Doctoral thesis

On Some Lower Bounds in Arithmetic Circuit Complexity

A thesis submitted in partial fulfilment of the requirement of the degree of

Doctor of Philosophy (Ph.D) in Computer Science

by

Suryajith Chillara

Declaration

I declare that the thesis "On some lower bounds in arithmetic circuit complexity" submitted by me for the degree of Doctor of Philosophy is the record of work carried out by me during the period from August 2013 to May 2017 under the guidance of Prof. Partha Mukhopadhyay. This work has not formed the basis for the award of any degree, diploma, associateship, fellowship, titles in this or any other university or other similar institution of higher learning.

July 2017 Suryajith Chillara

Chennai Mathematical Institute
Plot H1, SIPCOT IT Park, Siruseri,
Kelambakkam P.O. 603103
India

Certificate

This is to certify that the Ph.D. thesis submitted by Suryajith Chillara to Chennai Mathematical Institute, titled "On some lower bounds in arithmetic circuit complexity" is a record of *bona fide* research work done during the period from August 2013 till May 2017 under my guidance and supervision. The research work presented in this thesis has not formed the basis for the award of any degree, diploma, associateship, fellowship, titles in this institute or any other university or institution of higher learning.

It is further certified that the thesis represents independent work by the candidate and collaboration when existed was neccessitated by the nature and scope of problems dealt with.

Partha Mukhopadhyay
(Thesis superviser)

Chennai Mathematical Institute
Plot H1, SIPCOT IT Park, Siruseri,
Kelambakkam P.O. 603103
India

Synopsis

The computational model

Arithmetic circuits are a most natural and a standard model for computing polynomials. The notion of efficiency translates to a polynomial sized arithmetic circuit for the polynomial computation. This gives rise to the notion of an algebraic analog of P vs NP, formally known as VP vs VNP. In this algebraic setting, the goal is to show that explicit polynomials in VNP do not have polynomial sized arithmetic circuits. In particular, we would like to study the Permanent polynomial of a $n \times n$ matrix which Valiant showed that it is a canonical polynomial for VNP. Valiant also conjectured that the permanent polynomial does not have polynomial sized arithmetic circuits [Val79]. Interestingly, a slight variant of this polynomial which is the determinant polynomial, can be computed by a small sized arithmetic circuit [MV97].

In general, proving circuit size lower bounds for arbitrary circuits is difficult. The best known lower bound is $\Omega(n \log n)$ [BS83] for a n-variate polynomial of degree n. Thus, restricted models of circuits were looked at with a hope that the success there would throw some light on the general models.

Background

In a surprising result, Agrawal and Vinay [AV08] showed that proving exponential circuit size lower bounds for depth four circuits implies exponential circuit size lower bounds for general arithmetic circuits. In particular, given a polynomial (or sub-exponential) sized general arithmetic circuit, it can be transformed into a depth four circuit of sub-exponential size. This initiated the effort to prove exponential lower bounds for depth four circuits (cf. the survey by Saptharishi [Sap15] for a series of related results). Koiran [Koi12] and Tavenas [Tav15] carefully analyzed the chasm shown in [AV08] and came up with an *improved* depth reduction. Gupta et al. showed that such a chasm exists even at depth three, over $\mathbb{Q}$ [GKKS13].

It is known that any polynomial in VP could be written as a determinant of a quasi-polynomial sized matrix whose entries are linear polynomials. The minimum size of the matrix in such an expression is called the determinantal complexity. Another approach to settle the Valiant's hypothesis is by showing that the determinantal complexity of an explicit polynomial is super quasi-polynomial. But, the best known lower bounds for the permanent polynomial are quadratic [MR04, CCL08, Yab15].

Contributions made as a part of this doctoral thesis

Lower bounds for depth four circuits of bounded fan-in.

Kayal et al. [KSS14] showed a matching lower bound against Tavenas' bound for depth four circuits of bounded fan-in for a polynomial in VNP. Fournier et al. [FLMS14] proved the same bound against a polynomial in VP. The key technical ingredient in those proof is the method of shifted partial derivatives [Kay12]. We showed that these exciting current known circuit size lower bounds for bounded depth circuits can be unified in a simple way [CM14a]. We show that using a combinatorial technique. Interestingly, the technique involves nothing beyond binomial estimates and property of derivatives. This proof technique allows one to prove a similar lower bound for any polynomial that

exhibits this combinatorial property we mention.

On the limits of depth reduction at depth 3 over small finite fields

Looking at the current exciting chasm at depth four and the subsequent results against it, one might wonder if it is sufficient to prove circuit size lower bounds at depth three. A recent result justifies that it is true over $\mathbb{Q}$ [GKKS13]. Strong enough lower bounds here can prove valiant's conjecture. However, we observe in [CM14b] that it is not true over fixed-size finite fields. A strong exponential circuit size lower bound is already known for the determinant polynomial [GK98]. We also improved the situation by proving depth three circuit size lower bounds for two seemingly different polynomials (one which is built from the combinatorial design of Nisan and Wigderson [NW94] and the other is again the iterated matrix multiplication polynomial) over fixed-size finite fields. This strengthens the already known lower bounds of Grigoriev and Karpinski, over fixed-size finite fields.

Exponential lower bounds for bounded fan-in depth five circuits

Let us now consider the arithmetic circuits that use just the addition gates and the powering gates. We shall refer to such circuits as powering circuits. In [CKW11] Chen et al. posed the following open question. Can the monomial $x_1x_2\ldots x_n$ be efficiently computed by a constant depth powering circuit?

