X)$ and $DET(A)$. The rank for $PERM_n(X)$ is roughly quadratic and the rank for $DET(A)$ is $dim(A)$.

To conclude this expository section, we restate the fundamental thrust of arithmetic circuit complexity, which is to improve upon the lower bounds on the dimension of a matrix A with entries that are linear functions in $\{x_{i,j}\}$ such that $DET(A) = PERM_n(X)$.

1.1.4 Circuit Classes

As it is difficult to prove strong bounds for general circuits, we often work with circuits of bounded depth, multilinear circuits, homogeneous circuits, non-commutative circuits, etc. We will run through the more popular circuit classes.

The bounded depth circuit has bounded depth and unbounded fan-in. The two most popular are the depth-3 or $\sum\prod\sum$ circuits and the depth-4 or $\sum\prod\sum\prod$ circuits. In this work we also discuss recent results related to depth-5 circuits $\sum\prod\sum\prod\sum$ which is a far less common class of circuits to work with.

circuits have an addition gate as the output, with a middle layer of multiplication gates, and then a level of addition gates at the bottom. A circuit with s multiplication gates computes polynomials $\sum_{i=1}^{s}\prod_{j=1}^{d_i}\ell_{i,j}(x_1,...,x_n)$ where $\ell_{i,j}$ are linear functions. The significance of depth-3 circuits, is that we do not have strong lower bounds over fields of characteristic zero. In fact, the best lower bound for polynomials in VNP are almost cubic $\tilde{\Omega}(n^3)$. As it turns out, the best lower bounds in VP are also $\tilde{\Omega}(n^3)$ which is the primary result in this work. Also, for a restricted class of circuits, strong lower bounds imply super-polynomial lower bounds on the formula complexity of the permanent.

A common technique in lower bounding depth-3 and depth-5 circuits is to expand or project to depth-4 circuits which are notable for being the first depth for which we have no strong lower bounds over fields $\mathbb{F} \neq 2$. Furthermore, exponential lower bounds for circuits are equivalent to exponential lower bounds for general depth circuits. That is

to say, a polynomial can be computed by a sub-exponential general depth circuit if and only if it can be computed by a sub-exponential circuit, a result in [AV08].

Only a few papers have ever explored depth-5, for which there are superlinear lower bounds for large enough fields [Raz13a].

1.2 Structural Results

Depth reduction and strong lower bounds for constant depth circuits are the two ends of the machinery that works to separate VP and VNP. The majority of this work discusses lower bounding techniques over various circuit classes. However, this can feel vacuous without any depth reduction results to ground our concern for $\sum\prod\sum$, $\sum\prod\sum\prod$, and $\sum\prod\sum\prod\sum$ circuits which at first glance appear to be fairly artificial. After all, constant depth circuits have unbounded fan-in with alternating layers. In any case, it is important to know just how strong our lower bounds have to be to separate VP from VNP. The main result we present is the reduction of [VSBR83] for homogeneous polynomials computed by homogeneous circuits, which has been strengthened and extended by [AV08] [Koi10a]. We follow many of the proofs in [SY10] but change the order of presentation and omit lengthy calculations. We begin our discussion with homogeneity and then present a few results in depth reduction.

1.2.1 Homogenization

In this work we will explore homogeneous constant depth circuits [KS15]. Homogeneous circuits are a popular circuit model, not only because it is a naturally arising class of polynomials, but also that there exists strong homogenization results that justify the study of homogenized circuits.

In the study of homogeneous circuits, it turns out that we can decompose any computation to its homogeneous parts without increasing the size by too much. We follow the proof in [Str73]

Theorem 1.11. *[Str73]*

If f has an arithmetic circuit Φ of size s, then for every $r \in N$, there is a homogeneous circuit Ψ of size at most $O(r^2 s)$ computing $H_0[f], H_1[f], ..., H_r[f]$ where given a polynomial f, we denote by $H_i(f)$ its homogeneous part of degree i.

This transformation allows us to assume that circuits for families in VP have polynomial degrees as well, that all their intermediate computations are also low degree polynomials. We sketch the proof below.

Proof. We describe how to construct Ψ. For every gate v in Φ, we define $r+1$ gates in Ψ which we denote $(v,0),...,(v,r)$ so that (v,i) computes $H_i(\Phi_v)$. We construct Ψ inductively as follows. If v is an input gate, we can define (v,i) as an input gate with the appropriate properties. If $\Phi_v = \Phi_u + \Phi_w$ we define $\Psi_{(v,i)} = \Psi_{(u,i)} + \Psi_{(w,i)}$ for all i. If $\Phi_v = \Phi_u \times \Phi_w$ define $\Psi_{(v,i)} = \sum_{j=0}^{i} \Psi_{(u,j)} \times \Psi_{(w,i-j)}$. By inductive hypothesis Ψ computes what we desired. Every gate in Φ corresponds to at most $O((r+1)^2)$ gates in Ψ and so $|\Psi| = O(r^2 s)$ □

We have shown how to take a general circuit and transforms it into a homogeneous one that computes the same polynomial without a heavy penalty in size. However, the same can't be said of the change in depth. Indeed, the depth may be increased by a factor that is logarithmic in the degree.

1.2.2 Division Gates

Before moving on to depth reduction, a glaring feature of arithmetic circuits is that they do not include division gates. Would introducing $\div$ as an operator affect the power of arithmetic circuits. This is a natural question. If we have division gates than each gate computes a rational function instead of a polynomial. Thus we must stipulate that the circuit does not divide by the zero polynomial.

The first answer to this line of questioning was provided by Strassen [Str73] who showed that over infinite fields divisions do not add power to the model. The approach works when the field contains nonzeros for the polynomial that we divide by [BQH82]. This result was then extended to finite fields [HY09]. We will find that coping with division gates will require us to use the homogenization procedure described above.

Theorem 1.12. *If a polynomial $f \in \mathbb{F}[x_1,...,x_n]$ of degree r can be computed by an arithmetic circuit Φ of size s using the operations $+,-,\times,\div$ then there is a circuit Ψ of size $poly(s,r,n)$ that uses only the arithmetic operations $+,\times$ and computes f*

Proof. We will sketch the proof for large fields. First we show that if we slightly increasing the size of the circuit we can assume that the only division gates appear at the top of the circuit. In other words, we can assume that $f = h \div g$ where h and g are computed by the two children of the output gate, and the output gate is the only gate labelled by $\div$.

To see this, duplicate each gate v to two gates (v,numerator) and (v,denominator) such that (v,numerator) computes the numerator of the rational function that v computes and (v,denominator) computes the denominator of the function computed by v. This can be done using the following identities.

1. $h_1/g_1 + h_2/g_2 = (h_1g_2 + h_2g_1)/(g_1g_2)$
2. $h_1/g_1 \times h_2/g_2 = (h_1h_2)/(g_1g_2)$
3. $h_1/g_1 \div h_2/g_2 = (h_1g_2)/(h_2g_1)$

Finally, we show how to eliminate the division only division gate. If we express $f = h/g$ and notice that the field is large enough, $g(\alpha_1, ..., \alpha_n) \neq 0$ for some $\alpha_1, ..., \alpha_n \in \mathbb{F}$. If we translate the input and multiply by a field element, with a slight increase in size, we can assume that $g(0, 0, .., 0) = 1$. In this case,

$$f = \frac{h}{g} = \frac{h}{1-(1-g)} = \sum_{j=0}^{\infty} h(1-g)^j \tag{1.3}$$

where the last equality implies that every homogeneous part of f is the same as the homogeneous part of the right hand side. As $1 - g$ has no constants, the minimal degree of a monomial $(1 - g)^j$ is j. For every $i \in \{0, ..., r\}$

$$H_i(f) = \sum_{j=0}^{\infty} H_i(h(1-g)^j) = \sum_{j=0}^{i} H_i(h(1-g)^j) \tag{1.4}$$

So we can efficiently compute the homogeneous parts of f, using the homogeneous parts of the polynomials $\{h(1 - g)^j : j \in [r]\}$, each of which has a circuit of size $poly(r, s)$. Then we use 1.11 to complete the proof.

For small fields, we need to "simulate" a large field. We must do so without increasing the size by much. To make a field $\mathbb{F}$ larger we consider a field extension $\mathbb{E}$ of $\mathbb{F}$ that is large enough for our purposes, and we need $\mathbb{E}$ to be of size $poly(r, n)$. The point is that we can think of the elements of $\mathbb{E}$ as vectors in $\mathbb{F}$, and can therefore simulate the arithmetic operations over $\mathbb{E}$ by operations over $\mathbb{F}$. If $\bar{a} = \{a_1, ..., a_k\}$ and $\bar{b} = (b_1, ..., b_k)$ are two elements of $\mathbb{E} = \mathbb{F}^k$, where k is the degree of $\mathbb{E}$. Then the sum of $\bar{a}$ and $\bar{b}$ over $\mathbb{E}$ is just the entry wise sum of $\bar{a}$ and $\bar{b}$ as vectors in $\mathbb{F}^k$. The product of $\bar{a}$ and $\bar{b}$ over $\mathbb{E}$ can be defined coordinate-wise as $(\bar{a} \times \bar{b})_i = \lambda_i(\bar{a}, \bar{b})$ for $i \in [k]$ where λ_i is a fixed bilinear form (a $k \times k$ matrix with entries in $\mathbb{F}$). To simulate a circuit over $\mathbb{E}$ by a circuit over $\mathbb{F}$ we duplicate each gate k times and simulate the arithmetic operations over $\mathbb{E}$ by the arithmetic operations over $\mathbb{F}$. □

1.2.3 Depth Reduction

We present the following facts about partial derivatives [SY10].

Lemma 1.13. *Let v,w be two gates of a homogeneous circuit Φ. Denote with f_v and f_w the polynomial computed by v and w respectively.*

1. *Either $\partial_w f_v$ is zero or $\partial_w f_v$ is a homogeneous polynomial of total degree $deg(v) - deg(w)$*
2. *Assume that v is a product gate with children v_1 and v_2 such that $deg(v_1) \geq deg(v_2)$. If $deg(w) > deg(v)/2$ then $\partial_w f_v = f_{v_2} \cdot \partial_w f_{v_1}$.*
3. *Assume that v is a sum gate with children v_1, v_2. Then $\partial_w f_v = \partial_w f_{v_1} + \partial_w f_{v_w}$*

Proof. We outline the proof.

