Almost Cubic Bound for Depth Three Circuits in VP

Morris Yau *

November 20, 2016

Abstract

In "An Almost Cubic Lower Bound for \(\sum \prod \sum\) circuits in VP", [BLS16] present an infinite family of polynomials, \(\{P_n\}_{n \in \mathbb{Z}^+}\), with \(P_n\) on \(N = \Theta(n \text{ polylog}(n))\) variables with degree \(N\) being in VP such that every \(\sum \prod \sum\) circuit computing \(P_n\) is of size \(\Omega\left(\frac{N^3}{2 \log N}\right)\). We present a modified polynomial and perform a tighter analysis to obtain an \(\tilde{\Omega}(N^3)\) lower bound. More generally, we show that for every \(N\) and \(D\) satisfying \(\text{poly}(N) > D > \log^2 N\), there exist polynomials \(P_{N,D}\) on \(N\) variable of degree \(D\) in VP that can not be computed by circuits of size \(\tilde{\Omega}(N^2 D)\).

1 Introduction

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 wires. As such, depth three circuits capture "sums of products of linear polynomials". A recent line of work on depth reduction [AV08, Koi10, GKKS16, Tav15] has shown that moderately strong lower bounds for circuits of depth three implies a super-polynomial lower bound for general circuits. In addition, [Raz13] 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 build 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 circuits of 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 \(ES_{YM_n^d}(x_1, x_2, \ldots, x_N) = \)

*Harvard John A. Paulson School of Engineering and Applied Sciences, Department of Computer Science
\[ \sum_{S \subseteq [N], |S| = d} \prod_{i \in S} x_i \] 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, [GR00] proves an exponential lower bound for determinant and in [NW96] it was shown that any homogeneous depth three circuit computing \( ES_Y M^2_n \) has size \( \Omega(\frac{n}{d^2} y^d) \). More recently, in [KS15] a \( n^{\Omega(\sqrt{d})} \) lower bound was proved for depth three circuits, with bottom fan-in bounded by \( n^\epsilon \) 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 \( \Omega\left(\frac{n^3}{2^{\log n^2}}\right) \) 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 \( \tilde{\Omega}(N^2D) \). Setting \( D = N \), this recovers the VNP result up to a \( \log^4(N) \) 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 1.1.** There exists an explicit polynomial family \( P_{N,D} \) computable in VP on \( N \) variables of degree \( D \) satisfying \( \text{poly}(N) > D > \log^2 N \) such that any depth 3 circuit computing it has size \( \tilde{\Omega}(N^2D) \). 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].

## 2 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.

### 2.1 Basic Notation

The ideal generated by a set of polynomials of the ring \( P \) will be denoted \( (P) \). We use \( \text{poly}(N) \) to denote polynomial in \( N \) with an arbitrary constant exponent. A \( \Sigma \prod \Sigma \) circuit computes polynomials that are the sum of the product of at most \( Y \) affine linear forms. Similarly, a \( \Sigma \prod^Q \prod^R \Sigma \) 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 \( \Sigma \prod^Q \prod^R \Sigma \) circuit can be converted to a \( \Sigma \prod^Q R \Sigma \) circuit with constant factor overhead in size.
2.2 Shifted Partial Derivative Measure

As in previous works, we use a measure $\mu: \mathbb{F}[x] \to \mathbb{N}$ 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 $P \in \mathbb{F}[x_1, x_2, \ldots, x_N]$, let $\langle P \rangle_{\leq k}$ be the set of $k$'th order partials of $P$. Furthermore, let

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

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

$$\mu_{k, \ell} P = \dim(\langle P \rangle_{\leq k})$$

Adding the parameter $\ell$ produces this shifted derivative measure that introduces "leeway" into the measure of the "dimension of the partial derivatives" introduced in [NW96]

2.3 Circuits under Affine Projections

Given polynomial $P \in \mathbb{F}[x_1, x_2, \ldots, x_N]$ 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 $\mu_{k, \ell} P \circ A = \mu_{k, \ell} P \circ U_V$. 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 $\mu_{k, \ell} P \circ A = \mu_{k, \ell} P \circ U_V$ 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 2 constructs the embedded polynomial and section 4 details how the subspace restrictions are performed in practice.