Saptharishi [Sap15] observed that the monomial $x_1x_2\ldots x_n$ has non-trivial depth four and depth five powering circuits of size $2^{O(\sqrt{n})}$. Ideally, we would like to prove matching lower bounds but the current state of affairs is far from that. But we [CS17], along with the work of Engels et al. [ERS16] make some partial progress. We show that there are at least two restricted classes of depth five powering circuits which can not efficiently compute the monomial. Our models encapsulate the ones considered by Engels et al. [ERS16].

Determinantal complexity lower bounds for IMM polynomial

To resolve Valiant's hypothesis, proving $\mathrm{DetComp}(\mathsf{Perm}_n) = n^{\omega(\log n)}$ is sufficient. By improving upon the work of Von zur Gathen [vzG86], Cai [Cai90], Babai and Seress [vzG87], and Meshulam [Mes89], Mignon and Ressayre [MR04] proved a quadratic lower bound on the determinantal complexity of the permanent polynomial (cf. [CCL08, Yab15]). In this thesis, we observed that some basic determinantal complexity lower bounds can be proved using the dimension of the partial derivative space as a complexity measure. In particular, we showed that $\mathrm{DetComp}(\mathsf{IMM}_{n,d}) \geq \frac{dn}{2e}$ and $\mathrm{DetComp}(\mathsf{NW}_{n,\varepsilon n}) \geq \Omega(n^{1.5})$. Further, in [CM14a] we showed that the former can be improved to $\mathrm{DetComp}(\mathsf{IMM}_{n,d}) \geq \frac{dn}{2}$ using the rank of the Hessian matrix as a complexity measure. This was via an adaptation of the argument of Mignon and Ressayre, and Cai et al. [MR04, CCL08] to the iterated matrix multiplication polynomial. Since $\mathsf{IMM}_{n,d}$ is a generic algebraic branching program, $\mathrm{DetComp}(\mathsf{IMM}_{n,d}) < O(dn)$ thus making it tight up to a constant factor. This improves upon the previous bound of Jansen [Jan11]. We showed that any arithmetic formula computing the $\mathsf{IMM}_{n,d}(X)$ polynomial must be of size $\Omega(dn^3)$. This is through an adaptation of the argument of Kalorkoti [Kal85].

Tensor rank upper bounds for small formulas

Proving arithmetic formula size lower bounds has proven to be difficult. The best known lower bound is quadratic in the number of variables, due to Kalorkoti [Kal85]. In a surprising result, Raz [Raz10] showed that the existence of a connection between arithmetic formula lower bounds and the explicit tensors of high rank. Specifically, if a general arithmetic formula of polynomial size (say n^c) computes a set multilinear polynomial of degree $d \in (\omega(1), O(\log n / \log\log n))$ over nd variables such that the associated tensor of the polynomial is of high rank then the tensor rank is non-trivially bounded on the above to $n^{d(1-1/2^{O(c)})}$. In [CKSV16], we show that such a connection holds for a wider range of parameter d, $d \in (\omega(1), 2^{o(\log n)})$ if we start with the homo-

geneous arithmetic formulas. This helps us extend the results of Raz that hold over the range $\omega(1) \leq d \leq O(\log n / \log\log n)$ to $\omega(1) \leq d \leq O(\log n)$ since formulas of degree $O(\log n)$ can be homogenized without much overhead.

To my parents

Contents

Part I.

Preliminaries

1 Introduction

1.1. Algebraic complexity

One of the main goals in the theoretical study of computer science, is to better understand the notion of efficient computation. In particular, we would like to understand the ease or complexity of a few natural computations that we encounter. Turing [Tur37] introduced *a-machines* and thus formalized the notion of a computation as a run of a machine. The word *efficiency* now refers to a run that takes time that is polynomial in the size of the input. Shannon considered *Boolean circuits* as a model of computation for Boolean functions [Sha49]. It is important to note that Boolean circuits can be simulated by Turing machines. Later, researchers considered algebraic counterparts of the Boolean circuits, called Algebraic circuits, also referred to as Arithmetic circuits. By Boolean circuits computing a Boolean function, we mean that the Boolean circuit computes the evaluation table of the function. It is quite possible that two Boolean functions share the same evaluation table. But in the case of Arithmetic circuits, we are interested in the syntactic computation of the polynomials which is a stronger notion than functional computation. The underlying algebraic structure helps us understand the computations. Considering this, Valiant proposed that any circuit theoretic proof for $\mathsf{P} \neq \mathsf{NP}$ would have to be preceded by an analogous result in this more constrained arithmetic model [Val92].

Definition 1.1.1. *An arithmetic circuit is a* directed acyclic graph *with inputs which are*

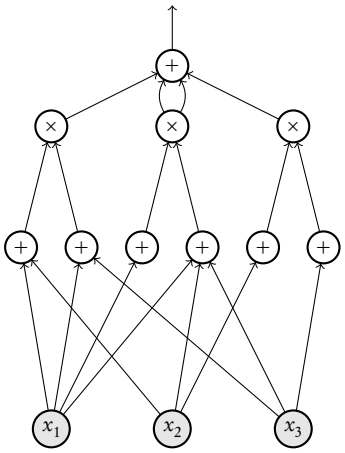

Figure 1.1.: An arithmetic circuit computing a polynomial over $\mathbb{F}[x_1, x_2, x_3]$.

the variables or constants as the leaves, the addition (+) and the multiplication (×) operations as internal nodes and the output node computes the polynomial. ◊

The complexity measures associated with this model are the size and the depth of the circuit which correspond to the number of arithmetic operations performed and the longest path from the leaf to a node respectively. If the out-degree of every node in the arithmetic circuit is at most 1, it is called an arithmetic formula.