1. For the first point, notice that f_v and f_w are both homogeneous. As $f_v = f_{v,w}|_{y=f_w}$, and by definition of $\partial_w f_v$, it follows that if $\partial_w f_v$ is nonzero then it is of degree $deg(v) - deg(w)$.
2. For the second point, since v is a product gate, $deg(v) = deg(v_1) + deg(v_2)$. By assumption we have $deg(v_2) < deg(w)$. Since the circuit is homogeneous, the gate w is not in the sub-circuit rooted at v_2. Hence $\partial_w f_{v_2} = 0$.
3. The third point follows from the definition of $\partial_w f_v$

□

Now we use these facts about partial derivatives to perform circuit depth reductions. Namely, without loss of generality, we can assume that polynomial size circuits of polynomial degree are of poly-logarithmic depth [VSBR83]. This directly implies that the determinant of an $n \times n$ matrix can be computed by a polynomial size circuit of depth $O(\log^2 n)$.

Theorem 1.14. *[VSBR83] For every homogeneous degree r polynomial f computed by a circuit Φ of size s, there is a homogeneous circuit Ψ of size $poly(r, s)$ computing f with the following additional structure.*

1. *The circuit Ψ has alternating levels of sum and product gates.*
2. *Each product gate v in Ψ computes the product of five polynomials, each of degree at most $2deg(v)/3$.*

3. *Sum gates have arbitrary fan-in. In particular, the number of levels in* Ψ *is* $O(\log r)$.

This theorem implies that if f is a polynomial of degree r computed by size s circuit, then f can be computed by a size $\text{poly}(s, r)$ circuit with fan-in at most two of depth $O(\log r(\log r + \log s))$.

To proceed we'll have to define some notation. For an integer $m \in \mathbb{N}$, denote by G_m the set of multiplication gates t in Φ with children t_1 and t_2 such that $m < deg(t)$ and $deg(t_1)$, $deg(t_2) \leq m$

Before we can prove theorem 1.14 we state the following two lemmas without proof.

Lemma 1.15. *Let* $m > 0$ *be an integer, and let* v, w *be two gates so that* $deg(w) \leq m < deg(v) < 2deg(w)$. *Then,*

$$f_v = \sum_{t \in G_m} f_t \times \partial_t f_v \text{ and } \partial_w f_v = \sum_{t \in G_m} \partial_w f_t \times \partial_t f_v \tag{1.5}$$

We can now prove theorem 1.14

Proof. First, assume without loss of generality that $s \geq n$. Second, we can assume without loss of generality that the circuit computing f, Φ, is a homogeneous arithmetic circuit of size $s' = O(r^2 s)$. To prove the theorem we construct Ψ. The construction is done in steps where at the i'th step we do the following:

1. Compute all polynomials f_v, for gates v in Φ such that $2^{i-1} < deg(v) \leq 2^i$
2. After we finish the first part, we compute all polynomials $\partial_w f_v$, for all appropriate gates v, w so that

$$2^{i-1} < deg(v) - deg(w) \leq 2^i \text{ and } deg(v) < 2deg(w) \tag{1.6}$$

We show that we can perform the first part of the i'th step by adding a layer consisting of $O(s'^2)$ product gates, of fan-in at most 3, and a layer of $O(s')$ sum gates, each of fan-in $O(s')$. Similarly, for the second part we need to add a layer of $O(s'^3)$ product gates, of fan-in at most 3, and a layer of $O(s'^2)$ sum gates, each of fan-in $O(s')$ to the circuit that was computed in the first part. This procedure increases the size by $O(s'^3)$ and the depth by a constant factor.

Now we construct Ψ. Gates in Ψ are denoted by v, t, and w. For a gate v, we denote by v' the gate in Ψ computing f_v. For two appropriate gates, v, w, we denote by (w, v) the gate in Ψ computing $\partial_w f_v$. We then proceed inductively.

For every gate v in Φ satisfying $deg(v) \leq 1$, f_v is linear. So, since $s' \geq n$, we can compute f_v with a linear arithmetic circuit of size $O(s')$ and depth $O(1)$. For every two gates v and w in Φ such that $deg(v) - deg(w) \leq 1$, lemma 1.13 implies that $\partial_w f_v$ is linear. Therefore, we can compute $\partial_w f_v$ with a linear circuit of size $O(s')$ and constant depth.

We repeat this process i times. For the $i+1$ step we first compute the $f'_v s$. Let v be a gate of degree $2^i < deg(v) \leq 2^{i+1}$, and denote $m = 2^i$.

Now we know that if a gate t is not in Φ_v, then $\partial_t f_v = 0$. So lemma 1.15 gives us

$$f_v = \sum_{t \in T_v} f_t \times \partial_t f_v = \sum_{t \in T_v} f_{t_1} \times f_{t_2} \times \partial_t f_v \tag{1.7}$$

where T_v is the set of gates $t \in G_m$, with children t_1 and t_2, such that t is in Φ_v. We can see that $m < deg(t) \leq 2m$ and $deg(t_1), deg(t_2) \leq m$. This implies that $deg(v) - deg(t) \leq 2^{i+1} - 2^i = 2^i$ and $deg(v) \leq 2^{i+1} < 2deg(t)$. So we find that f_{t_1}, f_{t_2}, and $\partial_t f_v$ have been computed. Then we compute f_v using equation 1.7 by adding one layer of $O(s')$ product gates with fan-in 3 and a sum gate with fan-in $O(s')$.

Now to compute the $\partial_w f_v$'s, let v and w be two gates so that

$$2^i < deg(v) - deg(w) \leq 2^{i+1} \text{ and } deg(v) < 2deg(w) \tag{1.8}$$

Now, we know $m = 2^i + deg(w)$. Therefore, $deg(w) \leq m < deg(v) < 2deg(w)$. We also know that if a gate t is not in Φ_v then $\partial_t f_v = 0$. Also by lemma 1.13, if a gate t admits $deg(t) > deg(v)$, then $\partial_t f_v = 0$. So by lemma 1.15

$$\partial_w f_v = \sum_{t \in T_v} \partial_w f_t \cdot \partial_t f_v \tag{1.9}$$

For $t \in T_v$, we have $deg(t) \leq deg(v) < 2deg(w)$. Denote the children of $t \in T$ by t_1 and t_2, and assume without loss of generality that w is in Φ_{t_1}. We notice that $deg(w) \leq deg(t_1)$ and $deg(t_1) \geq deg(t_2)$. So using the second item of lemma 1.13 we have the following equation

$$\partial_w f_v = \sum_{t \in T_v} f_{t_2} \cdot \partial_w f_{t_1} \cdot \partial_t f_v \tag{1.10}$$

We now show that all the polynomials $f_{t_2}, \partial_w f_{t_1}$, and $\partial_t f_v$ were already computed, including the first part of the $i+1$ step described above.

Since $deg(v) \leq 2^{i+1} + deg(w) \leq 2^{i+1} + deg(t_1) = 2^{i+1} + deg(t) - deg(t_2)$, it holds that $deg(t_2) \leq 2^{i+1} + deg(t) - deg(v) \leq 2^{i+1}$. So we conclude that f_{t_2} has been computed.

Likewise, $deg(t_1) \leq m = 2^i + deg(w)$, so we have $deg(t_1) - deg(w) \leq 2^i$. Since $deg(t_1) \leq deg(t) \leq deg(v) < 2deg(w)$, the polynomial $\partial_w f_{t_1}$ has also been computed.

Finally, $deg(t) > m = 2^i + deg(2)$, we get $deg(v) - deg(t) < deg(v) - 2^i - deg(w) \leq 2^{i+1} - 2^i = 2^i$. Since $deg(v) \leq 2^{i+1} + deg(w) \leq 2deg(t)$, the polynomial $\partial_t f_v$ has been computed.

Now using equation 1.10 we can compute $\partial_w f_v$ by adding $O(s')$ product gates of fan-in at most 3 and a sum gate of fan-in $O(s')$.

To conclude the proof of theorem 1.14, we address the problem that not every product gate has children of sufficiently low degree. Completing the proof involves increasing the fan-in of each multiplication gate to 5 by replacing f_{t_1} in 1.7 with the sum of products of lower degree polynomials. This concludes the proof. □

Building on this foundational result, [AV08] proved that exponential lower bounds for $\sum\prod\sum\prod$ circuits computing polynomials of degree $\Omega(n)$ imply exponential lower bounds for the size of general arithmetic circuits. This is presented as a corollary to the following theorem that we state and prove as a sketch.

Theorem 1.16. *Let $f(x_1, x_2, ..., x_n)$ be a polynomial of degree $r = O(n)$ over $\mathbb{F}$. If there exists a circuit of size $s = 2^{o(r+r\log(n/r))}$ for f, then there exists a $\sum\prod\sum\prod$ circuit of size $2^{o(r+r\log(n/r))}$ for f as well. Furthermore, the fan-in of the top layer of product gates is bounded by $\ell(n)$, where ℓ is any sufficiently slowly growing function tending to infinity with n, and the fan-in of the bottom layer of product gates is bounded by $o(r)$.*

Proof. Break the circuits ensured by Theorem 1.14 into two parts: first of the topmost t levels of multiplication gates together with the addition gates above them, and the second is the rest of the circuit. The polynomial computed by the first part is a polynomial of degree at most 5^t in $poly(r, s)$ variables, and therefore can be computed by a depth-2 circuit of size $O(5^t\binom{poly(r,s)+5^t}{5^t})$. The second circuit computes a polynomial of degree at most $D = deg(f)/(3/2)^t$ and therefore has a depth-2 circuit of size on the order of $\binom{n+D}{D}$. When we compose the two circuit parts we obtain depth-4 circuit of size $O(5^t\binom{poly(r,s)+5^t}{5^t}\binom{n+D}{D})$ computing f. We can complete the proof by optimizing over t. □

An extension by [Koi10a] relates depth reduction back to VP and VNP.

Theorem 1.17. *Let f be an n-variate polynomial of degree r that can be computed by a polynomial size arithmetic circuit. Then f can be computed by a $\sum\prod\sum\prod$ circuit with $n^{O(\log r)}$ addition gates and $n^{O(\sqrt{r}\log r)}$ multiplication gates. Furthermore, multiplication gates have fan-in at most $\sqrt{3r} + 1$.*

Which yields the following corollary

Corollary 1.18. *A $2^{n^{1/2+\epsilon}}$ lower bound on the depth-4 complexity of the permanent of n by n matrices is sufficient for separating VP from VNP.*

Our best lower bounds over $char(\mathbb{F}) \neq 2$ is n^{1+c} for some constant $c > 0$ [Raz10] which gives rise to the following open problem.