3 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.
3.1 Polynomial Construction

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

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots \\ x_{b1} & x_{b2} & \cdots & x_{bn} \end{bmatrix}$$ (3)

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

$$Permute(X) = b \prod_{i=1}^{b} \sum_{j=1}^{n} \sum_{J \in [n]^b} x_{j_{J1} j_{J2} \ldots j_{Jb}}^p$$ (4)

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.

3.2 Bounding Measure for Target Polynomial

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

**Lemma 3.1.** For all $k, \ell \in \mathbb{N}$ let the shifted partial derivative measure $\mu_{k,\ell f} = \dim((f)^{k}_{\leq \ell})$. Then for $k < Q$ the following lower bounds the size of $s$

$$
\frac{\mu_{k,\ell f}}{\binom{Q+k}{k} \binom{N+\ell+k(R-1)}{\ell+k(R-1)}} \leq s
$$ (5)

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 $\ell$ 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 3.2.** Let $f \in \mathbb{F}[x_1, x_2, \ldots, x_N]$ be a polynomial, then the following inequality lower bounds the shifted partial derivative measure $\mu_{k,\ell f}$ for all $k, \ell \in \mathbb{N}$. If $S \subseteq \partial_k(f)$ is a set of monomials satisfying for distinct $g, g' \in S$, $|g \Delta g'| \geq \tau$ then

$$
|S| \binom{N+\ell}{\ell} - |S|^2 \binom{N+\ell-\tau}{\ell-\tau} \leq \mu_{k,\ell f}
$$ (6)
Putting Lemma 0.1 and 0.2 together we obtain

\[ |S|^{(N+\ell)_{\ell}} - |S|^{2(\frac{N+\ell}{\ell})^{-r}} \leq s \]

we set \( k \) permutating

We apply standard binomial inequalities to obtain

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

\[ \delta_j \text{Permute}(X) = \frac{\partial}{\partial x_{j_1}} \frac{\partial}{\partial x_{j_2}} \ldots \frac{\partial}{\partial x_{j_k}} \]

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

### 3.3 Calculation

**Lemma 3.3.** For any \( \sum [Q] \prod [R] \sum \text{ circuit computing } \text{Permute}(X) \), if we set the values for the circuit parameters \( Q = n^{1 - \frac{1}{6m_n}} \), \( R = \frac{1}{\log n} \) and the shifted derivative parameters \( k = \log n, l = \frac{n \log n}{2 \log^2 n - 1} \), then the top fan-in \( s \) is greater than \( N^4 \). Adjusting the constant in the definition of \( Q \) gives us an \( \text{poly}(N) \) bound of arbitrary constant degree.

**Proof.** Plugging these parameters into (5) we find

\[ n^k (\frac{N+\ell}{\ell}) - n^{2k} (\frac{N+\ell}{\ell})^{-r} \leq s \]

We apply standard binomial inequalities to obtain

\[ n^k (\frac{N+\ell}{\ell}) - n^{2k} (\frac{N+\ell}{\ell})^{-r} \leq s \]

And remove the \( (\frac{N+\ell}{\ell}) \) term to obtain

\[ \frac{n^k}{(Q+k) (\frac{N+\ell}{\ell})^{k(R-1)}} \leq s \]

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

\[ n^k - n^{2k} (\frac{N+\ell}{\ell})^{-r} = \frac{1}{2} n^k \]
The denominator reduces to
\[ s \left( \frac{Q + k}{k} \right) \left( \frac{N + \ell}{k} \right) = s \left( \frac{Q + k}{k} \right) n^{2R - \frac{kR}{\ell}} \]

(13)

Now combining numerator and denominator we obtain
\[ s \geq \frac{1}{n^k} + \frac{1}{n^k} \geq \frac{1}{2n^k} \geq n^{(1 - \frac{2R}{\ell})k} n = \frac{1}{2} n^k \]

(14)

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. \(\square\)

4 Putting it Together

We give 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 \(\tilde{\Omega}(NQR) = \tilde{\Omega}(N^2D) = \tilde{\Omega}(N^2D')\) 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.