Arithmetic circuits are a most natural and a standard model for computing polynomials. The notion of efficiency here translates to a polynomial sized arithmetic circuit for that computes the polynomial syntactically. This gives rise to the notion of an algebraic analog of P vs NP, VP vs VNP.

1.2. Algebraic complexity classes

Algebraic-P or VP is the class of polynomials which can be computed arithmetic circuits of polynomial size and polynomial degree. Algebraic-NP or VNP is the class of polynomials where given a monomial, its coefficient in the polynomial can be computed *efficiently*. Also, a polynomial is in VNP if it can be expressed as

$$f_n(X) = \sum_{Y \in \{0,1\}^m} g_{n+m}(X, Y)$$

where $m = |Y| = \text{poly}(n)$ and g_{n+m} is a polynomial family in VP.

In this algebraic setting, the goal is to show that explicit polynomials in VNP do not have polynomial sized arithmetic circuits (cf. [SY10]). In particular, we would like to study the Permanent polynomial of a $n \times n$ matrix (denoted by Perm_n) as Valiant showed that this polynomial is a complete polynomial for VNP.

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

where σ varies over S_n, a set of all the permutations of $[n]$. The notion of completeness in this setting is defined via the polynomial projections.

Definition 1.2.1. *A polynomial $f(x_1,\ldots,x_n)$ is said to be a projection of the polynomial $g(y_1,\ldots,y_m)$ if there exist linear polynomials $\{\ell_1,\ldots,\ell_m\}$ over $\mathbb{F}[x_1,\ldots,x_n]$ such that $f = g(\ell_1,\ldots,\ell_m)$.* ◊

Valiant conjectured that the permanent polynomial does not have polynomial sized arithmetic circuits [Val79]. Interestingly, a *slight variant* of this polynomial which is the determinant polynomial, can be computed by a small sized arithmetic circuit [MV97]. The determinant polynomial of a $n \times n$ matrix is defined as

$$\text{Det}_n = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^{n} x_{i,\sigma(i)}$$

where σ varies over S_n, a set of all the permutations of $[n]$ and $\text{sgn}(\sigma)$ is the sign of the permutation σ.

Conjecture 1.2.2 (Valiant)**.** VP $\subsetneq$ VNP.

1.3. Lower bounds

To prove Conjecture 1.2.2, one needs to prove super polynomial lower bounds for polynomials in VP. In general, proving circuit size lower bounds for arbitrary circuits is difficult. The best known lower bound is sub quadratic. Baur and Strassen [BS83] proved

a lower bound of $\Omega(d \log n)$ for a n-variate polynomial of degree d, $f = x_1^d + \cdots + x_n^d$. Thus, some restricted models of circuits were looked at with a hope that the success there would throw some light on the general models.

1.3.1. Constant depth circuits

In a surprising result, Agrawal and Vinay [AV08] showed that proving exponential circuit size lower bounds for depth four circuits implies exponential circuit size lower bounds for general arithmetic circuits. In particular, given a polynomial (or sub-exponential) sized general arithmetic circuit, it can be transformed into a depth four circuit of sub-exponential size. This initiated the effort to prove exponential lower bounds for depth four circuits (cf. the survey by Saptharishi [Sap15] for a series of related results). Koiran [Koi12] and Tavenas [Tav15] carefully analyzed the chasm shown in [AV08] and came up with an *improved* depth reduction. The chasm tells us that proving a size lower bound of $n^{\omega(d/t)}$ against the $\Sigma\Pi\Sigma\Pi^{[t]}$ would imply super polynomial lower bounds against general circuits. Towards proving such lower bounds Gupta et al. [GKKS14] proved a lower bound of $2^{\Omega(n/t)}$ against $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits computing the determinant or the permanent polynomial over a $n \times n$ matrix. They used the dimension of the *shifted partial derivate* space as the complexity measure [Kay12]. Kayal, Saha and Saptharishi [KSS14] pushed the bound of Gupta et al. [GKKS14] to $n^{\Omega(d/t)}$ for $\Sigma\Pi\Sigma\Pi^{[t]}$ circuits computing an explicit polynomial in VNP of degree d over $n^{O(1)}$-variables. This candidate polynomial is based on the combinatorial designs of Nisan and Wigderson [NW94]. In another surprising result Fournier et al. [FLMS14] proved that a *matching* bound could also be obtained for a polynomial in VP, the iterated matrix multiplication polynomial[1]. This tells us that the bound of Tavenas is tight (up to a constant in the exponent).

[1] In fact, this polynomial is also a canonical polynomial for *Algebraic branching programs*.

1.3.2. Unified analysis of lower bounds against bounded fan-in depth-4 circuits

One of the main motivations of our study comes from this interesting fact that two seemingly different polynomials $\mathsf{NW}_{n,r} \in \mathsf{VNP}$ and $\mathsf{IMM}_{n,n} \in \mathsf{VP}$ behave very similarly as far as the $2^{\Omega(\sqrt{n}\log n)}$-size lower bound for $\Sigma\Pi\Sigma\Pi^{[\sqrt{n}]}$ circuits are concerned. In [CM14a], we sought a conceptual reason for this behaviour. We identify a simple combinatorial property such that any $n^{O(1)}$-variate polynomial of degree d that satisfies it would require $2^{\Omega(\sqrt{d}\log n)}$ size $\Sigma\Pi\Sigma\Pi^{[\sqrt{d}]}$ circuits. We call this *Leading Monomial Distance Property*. In particular, it does not matter whether the polynomial is easy (i.e. in VP) or hard (i.e. the polynomial is in VNP but not known to be in VP). As a result of this abstraction we present a simple *unified* analysis of the bounded fan-in depth four circuit size lower bounds for $\mathsf{NW}_{n,r}(X)$ and $\mathsf{IMM}_{n,d}(X)$. It is now well understood that the current known techniques can not help with proving strong enough lower bounds. However, some very interesting lower bounds were proved in the recent past over some related models (cf. [Sap15]).