Open Problem: Prove super-polynomial lower bounds for depth-4 circuits of fields of $char(\mathbb{F}) \neq 2$

After establishing depth reduction results, we can see how general depth-4 circuits can be, which buys us a significant amount of structure. Thus, depth reduction lies at the core of the justification for the constant depth circuit bounds we will be presenting for the rest of this work.

The Partial Derivative Measure

1.2.4 Introduction

We present some of the seminal lower bounds in arithmetic circuit complexity. One the way we provide motivation for the technique of lower bounding by partial derivative measure. We flesh the technique out for the quadratic lower bound on determinants for depth-3 circuits by [SW02] which held for 15 years until [KST16].

1.2.5 A Hard Polynomial over Finite Fields

Since it is so difficult to prove super-polynomial lower bounds for polynomials in VNP, it is valuable to see what sort of polynomial is extremely hard to compute. It turns out the MOD_q function has exponential lower bounds for $\sum\prod\sum$ circuits over small fields. MOD_q is defined as $MOD_q(x_1, ..., x_n) = 1$ iff $\sum_i x_i = 0$ mod q, thinking of x_i as an integer.

Theorem 1.19. *[GR00a] Let $p \neq q$ be two primes. Then, every $\sum\prod\sum$ circuit computing the n-varaite MOD_q function over $\mathbb{F}_p$ must be of size at least $2^{\Omega(n)}$*

Proof. Let Φ be a $\sum\prod\sum$ circuit over the field with p elements $\mathbb{F}_p$. The general idea is to partition the product gates of Φ into two sets, and to find the weakness of each of these sets. We partition the gates of Φ according to their rank: Every product gate v in Φ of degree d_v multiplies d_v linear functions. Define the rank of v as the dimension of the span of these d_v linear functions. Partition the product gates of Φ into V_{HIGH} and V_{LOW}, where V_{HIGH} is the set of product gates of rank at least R and V_{LOW} to contain the low rank gates.

Now we discuss the weakness of $\sum\prod\sum$ circuits over finite fields. If we substitute random field elements as inputs, each gate in V_{HIGH} is nonzero with probability at most $(1 - 1/p)^R = 2^{-c_p R}$. In addition, each product gate in V_{LOW} can be represented by a low degree polynomial of degree at most pR. Then if we use union bound we find that

a $\sum\prod\sum$ circuit Φ of size s can be approximated by a low degree polynomial. Indeed, there exists a polynomial $g \in \mathbb{F}_p[x_1, ..., x_n]$ of degree at most pR so that

$$Pr_a[\Phi(a) \neq g(a)] \leq 2^{-c_p R} s \tag{1.11}$$

where a is uniform in $\mathbb{F}_p^n$.

The MOD_q function cannot be well approximated by low degree polynomials over the Boolean cube. This property should complete the proof, provided that the result over the Boolean cube can imply the needed property over $\mathbb{F}_p$. The final peice of the proof is by Grigoriev and Raborov who used the observation that the uniform distribution over $\mathbb{F}_p^n$ is a convex combination of distributions that are uniform over translates of $\mathbb{F}_2^n$. □

1.2.6 Motivating Example

The following example proven in [NW96] motivates the use of complexity measures in proving lower bounds on arithmetic circuits, variations of which are the focus of this work.

Theorem 1.20. *Any homogeneous depth 3 circuit computing the 2d'th symmetric polynomial on n variables over a field of characteristic zero requires* $\Omega((n/4d)^d)$ *size.*

The following is both a proof and a general discussion of the technique for proving theorem 1.20

Proof. Let F be a field of characteristic 0. We consider polynomials in n variables X. For any set of polynomials $V \subseteq \mathbb{F}[X]$ we use $dim(V)$ to denote the dimension of the linear span of V i.e maximum number of linearly independent polynomials over $\mathbb{F}$ in V.

Let f be a polynomial. We let $\partial(f)$ denote the set of all partial derivatives of all orders of f. To belabor this point, a single monomial that is the product of k variables will have 2^k different partial derivatives. The linearity, sum and product formulae for partial derivatives upper bound the ability of the different circuit operations to increase the dimension of the set.

Proposition 1.21. *For every* $f_1, f_2, ..., f_r \in \mathbb{F}[X]$ *and* $\alpha \in \mathbb{F}$, $\alpha \neq 0$ *we have:*

- $dim(\partial(\alpha(f_1))) = dim(\partial(f_1))$
- $dim(\partial(\sum_i(f_i))) = \sum_i dim(\partial(f_i))$
- $dim(\partial \prod_i f_i) = \prod_i dim(\partial(f_i))$

This proposition bounds the dimension of the output of depth-3 circuits.

Lemma 1.22. *Let f be computed by a depth-3 circuits with fan-in s to the top gate, and fan-in d or less at every multiplication gate. Then $dim(\partial(f)) \leq s2^d$*

Proof. Every linear function g satisfy $\partial(g) = \{1, g\}$ and so we conclude $dim(\partial(g)) \leq 2$. The lemma follows by proposition 1.21. □

We conclude the proof of theorem 1.20 from lemma 1.22, the fact that homogeneous circuits computing a degree d polynomial cannot have multiplication fan-in exceeding d, and a lower bound on the dimension of the partials of symmetric functions below.

Lemma 1.23. *Let SYM_n^d denote the d'th elementary symmetric polynomial. Then*

$$dim(\partial(SYM_n^d)) \geq \binom{n}{d} \tag{1.12}$$

Proof. For a subset $R \subseteq [n]$ we let SYM_R^d be the d'th symmetric polynomial over variables in R. Therefore, $SYM_n^d = SYM_{[n]}^d$. Now let S and T range over the set I, which we order as vectors by fixing an order on the elements of I. The vector U is comprised of all monomials of length d, i.e $U_S = \prod_{i \in S} x_i$. The vector V contains the partial derivatives of $SYM_{[n]}^d$ with respect to d-monomials, which can be calculated to be $V_T = SYM_{[n]}^{2d}$ obtained by assigning zeros to all variables in T. Now we must lower bound $dim(V)$. We know that $V = DU$ where D is the $I \times I$ disjointness matrix $D_{T,S} = 1$ if $S \cup T =$ and 0 otherwise. Since D is full rank and all monomials in U are independent we have

$$dim(\partial(SYM_n^{2d})) \geq dim(V) = dim(U) = \binom{n}{d} \tag{1.13}$$

□

□

1.2.7 Determinant Lower Bound

Next to the permanent, the determinant is the second most studied polynomial in arithmetic circuit complexity, and depth-3 circuits being the focus of this work it is fitting that we present a lower bound for depth-3 circuits computing the determinant. The following determinantal lower bound was proven in [SW02] and held for 15 years. The recent improvements to this result are the main focus of this work. Finally, the proof uses

canonical techniques of circuit lower bounds, and it's instructive to see them deployed in a well contained setting.

Theorem 1.24. *Every $\sum\prod\sum$ circuit computing the determinant of an $n \times n$ matrix must be of size at least $\Omega(n^4/\log n)$*

Note that there are n^2 variables in the matrix so this is actually a quadratic lower bound for depth-3 circuits.

Proof. First we eliminate gates of high degree by restricting our inputs to an affine subspace. This makes certain linear functions vanish. Secondly, we use the partial derivative method. In essence, we claim that the space of partial derivatives of the modified circuit is of low rank whereas the determinant has high rank. To prove the determinant has high rank we use the fact that the determinant is downward-self-reducible in the strong sense.

For a polynomial $f \in \mathbb{F}[x_1, ..., x_n]$ we denote with $\partial_{x_i}(f)$ the partial derivative of f with respect to x_i. We also define $\partial_{[x_1,x_2]}(f)$ as $\partial_{x_1}\partial_{x_2}(f) = \partial_{x_2}\partial_{x_1}(f)$. Similarly, for any $S \subseteq [n]$, we can define $\partial_S(f)$. Finally, for integer k we denote $\Gamma_k f$ to be the dimension of the vector space over $\mathbb{F}$ spanned by all polynomials of the form $\partial_S(f)$ where $|S| = k$ set. And we let $\partial_S(f)$ refer to the space.

We will use the following claims

Claim 1.25. *For every product gate v of degree r in a $\sum\prod\sum$ circuit, $\Gamma_k(f)$ is at most $\binom{r}{k}$ where f_v is the polynomial that v computes*

Proof. Assume $f_v = \prod_{i=1}^{r} \ell_i$ for linear forms ℓ_i. The space $\partial_S(f)$ is spanned by $\{\prod_{i\in T} \ell_i : T \subseteq [r], |T| = r - k\}$ whose size is at most $\binom{r}{k}$. □

Claim 1.26. *For every k, $\Gamma_k(DET_n) \geq \binom{n}{k}^2$ where DET_n is the determinant of an $n\times n$ matrix which we denote X.*

Proof. For every two sets $R = \{r_1 < r_2 < ... < r_k\}$ and $C = \{c_1 < c_2 < ... < c_k\}$ of $[n]$, let $S(R, C) = \{x_{r_1,c_1}, ..., x_{r_k,c_k}\}$. The polynomial $\partial_{S(R,C)}(DET_n)$ is linearly independent over $\mathbb{F}$. The number of such pairs (R, C) is $\binom{n}{k}^2$. □

The claims tell us that if all product gates in a $\sum\prod\sum$ circuit Φ computing DET_n are of degree at most D, then the size of Φ is at least $\binom{n}{k}^2/\binom{D}{k}$ for every k. If we choose $k = n^2/(D \cdot e)$, then using Stirling's approximation we find that the seize of Φ

is $|\Phi| > \Omega(e^{n^2/(D\cdot e)})$. Therefore, if D is linear in n then we get an exponential lower bound. The proof then requires us to massage the circuit into this form.

Idea: If there are "many" gates of "high" degree then the circuit has many wires. If the circuit has a few gates of high degree then by restricting the inputs to a subspace we can eliminate all gates and still get a circuit computing the determinant of a slightly smaller matrix.