4.1 Polynomial Embedding

Permute(X) takes \(N = n \log n\) variables. We now introduce the formal variables \(W = \{w_1, w_2, \ldots, w_{2N}\}\) and \(U = \{U_1, U_2, \ldots, U_N\}\). Where each \(U_i \in U\) is a collection of \(q\) variables \(U_i = \{u_i_1, u_i_2, \ldots, u_i_q\}\) for \(q = \text{const}\) for constant factor \(C\). Now let \(M = \{m_1, m_2, \ldots, m_{2N}\}\) be \(2N\) pairwise distinct subsets of \([\text{const}]\) 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 \(\phi_i(U_j) = \prod_{y \in m_i} u_{yj}\). Now we are ready to define \(F_{N', D'}(U, W)\).

Let \(V\) be a set of \(N\) formal variables, defined as follows
\[ V = \begin{bmatrix} \phi_1(U_1) & \phi_2(U_1) & \ldots & \phi_{2N}(U_1) \\ \phi_1(U_2) & \phi_2(U_2) & \ldots & \phi_{2N}(U_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(U_N) & \phi_2(U_N) & \ldots & \phi_{2N}(U_N) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{2N} \end{bmatrix} \]

(15)

Then we define
\[ F_{N', D'}(U, W) = \text{Permute}(V) \]

(16)

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 $F_{N', D'}(U, W) = \text{Permute}(\chi_A(W))$ 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 4.1.** For each subset $A \in [2N]$ satisfying $|A| = N$, there exists a setting of the variables in $U$ such that $F_{N', D'}(U, W) = \text{Permute}(\chi_A(W))$ where $\chi_A(W)$ selects $N$ variables from $W$ corresponding to $A$.

### 4.2 Affine Subspace Restriction

Here we finish proving Theorem 0.1. For any $\Sigma [\Sigma \Sigma]$ 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 $QR = n^{1 - \frac{v}{\log n}} = \frac{nD}{32\log^2 n}$, then we have an $N \cdot n\frac{D}{32\log^2 n} = \Omega\left(\frac{N^2 D}{\text{polylog}(N)}\right)$ lower bound on the number of wires in the circuit and we’re done.

**case 2:** Consider a $\Sigma [\Sigma \Sigma]$ 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 $L(U, W) = aw_i + L'(U, W)$ where $L'(U, W)$ is an affine linear form not involving $w_i$. Then rewiring the circuit so that $w_i = \frac{-1}{a}L'(U, W)$ 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 $\Sigma [\Sigma \Sigma] [\Sigma \Sigma]$ circuit which we then pull apart to a $\Sigma [\Sigma \Sigma] [\Sigma \Sigma]$ 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 $F_{N', D'}(U, W) = \text{Permute}(\chi_A(W))$. Taken together, we have a $\Sigma [\Sigma \Sigma \Sigma]$ circuit with some top fan-in $s'$ computing our hard polynomial $\text{Permute}(\chi_A(W))$. In the process of converting from $\Sigma [\Sigma \Sigma]$ to $\Sigma [\Sigma \Sigma \Sigma]$ 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 $\Sigma [\Sigma \Sigma]$ circuit computing $F_{N', D'}(U, W)$ is greater than $\min\left(\frac{N^2 D}{\text{polylog}(N)}, N^4\right) = \Omega(N^{2D'})$ where we understand that $N^4$ can be any $\text{poly}(N)$. As a final comment, the $\tilde{\Omega}$ hides a $\log^6 N$ factor, whereas the VNP result in [KST16] is almost cubic by a $\log^2 N$ factor. Removal of this final $\log^4 N$ can be done if we do not incur the $\log N$ costs in the polynomial embedding, which would bring the VP result to match the VNP bounds.
References