1.3.3. On the limits of depth reduction at depth three over finite fields

Looking at the current exciting chasm at depth four and the subsequent results against it, one might wonder if it is sufficient to prove circuit size lower bounds at depth three. A recent result justifies that it is true over $\mathbb{Q}$ [GKKS13]. Gupta et al. showed that such a chasm exists even at depth three, over $\mathbb{Q}$ (or fields of large characteristic) [GKKS13].

Theorem 1.3.1 ([GKKS13]). *Let f be a polynomial of degree d over n variables and is computed by an arithmetic circuit C of size s. Then f can also be computed by a non-homogeneous $\Sigma\Pi\Sigma$ circuit of size* $\exp\left(O\left(\sqrt{d\log(sd)\log n}\right)\right)$.

Strong enough lower bounds here can prove valiant's conjecture. However, we proved that such a chasm is not present over fixed-size[2] finite fields [CM14b]. A strong exponential circuit size lower bound is already known for the determinant polynomial [GK98]. We improved the situation on the lower bounds front by proving depth three circuit

[2]By fixed size, we mean that the size of the finite field does not grow with the number of variables.

size lower bounds for two seemingly different polynomials (one which is built from the combinatorial design of Nisan and Wigderson [NW94] and the other is again the iterated matrix multiplication polynomial) over fixed-size finite fields. In fact, such a bound can be obtained for any polynomial, either in VP or VNP, if it exhibits a combinatorial property that we call the *downward closed monomials*. This strengthens the already known lower bounds of Grigoriev and Karpinski [GK98], over fixed-size finite fields.

1.3.4. Powering circuits

Let us now consider the arithmetic circuits that just use the addition gates and the powering gates. A powering gate takes in the tuple (f, d) as the input and output the polynomial f^d. It is denoted by $\wedge$. The expression in the form of the sum of powers of linear polynomials is a depth three powering circuit, a restriction of the general depth three circuits. Since there exists a depth three powering circuit of size at most $n2^{n-1}$ to compute a monomial, this computational model is *universal*[3]. In fact, there is a powering circuit of depth $(d+1)$ and size $O\left(n^d \cdot 2^{d \cdot n^{\frac{1}{d}}}\right)$. In [CKW11] Chen et al. posed the following open question.

Question 1.3.2 ([CKW11])**.** *Can the monomial $x_1 x_2 \dots x_n$ be efficiently computed by a constant depth powering circuit?*

This is the question that motivates our work. We show that there are at least two restricted classes of depth five powering circuits can not efficiently compute the monomial. Our models encapsulate the models considered by Engels et al. [ERS16].

Saptharishi[4] [Sap15] observed that the monomial $x_1 x_2 \dots x_n$ has non-trivial $\Sigma\wedge\Sigma\wedge$ and $\Sigma\wedge\Sigma\wedge\Sigma$ circuits of size $2^{O(\sqrt{n})}$ (cf. Lemma 6.4.1). Ideally, we would like to prove matching lower bounds but the current state of affairs is far away from that. But we, along with the work of Engels et al. [ERS16] make partial progress.

Engels et al. [ERS16] consider the depth five powering circuits which compute the polynomials of the form $g = \sum_{i=1}^{s} f_i^{\alpha_i}$ where $f_i = (\ell_1^d + \dots + \ell_n^d + c_i)$, ℓ_is are homoge-

[3] A computational model is said to be universal if it can simulate *any* computation.
[4] Saptharishi attributes the observation to Forbes.

neous linear polynomials, c_is are non-zero elements, and d is at least 21. We improve upon [ERS16] by extending it to a model that encapsulates their model and more.

1.4. Determinantal complexity

Let us recall that the determinant polynomial characterizes the class VP with respect to the quasi-polynomial projections.

Definition 1.4.1. *The determinantal complexity of a polynomial f, over n variables, is the minimum m such that there are affine linear polynomials $A_{k,\ell}$, $1 \leq k,\ell \leq m$ defined over the same set of variables and $f = \det((A_{k,\ell})_{1\leq k,\ell\leq m})$. It is denoted by* $\mathrm{DetComp}(f)$. ◊

To resolve Valiant's hypothesis, proving $\mathrm{DetComp}(\mathsf{Perm}_n) = n^{\omega(\log n)}$ is sufficient. By improving upon [vzG86, vzG87, Mes89, Cai90], Mignon and Ressayre [MR04] proved that $\mathrm{DetComp}(\mathsf{Perm}_n) \geq \frac{n^2}{2}$ over the fields of characteristic zero, using algebraic geometry. Subsequently, Cai et al. [CCL08] extended the result of Mignon and Ressayre to all fields of characteristic $\neq 2$. They also provided a simpler analysis of the same. Recently, Yabe [Yab15] improved upon the work of Mignon and Ressayre to show a lower bound of $(n-1)^2+1$ over the reals.