So far we know that when all product gates are of small degree, the circuit must be large. If all the product gates containing a variable $x_{1,n}$ are of degree at most D then the size of Φ is at least $\Omega(e^{(n-1)^2/(D\cdot e)})$. Indeed, if we consider the circuit where we eliminate all product gates not containing $x_{1,n}$ in Φ and only maintain the gates of Φ containing $x_{1,n}$. Let's call this new circuit Ψ, which computes DET_{n-1} and all its gates are of degree at most D. Hence a lower bound on the size of Ψ and hence on the size of Φ follows. We can also generalize this reasoning to variables $x_{1,n-1}, x_{2,n-2}, ..., x_{1,1}$.

On the other hand, if there is a gate $v_{1,n}$ of degree at least D that contains $x_{1,n}$ then there is a linear function ℓ_1 so that the restriction of the polynomial $\Phi_{v_{1,n}}$ to the subspace $x_{1,n} = \ell_1$ is zero. Applying this restriction, we can eliminate at least D wires from Φ. Applying this line of argument for $x_{1,n-1}, x_{2,n-2}, ..., x_{1,2}$, if at some point we find that all gates containing $x_{1,i}$ are of degree D then we proceed as in our former argument. Otherwise we find an appropriate substitution to $x_{1,i}$. At the end of the process we get that either there are at least $\Omega(e^{n^2/(D\cdot e)})$ gates in Φ or that we eliminated $(n-1)D$ wires. In the latter case we set $x_{1,1} = 1$ and $x_{i,1} = 0$ for all $i \in \{2, ..., n\}$ which yields a circuit computing DET_{n-1}. Finally, if we denote the size of the smallest $\sum\prod\sum$ circuit computing the determinant of an $n \times n$ matrix as $\sum\prod\sum(DET_n)$, then for every D,

$$\sum\prod\sum(DET_n) \geq max\Big(\Omega(e^{(n-1)^2/(D\cdot e)}), \sum\prod\sum(DET_{n-1}) + (n-1)D\Big) \quad (1.14)$$

If we choose $D = n^2/(4e \log n)$ the theorem follows by induction. □

Thus, almost quadratic lower bounds were known for polynomials in VNP since [SW02], but it would not be until [KST16] that an almost cubic lower bound would be proven for polynomial families in VNP. Nevertheless, the technique of lower bounding the dimension of the span of partial derivatives of a polynomial family, and then upper bounding the span of partial derivatives computed by a circuit model lays the foundations for the more sophisticated techniques that will be the focus of the remainder of this work.

Shifted Derivative Measure

1.3 Introduction

We present the results of [Yau16] A depth three $\sum\prod\sum$ circuit consists of a layer of sum gates, followed by a layer of multiplication gates, followed by a single sum gate that outputs the computation of the circuit. The fan-in is unbounded, and the circuit size is measured in terms of the number of edges. As such, depth three circuits capture "sums of products of linear polynomials". A recent line of work on depth reduction [AV08, Koi10b, GKKS16**?**] has shown that moderately strong lower bounds for circuits of depth three imply a super-polynomial lower bound for general circuits. In addition, [Raz13b] shows that a strong enough lower bound for set-multilinear depth three circuits implies a super- polynomial lower bound for general arithmetic formulas. These depth reduction results pave an avenue towards proving super-polynomial lower bounds for general circuits/formulas by leveraging the apparent simple structure of depth three circuits. Unfortunately, it is still an open problem to prove super-polynomial lower bounds for depth three circuits over fields of characteristic zero. Below we present some of the seminal results in depth three lower bounds.

In [SW02], Shpilka and Widgerson proved a $\Omega(n^2)$ depth three circuit computing the elementary symmetric polynomials $ES_{YM_n^d}(x_1, x_2, ..., x_n) = \sum_{S\subseteq[n],|S|=d}\prod_{i\in S} x_i$ on n variables and degree $d = \Theta(n)$. In the same paper, the authors prove a near quadratic lower bound for the determinant polynomial [SW02]. Restricting the circuit model (homogeneity, multilinearity) and restricting the field characteristic yields better results. Over fixed finite fields, [GR00a] prove an exponential lower bound for the determinant and in [NW96] it was shown that any homogeneous depth three circuit computing $ES_{YM_n^{2d}}$ has size $\Omega\big((\frac{n}{4d})^d\big)$. More recently, in [KS15] a $n^{\Omega(\sqrt{d})}$ lower bound was proven for depth three circuits, with bottom fan-in bounded by n^ϵ for any fixed $\epsilon < 1$, computing an explicit n-variate polynomial of degree d.

Despite success in many restricted settings (homogenous, degree bounded product gates) the lower bounds in general cases remain relatively weak. Recently [KST16] gave near

cubic $\tilde{\Omega}(n^3)$ lower bounds for a polynomial family in VNP, which was followed by [BLS16] who gave a $\Omega\left(\frac{n^3}{2^{\sqrt{\log n}}}\right)$ lower bounds for a polynomial family in VP.

In this work we strengthen the latter lower bound to get a polynomial in VP on N variables and degree D satisfying $\text{poly}(N) > D > \log^2 N$, with size lower bound $\tilde{\Omega}(N^2 D)$. Setting $D = N$, this recovers the VNP result up to a $log^5(N)$ factor. Along the way we present a simplified polynomial and a tighter analysis of its multiplicative complexity. We also expand on the trade off between circuit size as a joint function of the degree of the polynomial and the number of variables — something that does not seem to have been explicitly clarified before.

Our main result is as follows.

Theorem 1.27. *There exists an explicit polynomial family $P_{N,D}$ computable in VP on N variables of degree D satisfying poly$(N) > D > \log^2 N$ such that any depth 3 circuit computing it has size $\tilde{\Omega}(N^2 D)$. Setting $D = N$ as in previous works recovers, up to a $\log^4(N)$ factor the $\tilde{\Omega}(N^3)$ bound for polynomials in VNP [KST16]*

1.4 Preliminaries

We discuss some of the language and common techniques relating to arithmetic circuits. An extended treatment can be found in the survey [SY10] of Shpilka and Yehudayoff.

Our general organization is as follows. Section (3) constructs a "hard" polynomial and bounds its size for bounded fan-in circuits. Section (4) presents the embedding procedure producing a polynomial that can be analyzed for unbounded fan-in.

1.4.1 Basic Notation

The ideal generated by a set of polynomials of the ring P will be denoted $\langle P \rangle$. We use $poly(N)$ to denote polynomial in N with an arbitrary constant exponent. A $\sum \prod^Y \sum$ circuit computes polynomials that are the sum of the product of at most Y affine linear forms. Similarly, a $\sum \prod^Q \prod^R \sum$ circuit consists of a layer of sum gates, followed by two layers of product gates with fan-in bounded by R and Q respectively, followed by a final sum gate. We observe that each $\sum \prod^Q \prod^R \sum$ circuit can be converted to a $\sum \prod^{QR} \sum$ circuit with constant factor overhead in size.

1.4.2 Shifted Partial Derivative Measure

As in previous works, we use a measure $\mu : \mathbb{F}[\mathbf{x}] \to \mathbb{N}$ to capture weakness of a circuit model in opposition to a "hard" family of polynomials giving us a lower bound for the circuit family. Our choice of measure is the "dimension of the shifted partials" introduced in [Kay12]. For polynomial $P \in \mathbb{F}[x_1, x_2, ..., x_N]$, let $\langle P \rangle^{=k}$ be the set of k'th order partials of P. Furthermore, let

$$\langle P \rangle^{=k}_{\leq \ell} := \{f \cdot p | \forall \text{ monomials } f \text{ s.t } deg(f) \leq \ell, \forall p \in \langle P \rangle^{=k}\} \tag{1.15}$$

Then for $k, \ell \in \mathbb{N}$, the shifted derivative measure is defined to be

$$\mu_{k,\ell} P = dim(\langle P \rangle^{=k}_{\leq \ell}) \tag{1.16}$$

Adding the parameter ℓ produces this shifted derivative measure that introduces "leeway" into the measure of the "dimension of the partial derivatives" introdued in [NW96]

1.4.3 Circuits under Affine Projections

Given polynomial $P \in \mathbb{F}[x_1, x_2, ..., x_N]$ as above, let $A : \mathbb{F}^N \to \mathbb{F}^N$ be an affine linear transform, then it is easy to show that $\mu_{k,\ell} P \circ A \leq \mu_{k,\ell} P$. In which case if A is invertible, then $\mu_{k,\ell} P \circ A = \mu_{k,\ell} P$. The takeaway is that the shifted derivative measure is invariant under invertible affine transforms.

Now let V be a subspace of $\mathbb{F}^N$ and $V^\perp$ be its complement. Then if A is an affine projection onto the space V, then we say $P \circ A$ is a subspace restriction $P|_V$. If we let U_V be the orthogonal projection of $\mathbb{F}^N$ to V, by the above discussion we observe that $\mu_{k,\ell} P \circ U_V \circ A = \mu_{k,\ell} P \circ U_V$. This is useful for the following reason.

The central barrier to proving lower bounds for bounded depth circuits is the unbounded fan-in. The key idea is then to restrict the polynomial with an affine transform A to an affine subspace V so that the product gates with large fan-in can be pruned. We are then left with a bounded fan-in circuit which we can analyze. However, we must now compute the measure of the polynomial $P \circ A$. We do this precisely by noting that $\mu_{k,\ell} P \circ U_V \circ A = \mu_{k,\ell} P \circ U_V$ and construct P so that its shifted derivative measure is easy to compute under orthogonal affine restrictions. In some sense we are "embedding" a polynomial for which we can analyze its shifted derivative measure within P. Section 3 constructs the embedded polynomial and section 4 details how the subspace restrictions are performed in practice.

1.5 Embedded Polynomial

First we construct a polynomial in VP for which we can analyze its shifted derivative measure and bound its circuit size for constant depth circuits with bounded fan-in.

1.5.1 Polynomial Construction

Let X be a b-by-n matrix of formal variables as shown below.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{b1} & x_{b2} & \dots & x_{bn} \end{bmatrix} \tag{1.17}$$

Let $J = (j_1, j_2, .., j_b)$ for $J \in [n]^b$. Then define the function $Permute(X)$ to be

$$Permute(X) = \prod_{i=1}^{b} \sum_{j=1}^{n} x_{ij}^{\frac{D}{b}} = \sum_{J \in [n]^b} x_{1j_1}^{\frac{D}{b}} x_{2j_2}^{\frac{D}{b}} ... x_{bj_b}^{\frac{D}{b}} \tag{1.18}$$

Notice that $Permute(X)$ has $N = nb$ variables and has degree D. For $b = \log n$, $Permute(X)$ is in VP by inspecting the sum and product in the definition.