In this thesis, we observed that some basic determinantal complexity lower bounds can be proved using just the dimension of the partial derivative space as a complexity measure. In particular, we observed that $\mathrm{DetComp}(\mathsf{IMM}_{n,d}) \geq \frac{dn}{5}$ and $\mathrm{DetComp}(\mathsf{NW}_{n,\varepsilon n}) \geq \Omega(n^{1.5})$. Further, in [CM14a] we showed that the former can be improved to $\mathrm{DetComp}(\mathsf{IMM}_{n,d}) \geq \frac{dn}{2}$. Similar to the approach of [CCL08] and [MR04], we also use the the rank of Hessian matrix as our main technical tool.

Since $\mathsf{IMM}_{n,d}(X)$ has an algebraic branching program of size $O(dn)$ [Nis91], from the above theorem it follows that $\mathrm{DetComp}(\mathsf{IMM}_{n,d}(X)) = \Theta(dn)$. This improves upon the earlier bound of $\Theta(n)$ for the determinantal complexity of the iterated matrix multiplication polynomial for any constant $d > 1$ [Jan11].

1.5. Tensor rank

Proving arithmetic formula size lower bounds has been difficult as well. The best known lower bound for an explicit polynomial is quadratic [Kal85]. In a surprising result, Raz [Raz10] showed a connection between the size of the arithmetic formula computing a set multilinear polynomial and the rank of the *corresponding* tensor (over an appropriate range of parameters). More formally, let $f \in \mathbb{F}[X_1, X_2, \ldots, X_d]$ be a set multilinear polynomial of degree d over nd variables, where for every $i \in [d]$, X_i is a subset of variables of size n. In a natural way, f can be viewed as a tensor $f : [n]^d \to \mathbb{F}$. Raz [Raz10] showed if $\omega(1) \leq d \leq O(\log n / \log\log n)$ and f is computed by an arithmetic formula of size poly(n), then the rank of f as a tensor is far from n^{d-1} (the trivial upper bound). We know that there exist tensors $g : [n]^d \to \mathbb{F}$ of rank n^{d-1}/d. We showed that the tensor rank of f is far from n^{d-1} if f is computed by a *homogeneous* formula of polynomial size and d is such that $\omega(1) \leq d \leq n^{o(1)}$. We do this through a structured analysis of the chasms at depth four for homogeneous formulas. For general formulas, this helps extend the range of parameters from $\omega(1) \leq d \leq O(\log n / \log\log n)$ to $\omega(1) \leq d \leq O(\log n)$. This follows from the combination of our extension with the fact that formulas can be homogenized without much overhead when $d = O(\log n)$.

1.6. Organization of the thesis

Chapter 3 introduces the chasms and we provide an alternate analysis of the chasms of Agrawal and Vinay [AV08], Koiran [Koi12], and Tavenas [Tav15]. We further extend the chasm to homogeneous formulas. In Chapter 4, we show an unified analysis of the exponential lower bounds against the depth four circuits of bounded fan-in which compute explicit polynomials. In Chapter 5, we show that there is no chasm at depth three over fixed size finite fields. In Chapter 6, we show that there exist two restricted models of depth five powering circuits that can not compute a monomial. In Chapter 7, we show a *tight* lower bound on the determinantal complexity of the iterated matrix multiplication polynomial. In Chapter 8, we extend work of Raz [Raz10] and show a connection

between the size of a homogeneous formula computing a set-multilinear polynomial and the rank of the corresponding tensor, for an appropriate range of parameters. In Chapter 9, we conclude by laying out the directions and the associated challenges that lay ahead.

2

Preliminaries

2.1. Notation

- For any integer n, we shall use $[n]$ to denote the set $\{1,...,n\}$.
- We shall use the underbar with the letters to denote a tuple. For example $\underline{i} = (i_1, i_2, \ldots, i_n)$. We use the bold face to denote a vector, for example $\mathbf{v} = (v_1, v_2, \ldots, v_n)^T$.
- For $\mathbf{i} = \{i_1, i_2, \ldots, i_n\}$ and $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$, we shall use $\mathbf{x}^{\mathbf{i}}$ to denote the monomial $x_1^{i_1} x_2^{i_2} \ldots x_n^{i_n}$. The degree of the monomial $m = \mathbf{x}^{\mathbf{i}}$, denoted by $\deg(m)$ is $(i_1 + i_2 + \cdots + i_n)$. Similarly we use $\mathrm{var}(m)$ to denote the variables occurring in that monomial. That is, $\mathrm{var}(m) = \left\{x_{i_j} : i_j > 0,\ j \in [n]\right\}$.
- We use $\mathbf{x}^{=\ell}$ to denote all the monomials whose degree is equal to ℓ, that is, $\left\{\mathbf{x}^{\mathbf{i}} : i_1 + i_2 + \cdots + i_n = \ell\right\}$. Similarly, we use $\mathbf{x}^{\leq \ell}$ to denote the set of monomials whose degree is at most ℓ, $\left\{\mathbf{x}^{\mathbf{i}} : i_1 + i_2 + \cdots + i_n \leq \ell\right\}$.
- For a monomial $m = \mathbf{x}^{\mathbf{i}}$, we use $\partial_m f$ to denote the partial derivative of f with respect to the monomial m.

$$\partial_m f := \frac{\partial}{\partial x_{i_1}}\left(\frac{\partial}{\partial x_{i_2}} \ldots \left(\frac{\partial f}{\partial x_{i_n}}\right)\right)$$

- For a set of monomials S, we use $\partial_S f$ to denote the set of partial derivatives of f

with respect to every monomial in S, $\partial_S f := \{\partial_m f : m \in S\}$.

- We use $\partial^{=k} f$ to denote the set of partial derivatives of f with respect to monomials of degree equal to k, $\partial^{=k} f := \partial_{\mathbf{x}^{=k}} f$.

- For a set of polynomials $\{f_1, f_2, \ldots, f_m\}$, we use $\langle f_1, f_2, \ldots, f_m \rangle$ to denote the ideal generated by them.

$$\langle f_1, f_2, \ldots, f_m \rangle = \{a_1 f_1 + a_2 f_2 + \cdots + a_m f_m : a_1, \ldots, a_m \in \mathbb{F}[X]\}.$$

  We use $\langle f_1, f_2, \ldots, f_m \rangle_{\leq \ell}$ to denote the $\leq \ell$-degree combinations of $f_1, \ldots, f_m$.

$$\langle f_1, f_2, \ldots, f_m \rangle = \{a_1 f_1 + a_2 f_2 + \cdots + a_m f_m : \forall i \in [m], a_i \in \mathbb{F}[X]; \deg(a_i) \leq \ell\}.$$

- For a polynomial f, by $f|_{x \leftarrow a}$, we mean that every appearance of x in f is replaced by a. $f|_{\underline{\mathbf{x}} \leftarrow \underline{\mathbf{a}}}$ and $f|_{\mathbf{x} \leftarrow \mathbf{a}}$ are similarly defined.

- For a set X, we $X = Y \sqcup Z$ use to denote the partition of X into two disjoint sets Y and Z.

Arithmetic circuits

- Without loss of generality, we assume that the output gate is an addition gate and the circuits are layered with alternating addition and multiplication gates. We use Σ to represent a layer of addition gates and Π to represent a layer of multiplication gates. $\Sigma\Pi$ represents a polynomial expressed in the form of the sum of monomials. $\Sigma\Pi\Sigma$ represents polynomials that are expressible as sum of products of linear forms, also called depth three circuits. Similarly, we can define the depth four circuits $\Sigma\Pi\Sigma\Pi$.

- We use a superscript of a symbol $\{\Sigma, \Pi\}$ to denote the bound on the fan-in of the gates in the layer. For example, $\Pi^{[m]}$ denotes the layer of multiplication gates with

a fan-in of at most m. $\Sigma\Pi^{[D]}\Sigma\Pi^{[t]}$ represents depth four circuits whose multiplication gates have a fan-in bound of D and t respectively.

- We use the term hom in the superscript of a symbol $\{\Sigma, \Pi\}$ to mean that the operation is over homogeneous components feeding into the gates in that layer. For example, $\Sigma^{[\text{hom}]}$ represents the summation of homogeneous polynomials.

2.2. Approximation of the binomial coefficients

The following beautiful lemma (from [GKKS14]) is the key to the asymptotic estimates required for the lower bound analyses.

Lemma 2.2.1 (Lemma 6, [GKKS14]). *Let $a(n), f(n), g(n) : \mathbb{Z}_{\geq 0} \to \mathbb{Z}_{\geq 0}$ be the integer valued functions such that $(f+g) = o(a)$. Then,*

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

Proposition 2.2.2. *For all n, k such that $k \leq 0.5n$, the following bounds for the binomial coefficient $\binom{n}{k}$ hold.*

1. $\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{e\cdot n}{k}\right)^k$.
2. $\binom{n}{k} \sim 2^{n\cdot H\left(\frac{k}{n}\right)}$ *where H is the binary entropy function.*

2.3. Polynomial families

In this section, we shall define the polynomials that we talk about in this thesis.

2.3.1. Determinant

Definition 2.3.1. *The determinant polynomial of a $n \times n$ matrix is defined as*

$$\mathsf{Det}_n = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} x_{i,\sigma(i)}$$

where σ varies over S_n, a set of all the permutations of $[n]$ and $\operatorname{sgn}(\sigma)$ is the sign of the permutation σ. ◊

Valiant showed that the determinant polynomial is complete for the class VP with respect to quasi-polynomial projections [Bür00]. It is also known that determinant has a polynomial sized circuit, rather a polynomial sized *algebraic branching program* [MV97].

2.3.2. Permanent

Definition 2.3.2. *The permanent polynomial of a $n \times n$ matrix is defined as*

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

where σ varies over S_n, a set of all the permutations of $[n]$. ◊

Valiant showed that the permanent polynomial is complete for the class VNP over all fields of characteristic $\neq 2$ (cf. [Bür00]).

2.3.3. Nisan Wigderson polynomial

Definition 2.3.3 (Nisan-Wigderson polynomial)**.** *For integers $n > 0$ ranging over prime powers and an integer r, we define a polynomial family $\left\{\mathsf{NW}_{n,r}\right\}$ as follows.*

$$\mathsf{NW}_{n,r}(X) = \sum_{a(z) \in \mathbb{F}_n[z]} x_{1a(1)} x_{2a(2)} \cdots x_{na(n)}$$

where $a(z)$ runs over all univariate polynomials of degree $< r$ and $X = \left\{x_{ij} : (i,j) \in [n] \times [n]\right\}$. ◊

Proposition 4 in [Val79] tells us that if there is a polynomial time algorithm to test if the coefficient of a given monomial is 1 in the given polynomial $P(X) \in \mathbb{F}[X]$ with $\{0,1\}$ coefficients, then $P(X)$ is in VNP over $\mathbb{F}$. Given any monomial m over X, we can decide in polynomial time if it is a monomial in the polynomial $\{\mathsf{NW}_{n,r}\}_{n>0}$ by checking if it *conforms* to a univariate polynomial of degree at most r. Thus, $\{\mathsf{NW}_{n,r}\}_{n>0}$ is in VNP.