1.5.2 Bounding Measure for Target Polynomial

The first lemma is presented as Proposition 9 in [AG13]. If polynomial $f \in \mathbb{F}[x_1, ..., x_N]$ is of the form $f = \sum_{i=1}^{s} \prod_{j=1}^{Q} G_{ij}(x_1, x_2, ..., x_N)$ where each G_{ij} is a polynomial of degree no greater than R, then the following inequality bounds the size of s.

Lemma 1.28. *For all $k, \ell \in \mathbb{N}$ let the shifted partial derivative measure $\mu_{k,\ell} f = dim(\langle f \rangle^k_{\leq \ell})$. Then for $k < Q$ the following lower bounds the size of s*

$$\frac{\mu_{k,\ell} f}{\binom{Q+k}{k}\binom{N+\ell+k(R-1)}{\ell+k(R-1)}} \leq s \tag{1.19}$$

With respect to circuits, s is the size of the top fan-in which is what we'll be using as a lower bound for circuit size. Q can then be interpreted as the top layer product gate fan-in. So long as each product gate has a fan-in consisting of polynomials of degree no greater than R, the above lemma holds. Summarizing these remarks, we find that the left hand side of the inequality is dependent only on the circuit model, and that k and ℓ are chosen for analytical convenience.

The next lemma has several formulations. We will present the formulation in Lemma 3 of [CM13].

First, we define a distance metric between any pair of monomials g and g' of identical degree. Let h be the monomial of minimum degree divisible by both g and g'. Then let $|g\Delta g'| = deg(h) - deg(g)$ which is well defined because $deg(g) = deg(g')$.

Lemma 1.29. *Let $f \in \mathbb{F}[x_1, x_2, .., x_N]$ be a polynomial, then the following inequality lower bounds the shifted partial derivative measure $\mu_{k,l}f$ for all $k, l \in \mathbb{N}$. If $S \subseteq \partial_k\langle f\rangle$ is a set of monomials satisfying for distinct $g, g' \in S$, $|g\Delta g'| \geq \tau$ then*

$$|S|\binom{N+\ell}{\ell} - |S|^2\binom{N+\ell-\tau}{\ell-\tau} \leq \mu_{k,\ell}f \tag{1.20}$$

Putting Lemma 0.1 and 0.2 together we obtain

$$\frac{|S|\binom{N+\ell}{\ell} - |S|^2\binom{N+\ell-\tau}{\ell-\tau}}{\binom{Q+k}{k}\binom{N+\ell+k(R-1)}{\ell+k(R-1)}} \leq s \tag{1.21}$$

Now we must determine the size of a set S satisfying the properties of Lemma 0.2 with a corresponding minimum distance τ for our polynomial $Permute(X)$.
Consider the following, we set $k = b = \log n$, and define $\partial_J Permute(X)$ for $J = (j_1, j_2, ..., j_k) \in [n]^k$ to be the k'th order derivative obtained by differentiating $Permute(X)$ by $x_{1j_1}x_{2j_2}...x_{kj_k}$. Then

$$\partial_J Permute(X) = x_{1j_1}^{\frac{D}{\log n}-1} x_{2j_2}^{\frac{D}{\log n}-1} ... x_{kj_k}^{\frac{D}{\log n}-1} \tag{1.22}$$

Then we define $S := \{\partial_J Permute(X) | \forall J \in [n]^k\}$ which gives us $|S| = n^k$. Furthermore, for any distinct $J, J' \in [n]^k$, J and J' differ in some coordinate j_i implying $\tau = \frac{D}{\log n} - 1$. Armed with our values of $|S|$ and τ, we can set the circuit parameters Q, R and the shifted derivative parameters k, ℓ and compute a lower bound on bounded fan-in depth four circuits.

1.5.3 Calculation

Lemma 1.30. *For any $\sum\prod^Q\prod^R\sum$ circuit computing $Permute(X)$, if we set the values for the circuit parameters $Q = n^{1-\frac{5}{\log n}}$, $R = \frac{\tau}{\log^2 n}$ and the shifted derivative parameters $k = \log n$, $l = \frac{n \log n}{2^{\frac{\log^2 n+1}{\tau}}-1}$, then the top fan-in s is greater than N^4. Adjusting the constant in the definition of Q gives us an $poly(N)$ bound of arbitrary constant degree.*

Proof. Plugging these parameters into (5) we find

$$\frac{n^k\binom{N+\ell}{\ell} - n^{2k}\binom{N+\ell-\tau}{\ell-\tau}}{\binom{Q+k}{k}\binom{N+\ell+k(R-1)}{\ell+k(R-1)}} \leq s \tag{1.23}$$

We apply standard binomial inequalities to obtain

$$\frac{n^k\binom{N+\ell}{\ell} - n^{2k}\binom{N+\ell}{\ell}\left(\frac{N+\ell}{\ell}\right)^{-\tau}}{\binom{Q+k}{k}\binom{N+\ell}{\ell}\left(\frac{N+\ell}{\ell}\right)^{k(R-1)}} \leq s \tag{1.24}$$

And remove the $\binom{N+\ell}{\ell}$ term to obtain

$$\frac{n^k - n^{2k}\left(\frac{N+\ell}{\ell}\right)^{-\tau}}{\binom{Q+k}{k}\left(\frac{N+\ell}{\ell}\right)^{k(R-1)}} \leq s \tag{1.25}$$

Now our setting of ℓ gives us $\left(\frac{N+\ell}{\ell}\right)^{-\tau} = \frac{1}{2}n^{-k}$ so that the numerator reduces to

$$n^k - n^{2k}\left(\frac{N+\ell}{\ell}\right)^{-\tau} = \frac{1}{2}n^k \tag{1.26}$$

The denominator reduces to

$$\binom{Q+k}{k}\left(\frac{N+\ell}{\ell}\right)^{kR} = \binom{Q+k}{k}n^{\frac{k^2R}{\tau}}2^{\frac{kR}{\tau}} \tag{1.27}$$

Now combining numerator and denominator we obtain

$$s \geq \frac{\frac{1}{2}n^k}{\binom{Q+k}{k}n^{\frac{k^2R}{\tau}}2^{\frac{kR}{\tau}}} \geq \frac{\frac{1}{2}n^k}{\binom{Q+k}{k}n} \geq \frac{\frac{1}{2}n^k}{Q^k n} \geq \frac{\frac{1}{2}n^k}{n^{(1-\frac{5}{\log n})k}n} = \frac{1}{2}n^4 \tag{1.28}$$

This concludes our analysis of $Permute(X)$. We can obtain any polynomial lower bound by adjusting the constant parameter 5 in the setting of Q which is all we need for the subspace restrictions detailed next. □

1.6 Putting it Together

We present the technique of subspace restrictions following the general presentation in [BLS16, KST16]. The proof idea is to construct an explicit polynomial $F_{N',D'}$ in VP with $N' = \Theta(N \log N)$ variables and degree $D' = \Theta(D \log N)$ where any circuit computing $F_{N',D'}$ satisfies the property that restricting any N product gates yields a circuit computing $Permute(X)$. So long as $Permute(X)$ must be computed by a

$poly(N)$ sized circuit with some large constant degree, then $F_{N',D'}$ must be computed by a $\Omega(NQR) = \tilde{\Omega}(N^2D) = \tilde{\Omega}(N'^2D')$ sized circuit. Note, it is for $F_{N',D'}$, not $Permute(X)$, for which we produce our almost cubic lower bound. First we present the construction of $F_{N',D'}$, then we present the subspace restriction procedure, and finally we prove Theorem 0.1.

1.6.1 Polynomial Embedding

$Permute(X)$ takes $N = n \log n$ variables. We now introduce the formal variables $W = \{w_1, w_2, ..., w_{2N}\}$ and $U = \{U_1, U_2, ..., U_N\}$. Where each $U_i \in U$ is a collection of q variables $U_i = \{u_{i1}, u_{i2}, ..., u_{iq}\}$ for $q = C \log n$ for constant factor C. Now let $M = \{m_1, m_2, ..., m_{2N}\}$ be $2N$ pairwise distinct subsets of $[C \log n]$ where each $m_i \in M$ is of size $|m_i| = C' \log n$. . Then for $i \in [2N]$ and $j \in N$, we define $\phi_i(U_j) = \prod_{y \in m_i} u_{jy}$. Now we are ready to define $F_{N',D'}(U, W)$. Let V be a set of N formal variables, defined as follows

$$V = \begin{bmatrix} \phi_1(U_1) & \phi_2(U_1) & \dots & \phi_{2N}(U_1) \\ \phi_1(U_2) & \phi_2(U_2) & \dots & \phi_{2N}(U_2) \\ \dots & \dots & \dots & \dots \\ \phi_1(U_N) & \phi_2(U_N) & \dots & \phi_{2N}(U_N) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{2N} \end{bmatrix} \tag{1.29}$$

Then we define

$$F_{N',D'}(U, W) = Permute(V) \tag{1.30}$$

Their is slight notational abuse since we initially defined $Permute$ to be a function taking a matrix of N variables but V is a vector. It is to be understood that in writing $Permute(V)$ we implicitly arrange V into a matrix. First we observe that $F_{N',D'}(U, W)$ has $N' = CN \log N + 2N = \Theta(N \log N)$ variables. Furthermore, the degree $D' = C'D \log N = \Theta(D \log N)$. Since the sets $m_i \in M$ are pairwise distinct, for each subset $A \in [2N]$ satisfying $|A| = N$ there exists a setting of the variables in U such that
$F_{N',D'}(U, W) = Permute(\chi_A(W))$ where $\chi_A(W)$ selects N variables from W corresponding to A. Therefore, we call the W's "relevant" variables and the U's "indicator" variables that we eventually set to be $\{0, 1\}$. We restate this critical property in the following lemma.