2.3.4. Iterated matrix multiplication

Definition 2.3.4 (Iterated matrix multiplication polynomial)**.** *The iterated matrix multiplication polynomial over d generic $n \times n$ matrices $X_1, X_2, \ldots, X_d$ is the $(1,1)$th entry in the product of these matrices. More formally, let $X_1, X_2, \ldots, X_d$ be d generic $n \times n$ matrices over disjoint sets of variables. For any $k \in [d]$, let $x_{ij}^{(k)}$ be the variable in X_k indexed by $(i,j) \in [n] \times [n]$. Then the iterated matrix multiplication polynomial, denoted by the family $\{\mathsf{IMM}_{n,d}\}$, is defined as follows.*

$$\mathsf{IMM}_{n,d}(X) = \sum_{i_1, i_2, \ldots, i_{d-1} \in [n]} x_{1 i_1}^{(1)} x_{i_1 i_2}^{(2)} \ldots x_{i_{(d-2)} i_{(d-1)}}^{(d-1)} x_{i_{(d-1)} 1}^{(d)}.$$

◊

This is a canonical polynomial for the *algebraic branching programs* and thus is in VP.

3

Depth reduction for arithmetic circuits

3.1. Introduction

Brent [Bre74] first showed that the arithmetic formulas of size s and of unrestricted depth can be transformed in to arithmetic formulas of depth $O(\log s)$ and of size $\text{poly}(s)$. This construction is very specific to the formulas and it does not extend to the arithmetic circuits that are not formulas. For the arithmetic circuits, Hyafil [Hya79] showed that any arithmetic circuit of polynomial size can be transformed into an arithmetic circuit of depth $O(\log^2 n)$ and of quasi-polynomial size. By improving upon the result of Hyafil, Valiant et al. [VSBR83] proved that any arithmetic circuit of polynomial circuit size (say n^c for a constant c) can be simulated by a $O(\log^2 n)$ depth arithmetic circuit of polynomial size. This can be formalized as $\mathsf{VP} = \mathsf{VNC}^2$. Furthermore, the corresponding $O(\log^2 n)$ depth circuit is highly structured (see Theorem 3.2.1).

In a surprising result, Agrawal and Vinay [AV08] showed that proving exponential circuit size lower bounds for depth four circuits implies exponential circuit size lower bounds for general arithmetic circuits. In particular, given a polynomial (or sub-exponential) sized general arithmetic circuit, it can be transformed into a depth four circuit of sub-exponential size. This initiated the effort to prove exponential lower bounds for depth four circuits (cf. the survey by Saptharishi [Sap15]). Koiran [Koi12] and Tavenas [Tav15] carefully analyzed the chasm shown in [AV08] and came up with an *improved* depth reduction.

In this chapter we shall present an alternate proof to Tavenas' depth reduction

[CKSV16]. Using this new technique, we arrive at *structured* depth four circuits. The starting point of this proof is the surprising depth reduction result of Valiant et al. [VSBR83] (and Allender et al. [AJMV98]). We shall also prove a structured depth reduction theorem for homogeneous formulas. This builds on the log product decomposition theorem for the homogeneous formulas of Hrubes and Yehudayoff [HY11].

We shall use this extra structure of the depth four formulas presented in Section 3.3 to prove non-trivial bounds on the tensor rank of formulas computing *set-multilinear* polynomials, in Chapter 8.

3.2. Depth reduction for arithmetic circuits to depth four

Let us formally recall the result of Valiant et al. [VSBR83] and Allender et al. [AJMV98].

Theorem 3.2.1 ([VSBR83, AJMV98]). *Let f be a polynomial of degree d over n variables, and is computed by an arithmetic circuit C_0 of size s. Then there is an arithmetic circuit C that also computes f, of size $s' = \mathrm{poly}(s,n,d)$ and depth $O(\log d)$. Furthermore, C is homogeneous, all multiplication gates have fan-in at most 5, and if u is any multiplication gate of C, then all its children v satisfy $\deg(v) \leq \deg(u)/2$.*

The circuit C obtained from Theorem 3.2.1 has addition gates of unbounded fan-in. It could be as large as $O(s)$. We can transform this further to obtain a bounded fan-in arithmetic circuit C' of size $\mathrm{poly}(s,n,d)$ and depth at most $O(\log s \log d)$. If $s = n^{O(1)}$ then the depth of the circuit obtained is $O(\log^2 n)$.

By improving upon the results of Agrawal and Vinay [AV08], and Koiran [Koi12], Tavenas [Tav15] proved the following theorem.

Theorem 3.2.2 ([AV08, Koi12, Tav15]). *Let f be a polynomial of degree d over n variables and is computed by an arithmetic circuit C of size s. Then, for any $0 < t \leq d$, f can also be computed by a homogeneous $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit of top fan-in $s^{O(d/t)}$ and size $s^{O(t+d/t)}$.*

Here we present an alternate proof of the above theorem through which we can construct more structured depth four circuits.

Proof of Theorem 4.1.1. From Theorem 3.2.1, we can assume that the circuit C we start with is of depth $O(\log d)$. If g is a polynomial computed at an arbitrary intermediate node of C, then from the structure of C we have the following homogeneous expression.

$$g = \sum_{i=1}^{s} g_{i1} \cdot g_{i2} \cdot g_{i3} \cdot g_{i4} \cdot g_{i5} \tag{3.2.3}$$

where each g_{ij} is computed by a node in C as well, and $\deg(g_{ij}) \leq \deg(g)/2$. In particular, if g were the output gate of the circuit that computes the polynomial f, then the RHS may be interpreted as a $\Sigma\Pi\Sigma\Pi^{[d/2]}$ circuit of top fan-in s. Let $t \in \mathbb{N}$ be a parameter. To obtain a $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit eventually, we shall perform the following steps on the output gate:

1. For each summand $g_{i1} \dots g_{ir}$ in the RHS, pick the gate g_{ij} with the largest degree (in case of a tie, pick the gate with the smaller index j). If g_{ij} has a degree that is greater than t, then expand g_{ij} in-place using (3.2.3).
2. Repeat this process until all of the g_{ij}'s on the RHS have a degree of at most t.