Lemma 1.31. *For each subset $A \in [2N]$ satisfying $|A| = N$, there exists a setting of the variables in U such that $F_{N',D'}(U, W) = Permute(\chi_A(W))$ where $\chi_A(W)$ selects N variables from W corresponding to A.*

1.6.2 Affine Subspace Restriction

Here we finish proving Theorem 0.1. For any $\sum\prod\sum$ circuit computing $F_{N',D'}(U,W)$ we say a product gate is "heavy" if its fan-in consists of more than QR sum gates that have a relevant variable $w_i \in W$ in their fan-in. Then there are two cases.

case 1: If there are more than $N = \Theta(n \log n)$ product gates with fan-in greater than $QR = n^{1-\frac{5}{\log n}} \frac{\tau}{\log^2 n} = \frac{nD}{32\log^3 n}$, then we have an $N\frac{nD}{32\log^3 n} = \Omega(\frac{N^2 D}{\text{polylog}(N)})$ lower bound on the number of wires in the circuit and we're done.

case 2: Consider a $\sum\prod\sum$ circuit with top fan-in s computing $F_{N',D'}(U,W)$. If there are fewer than N heavy product gates than we remove them in the following manner. Let $P(U,W)$ be a heavy product gate, then choose any sum gate $L(U,W)$ in the fan-in of $P(U,W)$ that is the affine sum of variables including a relevant $w_i \in W$. Therefore we can write $L(U,W) = \alpha w_i + L'(U,W)$ where $L'(U,W)$ is an affine linear form not involving w_i. Then rewiring the circuit so that $w_i = \frac{-1}{\alpha} L'(U,W)$ removes the sum gate $L(U,W)$ and the product gate $P(U,W)$. Repeating this process at most N times for all heavy product gates we are eventually left with a $\sum\prod^{QR}\sum$ circuit which we then pull apart to a $\sum\prod^{Q}\prod^{R}\sum$ circuit (Note: pulling the product apart does not change the size of the top fan-in). Now let $Y \in [2N]$ be the set of indices corresponding to the unrestricted variables in W, and let $A \subseteq Y$ be a subset of the unrestricted variables of size $|A| = N$. Then by lemma 0.5 we can set the U's so that $F_{N',D'}(U,W) = \mathit{Permute}(\chi_A(W))$. Taken together, we have a $\sum\prod^{Q}\prod^{R}\sum$ circuit with some top fan-in s' computing our hard polynomial $\mathit{Permute}(\chi_A(W))$. In the process of converting from $\sum\prod\sum$ to $\sum\prod^{Q}\prod^{R}\sum$ we have performed affine restrictions and set the variables in U, operations that can only decrease the size of the top fan-in. Therefore $s > s'$, and by lemma 0.4 we know $s > s' > N^4$.

Taking the minimum of case 1 and case 2 we obtain the size of any $\sum\prod\sum$ circuit computing $F_{N',D'}(U,W)$ is greater than $\min\left(\frac{N^2 D}{\text{polylog}(N)}, N^4\right) = \tilde{\Omega}\left(N'^2 D'\right)$ where we understand that N^4 can be any $poly(N)$. As a final comment, the $\tilde{\Omega}$ hides a $log^7 N$ factor, whereas the VNP result in [KST16] is almost cubic by a $\log^2 N$ factor. One avenue towards removal is avoiding the overhead in both variables and degree in the polynomial embedding.

Projected Shifted Derivative Measure

1.7 Introduction

The last idea we present in this work is that of lower bounding by projected shifted derivatives. The method was introduced in [KLSS14] to prove exponential lower bounds for depth-4 homogeneous formulas. We present an elaboration of this technique in the more recent paper [KS15] which proves exponential lower bounds for homogeneous depth-5 circuits over finite fields. [KS15] puts together many of the ideas that pervade arithmetic circuit complexity i.e taking advantage of multiplicities, projected shifted derivatives, multilinearization, the Nisan-Widgerson polynomial family, etc. Thus, we will end this work with an partial sketch of the [KS15] result, and discuss its connections to the depth reduction and homogeneity.

1.8 Projected Shifted Derivative

The main theorem we prove

Theorem 1.32. *[KS15] There is an explicit family of polynomials* $\{P_d : d \in \mathbb{N}\}$ *with* $Deg(P_d) = d$ *in the class* VNP *such that for any finite field* $\mathbb{F}_q$ *any homogeneous depth-5 circuit computing* P_d *must have size* $\exp(\Omega_q(\sqrt{d}))$.

Here the polynomial is chosen from the Nisan-Widgerson family introduced by [KSS14]. Instead of proving 1.32 we will prove an equally strong version where we replace homogeneity with a restricted fan-in for the top product gates.

Theorem 1.33. *There is an explicit family of polynomials* $\{P_d : d \in \mathbb{N}\}$ *with* $Deg(P_d) = d$ *in the class* VNP *such that for any finite field* $\mathbb{F}_q$ *any* $\sum\prod^{O(\sqrt{d})}\sum\prod\sum$ *circuit computing* P_d *must have size* $\exp(\Omega_q(\sqrt{d}))$.

For perspective, it is known that for characteristic zero fields a lower bound of $exp(\omega(d^{1/3}\log d))$ suffices to separate VP from VNP using the depth reduction results in [AV08, Koi10b, Tav15, GKKS16].

Now we may proceed with an outline of the proof. The key technique is to approximate the growth of the space of projected shifted partial derivatives of a polynomial. Instead of working with a space of formal polynomials, the idea is to treat them as a space of functions from $\mathbb{F}_q^n \to \mathbb{F}_q$. The primary advantage of "projecting" the derivatives is that we can choose a well engineered subset of $\mathbb{F}_q^n$ over which we can make a variety of simplifying assumptions.

1.8.1 Proof Idea

The proof proceeds as follows

1. Define a complexity measure $\Gamma : \mathbb{F}_q[x] \to \mathbb{N}$.
2. For all homogeneous depth-5 circuits C of size at most $\exp(\delta(\sqrt{d}))$ prove an upper bound on $\Gamma(C)$
3. For the target hard polynomial P, show that $\Gamma(P)$ is much larger than the upper bound provided in step 2.

We will present points (1) and (2) to explore how the projected derivative measure can be applied on a circuit. Point (3) is less instructive, and follows from more well known properties of the Nisan-Widgerson family of polynomials.

When working with the derivative and shifted derivative measure, we associate a linear space of polynomials to every polynomial in $\mathbb{F}_q[\mathbf{x}]$ and use the dimension of this space over $\mathbb{F}_q$ as a measure of complexity of the polynomial.

For the projected shifted derivative, the space of shifted derivatives is replaced by their evaluations on a subset of $\mathbb{F}_q^n$, where we give the evaluations some ordering, say lexicographically, to form vectors of evaluations. The precise definition is as follows.

Definition 1.34. Let k, l be some positive integer parameters and let $S \subset \mathbb{F}_q^n$. For any polynomial P define $\Gamma_{k,l,s}(P)$ as

$$\Gamma_{k,l,s}(P) := \mathrm{Dim}\{\mathrm{Eval}_S(\mathbf{x}^{=l}\partial^{=k}(P))\} \tag{1.31}$$

Next we define the Nisan-Widgerson polynomial family.

Definition 1.35. Let d, m, e be arbitrary parameters with m being a power of a prime, and $d, e \leq m$. 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 $d \leq m$ we have that $[d] \subseteq \mathbb{F}_m$. The Nisan-Widgerson polynomial with parameters d, m, e denoted by $NW_{d,m,e}$ is defined as

$$NW_{d,m,e}(\mathbf{x}) = \sum_{\substack{p(t)\in\mathbb{F}_m[t] \\ Deg(p)<e}} x_{1,p(1)}...x_{d,p(d)} \tag{1.32}$$

That is for every univariate polynomial $p(t) \in \mathbb{F}_m[t]$ of degree less than e we add one monomial that encodes the graph of p on the points $[d]$. This is a homogeneous, multilinear polynomial of degree d over dm variables with exactly m^e monomials.

Now we introduce some notation

1. For a polynomial $P \in \mathbb{F}_q[\mathbf{x}]$ and for a set $S \subseteq \mathbb{F}_q^n$ we shall denote by $\text{Eval}_S(P)$ the vector of the evaluation of P on points in S (say in lexicographic order). For a set of vectors V, their span over $\mathbb{F}_q$ will be denoted by $\text{Span}(V)$ and their dimension by $\text{Dim}(V)$.

2. We shall use $\mathcal{H}$ to denote the set $\{0,1\}^n \in \mathbb{F}_q^n$

3. A depth-5 circuit C computes a polynomial of the form

$$C = \sum_a \prod_b \sum_c \prod_d L_{a,b,c,d} \tag{1.33}$$

   Where $L_{a,b,c,d}$ are linear polynomials.

Furthermore we define the rank and terms of a circuit.

Definition 1.36. For a depth-5 circuit we shall denote by Terms(C) the set

$$Terms(C) := \Big\{ \prod_d L_{abcd} \Big\}_{a,b,c} \tag{1.34}$$

which are all products of linear polynomials computed by the bottommost product gates.

For any term $T = \prod_d L_d$, define $Rank(T)$ to be $Dim\{L_d\}_d$ which the maximum number of independent linear polynomials among the factors of T. For a depth-5 circuit C we shall use $\text{Rank}(C)$ to denote $\max_{T\in Terms(C)} Rank(T)$. For a parameter τ we shall use $\text{Terms}_{>\tau}(C)$ to refer to terms $T \in \text{Terms}(C)$ with $\text{Rank}(T) > \tau$.

1.9 Circuit Complexity Upper Bound

We will present the upper bounding of the projected shifted derivative measure on our circuit model. The bound hinges upon several insights.

1.9.1 Low rank gates are low-degree polynomials

The following lemma was presented in [GR00a, GR00b], to transform gates of low rank when working over a finite field.

Lemma 1.37. *[GR00a, GR00b] Let Q be a product of linear polynomials of rank at most τ. Then, there is a polynomial $\tilde{Q}$ of degre at most $(q-1)\tau$ such that $\tilde{Q} = Q(a)$ for all $\boldsymbol{a} \in \mathbb{F}_q^n$.*

Proof. Without loss of generality, we shall assume that the rank is equal to τ as the degree upper bound will only be better for a smaller rank and let $L_1, ..., L_\tau$ be linearly independent. Let

$$Q = \prod_{i=[\tau]} L_i \prod_{j \neq [\tau]} L_j \tag{1.35}$$

Here, each linear form in the second product term is in the linear span of the linear forms $\{L_i : i \in [\tau]\}$, and so can be expressed as their linear combination. Therefore, Q can be expressed as a polynomial in $\{L_i : i \in [\tau]\}$. Let $Q = f(L_1, L_2, ..., L_\tau)$. Since we are working over $\mathbb{F}_q$ it follows that for every choice of L_i and for every $\mathbf{a} \in F_q^n$, we have $L_i^q(\mathbf{a}) = L_i(\mathbf{a})$. So for every $a \in F_q^n$

$$f(L_1, L_2, ..., L_\tau)(\mathbf{a}) = [f(L_1, L_2, ..., L_\tau)] \mod \langle \{L_i^q - L_i : I = 1, ..., \tau\} \rangle](\mathbf{a}) \tag{1.36}$$

The lemma then follows by setting $\tilde{Q} := f(L_1, L_2, ..., L_\tau) \mod \langle \{L_i^q - L_i : I = 1, ..., \tau\} \rangle$ □

1.9.2 High rank gates are almost always zero

Let us assume that $\text{size}(C) \leq 2^{\sqrt{d}/100}$. Now we fix a threshold τ and all terms T with $\text{Rank}(T) > \tau$ we call high rank terms and the rest are low rank. Under a random evaluation in F_q^n every non-zero linear polynomial takes value zero with probability $1/q$. Thus, if we have a term that is a product of many independent linear polynomials, then with very high probability many of them will be set to zero. That is to say the term will vanish with high multiplicity at most points. We formalize this notion below.

Definition 1.38. Multiplicity: For any polynomial P and a point $\mathbf{a} \in \mathbb{F}_q^n$ we say that $\mathbf{a}$ vanishes with multiplicity t on P if $Q(\mathbf{a}) = 0$ for all $Q \in \partial^{\leq t-1}(P)$. In other words, $\mathbf{a}$ is a root of P and all its derivatives up to order $t-1$.

We shall denote by $\text{Mult}(P, \mathbf{a})$ the maximum t such that $\mathbf{a}$ vanishes on $\partial^{\leq t-1}(P)$

Now we would like to prove that high rank terms vanish with high multiplicity almost everywhere. We will use the fact that if T is a product of linear polynomials then $\mathbf{a}$ vanishes with multiplicity t on P if $\mathbf{a}$ vanishes on at least t factors of P.

Lemma 1.39. *Let $T = \prod_{i=1}^{d} L_i$ be a term of rank at least r. Then, for every $\delta > 0$,*

$$\Pr_{a \in \mathbb{F}_q^n}\Big[\textit{Mult}(T, \boldsymbol{a}) \leq (1-\delta)\frac{r}{q}\Big] \leq \exp\big(-\frac{\delta^2 r}{2q}\big) \tag{1.37}$$

Proof. Let $L_1, ..., L_r$ be linearly independent. Then, the evaluations of $L_1, ..., L_r$ at a point $a \in \mathbb{F}_q^n$ are also linearly independent and $\Pr_{a \in \mathbb{F}_q^n}[L_i(a) = 0] = 1/q$ for $i = 1, ..., r$. □

For $i = 1, ..., r$, we let Y_i be the indicator random variable that is one if $L_i(\mathbf{a}) = 0$ and zero otherwise. Then if we define $Y = \sum_{i \in [r]} Y_i$, we have by linearity of expectations

$$\mathbb{E}[Y] = \sum_{i \in [r]} \mathbb{E}[Y_i] = \frac{r}{q} \tag{1.38}$$

Since the events Y_i are linearly independent, by Chernoff Bound, we know that for every $\delta > 0$

$$Pr\Big[Y \leq (1-\delta)\frac{r}{q}\Big] \leq \exp\big(-\frac{\delta^2 r}{2q}\big) \tag{1.39}$$

Then we apply union bound on the high-rank gates in a small circuit to obtain

Corollary 1.40. *Let C be a depth-5 circuit over $\mathbb{F}_q$ such that $\textit{size}(C) \leq 2^{\sqrt{d}/100}$. Let $\tau = O(\sqrt{d})$ so that*

$$\exp\big(\frac{\tau}{8q}\big) < 2^{\sqrt{d}/50} \tag{1.40}$$

Then the union bound gives us

$$\Pr_{\boldsymbol{a} \in \mathbb{F}_q^n}\Big[\exists T \in \textit{Terms}_{>\tau}(C) : \textit{Mult}(T, \boldsymbol{a}) \leq \frac{\tau}{2q}\Big] \leq 2^{-\sqrt{d}/50} \tag{1.41}$$

Hence, we set our parameter $\tau = O(\sqrt{d})$ and $k = \tau/2q^3$

1.9.3 Projected Set

Next we discuss how we select the set of points S that we evaluate the shifted derivatives on. For a circuit of size $2^{\sqrt{d}/100}$, let $\mathcal{E}$ be a set of points $\mathbf{a}$ such that there is some $T \in \text{Terms}_{>\tau}(C)$ for which $\text{Mult}(T, \mathbf{a}) \leq k = \tau/2q^3$. Then by 1.40 we know that the "bad set" $\mathcal{E}$ is not too large. In fact,

$$|\mathcal{E}| = \delta \cdot q^n \text{ for some } \delta = \exp(-O(\sqrt{d})) \tag{1.42}$$

Let S be any subset of $\mathbb{F}_q^n/\mathcal{E}$ that is contained in a translate of a hypercube, that is there exists some $\mathbf{c} \in \mathbb{F}_q^n$ such that

$$S \subset (\mathbf{c} + \mathcal{H})/\mathcal{E} \tag{1.43}$$

We choose a translated hypercube to take advantage of the following fact, that evaluations of a polynomial on a hypercube can be approximated by a multilinear polynomial thus drastically reducing the degree of the polynomial computed by circuit C.

Lemma 1.41. *Fix a translate of a hypercube $\boldsymbol{c} + \mathcal{H}$. Then for every polynomial $Q \in \mathbb{F}_q[\boldsymbol{x}]$, there is a unique multilinear polynomial Q' such that $Deg(Q') \leq Deg(Q)$ and $Q'(\boldsymbol{a}) = Q(\boldsymbol{a})$ for every $\boldsymbol{a} \in \boldsymbol{c} + \mathcal{H}$*

Proof. If $\mathbf{a} \in \mathbf{c} + \mathcal{H}$ then for each $i \in [n]$ we have a_i to be either c_i or c_{i+1}. Thus it suffices to replace each x_i^2 by a linear polynomial in x_i that maps c_i to c_i^2 and $c_i + 1$ to $(c_i + 1)^2$. We want Q' to agree on all points on $\mathbf{c} + \mathcal{H}$ of degree at most $Deg(Q)$. One way to do this is to define $Q' = Q \mod I_c$ where I_c is the ideal defined by

$$I_c := \langle \{x_i^2 - (c_i^2 + (x_i - c_i)(2c_i + 1)) : i = 1, ..., n\} \rangle \tag{1.44}$$

It is easy to check that I_c vanishes on $\mathbf{c} + \mathcal{H}$ and any Q can be reduced to a multilinear polynomial modulo I_c. And since no multilinear polynomial can vanish on all of $\mathbf{c} + \mathcal{H}$, Q' is necessarily unique. □

We are now ready to present the upper bound of the complexity of the circuit C.

Lemma 1.42. *Let C be a depth-5 circuit, of formal degree at most $2d$ and size(C) $\leq 2^{\sqrt{d}/100}$, that computes an n-variate degree d polynomial. Let τ and k be chosen as above, and l be a parameter satisfying $l + k\tau q < n/2$. If S is any subset of $\mathbb{F}_q^n$ $\mathcal{E}$ that is contained in a translate of a hypercube, then*

$$\Gamma_{k,l,s} \leq 2^{\sqrt{d}/100} \binom{4d/\tau + 1}{k} \binom{n}{l + k\tau q} poly(n) \tag{1.45}$$

Proof. We can write the circuit as the sum of depth-4 circuit $C = R_1 + ... + R_s$, where $s \leq 2^{\sqrt{d}/100}$ and each R_i is a product of depth-3 circuits with $\text{Deg}(R_i) \leq 2d$. Since $\Gamma_{k,l,s}$ is subadditive, it suffices to show that for each R_i we have

$$\Gamma_{k,l,s}(R_i) \leq \binom{4d/\tau + 1}{k} \binom{n}{l + k\tau q} \text{poly}(n) \tag{1.46}$$

Then for each R_i, define the $R_i^{\leq \tau}$ as the polynomial obtained by removing all high rank terms satisfying $T \in \text{Terms}_{>\tau}(R_i)$ Then the lemma would follow from the following two claims.

Claim 1.43. *For every* $i \in [r]$, $\Gamma_{k,l,s}(R_i) = \Gamma_{k,l,s}(R_i^{\leq \tau})$

Claim 1.44. *For every* $i \in [r]$, $\Gamma_{k,l,s}(R_i^{\leq \tau}) \leq \binom{4d/\tau+1}{k}\binom{n}{l+k\tau q} poly(n)$

In essence, we can remove high rank gates because we are evaluating the measure on a set S that removes the "bad set", and then sub-additivity will reduces the problem to a sum over depth-4 circuit complexity. Thus we conclude this section with a proof of claim 1.43 and 1.44

Proof. 1.43 Let $R = Q_1...Q_m$. Let $T \in \text{Terms}_{>\tau}(C)$. We shall show if $R' = (Q_1 - T)Q_2...Q_m$, then for any k'th order partial derivative $\partial_{\mathbf{x}^\alpha}$,

$$\text{Eval}_S(\partial_{\mathbf{x}^\alpha}(R)) = \text{Eval}_S(\partial_{\mathbf{x}^\alpha}(R')) \tag{1.47}$$

Consider $R - R' = T \cdot Q_2...Q_m$. By the chain rule, every term in $\partial_{\mathbf{x}^\alpha}(R - R')$ is divisible by some k'-th order partial derivative of T with $k' \leq k$. Since we chose S so that every $\mathbf{a} \in S$ satisfies $\text{Mult}(T, \mathbf{a}) > k$, and hence $\mathbf{a}$ vanishes on $\partial^{\leq k}(T)$ for any $T \in \text{Terms}_{>\tau}(C)$. Thus, $\text{Eval}_S(\partial_{\mathbf{x}^\alpha}(R^{\leq \tau}))$ where $\text{Deg}(\mathbf{x}^\alpha) = k$, and $\Gamma_{k,l,s}(R) = \Gamma_{k,l,s}(R^{\leq \tau})$ □

Proof. 1.44 As this proof is fairly involved, we will present it as a proof sketch. Let $R^{\leq \tau} = Q_1...Q_d$ with each Q_i being a $\sum\prod\sum$ circuit. Some of these $Q_i's$ could have degree more than τ although their rank is bounded by τ Let us denote them $Q_1, ..., Q_m$. Then let us group the low degree gates together (multiplying them) to ensure that all of them have degree between $\tau/2$ and τ denoted $Q'_1, ..., Q'_r$. Then we can write

$$R^{\leq \tau} = Q_1...Q_m \cdot Q'_1..Q'_r \tag{1.48}$$

Note all the Q's and the Q' terms are low rank, and $m + r \leq 4d/\tau + 1$. Now that we've separated $R^{\leq \tau}$ into the product of high degree and low degree polynomials, we can use the fact that low rank gates are low degree polynomials. Then we evaluate these low degree polynomials over a subset of a hypercube so we can multilinearize the polynomial.

Taken together, we can severely restrict the complexity of R. Thus, we present the exact manipulations required to upper bound the projected shifted derivative measure on R.

$$\partial_{\mathbf{x}^\alpha}(R) \in \text{ Span } \left\{\partial_{\mathbf{x}^\beta}(Q_A)\cdot\partial_{\mathbf{x}^\gamma}(Q'_B)\cdot Q_{\bar{A}}\cdot Q'_{\bar{B}} : \mathbf{x}^\alpha = \mathbf{x}^\beta\cdot\mathbf{x}^\gamma, A\subseteq[m], B\subseteq[r], |A|+|B| = k\right\}$$

$$\in \partial_{\mathbf{x}^\alpha}(R) \in \text{ Span } \left\{\partial_{\mathbf{x}^\beta}(Q_A)\cdot x^{\leq k\tau}\cdot Q_{\bar{A}}\cdot Q'_{\bar{B}} : \mathbf{x}^\alpha = \mathbf{x}^\beta\cdot\mathbf{x}^\gamma, A\subseteq[m], B\subseteq[r], |A|+|B| = k\right\}$$

$$\implies \mathbf{x}^{=l}\partial_{\mathbf{x}^\alpha}(R) \subseteq \text{ Span } \left\{\partial_{\mathbf{x}^\beta}(Q_A)\cdot x^{\leq l+k\tau}\cdot Q_{\bar{A}}\cdot Q'_{\bar{B}} : \mathbf{x}^\alpha = \mathbf{x}^\beta\cdot\mathbf{x}^\gamma, A\subseteq[m], B\subseteq[r], |A|+|B| = k\right\}$$

$$\implies \text{Eval}_S(\mathbf{x}^{=l}\partial_{\mathbf{x}^\alpha}(R)) \subseteq \text{ Span } \left\{\text{Eval}_S(\partial_{\mathbf{x}^\beta}(Q_A)\cdot x^{\leq l+k\tau}\cdot Q_{\bar{A}}\cdot Q'_{\bar{B}}) : {}^{\mathbf{x}^\alpha=\mathbf{x}^\beta\cdot\mathbf{x}^\gamma}_{A\subseteq[m],B\subseteq[r],|A|+|B|=k}\right\}$$

Now consider the term $\partial_{\mathbf{x}^\beta}(Q_A)$, which is a linear combination of term $T_1, ..., T_{|A|}$ where each T_i is a product of linear polynomials and has rank at most τ. Then using lemma 1.37 on each of the $T'_i s$ we obtain

$$\text{Eval}_S(\partial_{\mathbf{x}^\beta}(Q_A)) \in \text{ Span } \left\{\text{Eval}_S(\ \mathbf{x}\ ^{\leq(q-1)\tau|A|})\right\}$$

Therefore,

$$\text{Eval}_S(\mathbf{x}^{=l}\partial_{\mathbf{x}^\alpha}(R)) \subseteq \text{ Span } \left\{\text{Eval}_S(\partial_{\mathbf{x}^\beta}(Q_A)\cdot x^{\leq l+k\tau}\cdot Q_{\bar{A}}\cdot Q'_{\bar{B}}) : {}^{\mathbf{x}^\alpha=\mathbf{x}^\beta\cdot\mathbf{x}^\gamma}_{A\subseteq[m],B\subseteq[r],|A|+|B|=k}\right\}$$

$$\implies \text{Eval}_S(\mathbf{x}^{=l}\partial_{\mathbf{x}^\alpha}(R)) \subseteq \text{ Span } \left\{\text{Eval}_S(\ \mathbf{x}\ ^{\leq l+k\tau+(q-1)k\tau}\cdot Q_{\bar{A}}\cdot Q'_{\bar{B}}) : {}^{\mathbf{x}^\alpha=\mathbf{x}^\beta\cdot\mathbf{x}^\gamma}_{A\subseteq[m],B\subseteq[r],|A|+|B|=k}\right\} \quad (1.49)$$

Now we apply the multilinearization lemma 1.41. To rehash, multilinearization of a polynomial f produces an equivalent multilinear polynomial that agrees with all evaluations on a translate of a hypercube. Hence,

$$\text{Eval}_S(\mathbf{x}^{=l}\partial_{\mathbf{x}^\alpha}(R)) \subseteq \text{ Span } \left\{\text{Eval}_S(\ \mathbf{x}\ ^{\leq l+qk\tau}_{mult}\cdot Q_{\bar{A}}\cdot Q'_{\bar{B}}) : {}^{\mathbf{x}^\alpha=\mathbf{x}^\beta\cdot\mathbf{x}^\gamma}_{A\subseteq[m],B\subseteq[r],|A|+|B|=k}\right\} \quad (1.50)$$

Finally, we use the fact that $m + r \leq 4d/\tau + 1$ to get the bound

$$\Gamma_{k,l,s}(R) \leq \binom{4d/\tau+1}{k}\binom{n}{l+k\tau q}\cdot n \quad (1.51)$$

□

Putting claim 1.43 and 1.44 together with the sub-additivity property of the projected shifted derivative measure concludes the lemma. □

This completes the treatment of upper bounding the projected derivative measure on depth-5 circuits with restricted top product gate fan-in which can be shown to be equivalent to depth-5 homogeneous circuits. The rest of the proof technique would require a strong lower bound on the projected derivative measure of the Nisan-Widgerson family of polynomials, which together with the lower bound on the circuit will give us the exponential lower bound as advertised.

Bibliography

[AG13] Neeraj Kayal Ramprasad Saptharishi Ankit Gupta, Pritish Kamath. Approaching the chasm at depth four. In *Conference on Computational Complexity*. IEEE, June 2013.

[AV08] Manindra Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, 00(undefined):67–75, 2008.

[BLS16] Nikhil Balaji, Nutan Limaye, and Srikanth Srinivasan. An almost cubic lower bound for ΣΠΣ circuits computing a polynomial in VP. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:143, 2016.

[BQH82] Allan Borodin, Joachim Von Zur Qathen, and John Hopcroft. Fast parallel matrix and gcd computations. In *in proceedings of the 23rd annual symposium on foundations of computer science (FOCS82)*, pages 65–71, 1982.

[CCL08] Jin-Yi Cai, Xi Chen, and Dong Li. A quadratic lower bound for the permanent and determinant problem over any characteristic not equal to 2. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 491–498, New York, NY, USA, 2008. ACM.

[CM13] Suryajith Chillara and Partha Mukhopadhyay. Depth-4 lower bounds, determinantal complexity: A unified approach. *arXiv preprint arXiv:1308.1640*, 2013.

[GKKS16] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic circuits: A chasm at depth 3. *SIAM Journal on Computing*, 45(3):1064–1079, 2016.

[GR00a] D. Grigoriev and A. Razborov. Exponential lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. *Applicable Algebra in Engineering, Communication and Computing*, 10(6):465–487, 2000.

[GR00b] D. Grigoriev and A. Razborov. Exponential lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. *Applicable Algebra in Engineering, Communication and Computing*, 10(6):465–487, 2000.

[HY09] Pavel Hrubeš and Amir Yehudayoff. Arithmetic complexity in algebraic extensions, 2009.

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

[KLSS14] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An exponential lower bound for homogeneous depth four arithmetic formulas. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 61–70, 2014.

[Koi10a] Pascal Koiran. Arithmetic circuits: the chasm at depth four gets wider. *CoRR*, abs/1006.4700, 2010.

[Koi10b] Pascal Koiran. Arithmetic circuits: the chasm at depth four gets wider. *CoRR*, abs/1006.4700, 2010.

[KS15] Mrinal Kumar and Ramprasad Saptharishi. An exponential lower bound for homogeneous depth-5 circuits over finite fields. *CoRR*, abs/1507.00177, 2015.

[KSS14] Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 146–153, 2014.

[KST16] Neeraj Kayal, Chandan Saha, and Sébastien Tavenas. An Almost Cubic Lower Bound for Depth Three Arithmetic Circuits. In Yuval Rabani Ioannis Chatzigiannakis, Michael Mitzenmacher and Davide Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 33:1–33:15, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[MR04] Thierry Mignon and Nicolas Ressayre. A quadratic bound for the determinant and permanent problem. In *International Mathematics Research Notices*, pages 2004–4241, 2004.

[NW96] Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1996.

[Raz10] Ran Raz. Elusive functions and lower bounds for arithmetic circuits. *Theory of Computing*, 6(7):135–177, 2010.

[Raz13a] Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. *J. ACM*, 60(6):40:1–40:15, November 2013.

[Raz13b] Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. *J. ACM*, 60(6):40:1–40:15, November 2013.

[Rys63] H.J. Ryser. *Combinatorial mathematics*. Carus mathematical monographs. Mathematical Association of America; distributed by Wiley [New York, 1963.

[Str73] Volker Strassen. Vermeidung von divisionen. *Journal fr die reine und angewandte Mathematik*, 264:184–202, 1973.

[SW02] Amir Shpilka and Avi Wigderson. Depth-3 arithmetic circuits over fields of characteristic zero. *Comput. Complex.*, 10(1):1–27, January 2002.

[SY10] Amir Shpilka and Amir Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Found. Trends Theor. Comput. Sci.*, 5(38211;4):207–388, March 2010.

[Tav15] Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. *Information and Computation*, 240:2–11, 2015.

[Val79] L. G. Valiant. Completeness classes in algebra. In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*, STOC '79, pages 249–261, New York, NY, USA, 1979. ACM.

[VSBR83] L. G. Valiant, S. Skyum, S. Berkowitz, and C. Rackoff. Fast parallel computation of polynomials using few processors. *SIAM J. Comput*, 1983.

[Yau16] Morris Yau. Almost cubic bound for depth three circuits in vp. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:187, 2016.