Each iteration of the above procedure increases the top fan-in by a multiplicative factor of s. If we could show that in $O(d/t)$ iterations, all the factors in each of the summands on the RHS have a degree of at most t, then we would have obtained an $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit of top fanin $s^{O(d/t)}$ that computes f.

Let us label a factor g_{ij} *bad* if its degree is greater than $t/8$. To bound the number of iterations of the above mentioned procedure, we will count the number of bad factors in each summand. Since the procedure always maintains homogeneity, the number of bad factors in any summand can at most be $8d/t$ (i.e., not too many). We will now show that each iteration *increases* the number of bad factors by at least one and hence the number of iterations must be bounded on the above by $8d/t$.

In (3.2.3), if $\deg(g) = k$, the largest degree factor of any summand on the RHS is at least $k/5$ (since the sum of the degrees of the five factors must add up to k) and it continues to be bad if $k > t$. But the largest degree factor can have a degree of at most

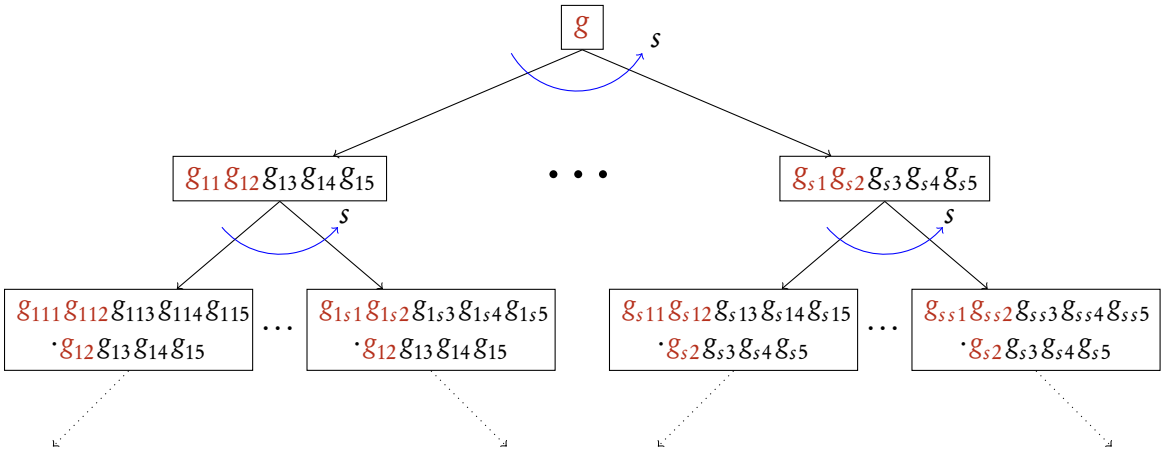

Figure 3.1.: Depth reduction analysis for arithmetic circuits

$k/2$. Hence the other four factors must together contribute at least $k/2$ to the degree. This implies that the second largest factor in each summand has a degree of at least $k/8$. This factor is bad too, if we started with a factor of degree greater than t. Therefore, as long as we are expanding factors of degree more than t using (3.2.3), we are guaranteed that its replacements have at least one additional bad factor. As argued earlier, we can never have more than $8d/t$ such factors in any summand and this bounds the number of iterations by $8d/t$.

Observe that the above procedure can be viewed as a tree, as described in Figure 3.1, where each node represents an intermediate summand in the iterative process. From (3.2.3) it is clear that the tree is s-ary. Furthermore, the number of *bad* factors strictly increases as we go down the tree (these are marked in red in Figure 3.1). Since the total number of bad factors in any node is at most $8(d/t)$, the depth of the tree can at most be $8(d/t)$. Therefore, the total number of leaves is at most $s^{(8d/t)}$. After the end of the procedure, each factor in every summand, is of degree at most t. Since any polynomial of degree at most t can be written as a sum of at most $n^{O(t)}$ monomials, the total size of the resulting $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit is at most $s^{O(t+d/t)}$ (since $s \geq n$). □

3.3. Depth reduction for homogeneous arithmetic formulas

In this section, we will show that the class of homogeneous formulas can be depth reduced to a more structured depth four circuit.

To quickly recap the earlier proof, we began with an equation of the form $g = \sum_i g_{i1} \cdot g_{i2} \cdot g_{i3} \cdot g_{i4} \cdot g_{i5}$ and recursively applied the same expansion on all the large degree factors in each of the summands. The only property that we really used was that in the above equation, there were at least two factors that had a *large* degree. In the case of homogeneous formulas, we shall now see that there are better expansions that we could use as a starting point.

Theorem 3.3.1 ([HY11])**.** *Let f be an n-variate polynomial of degree d that is computed by a homogeneous formula of size s. Then f can be expressed as*

$$f = \sum_{i=1}^{s} f_{i1} \cdot f_{i2} \cdots f_{ir} \tag{3.3.2}$$

where

1. *the expression is homogeneous,*
2. *for each i, j, we have $\left(\frac{1}{3}\right)^j d \leq \deg(f_{ij}) \leq \left(\frac{2}{3}\right)^j d$ and $r = \Theta(\log d)$,*
3. *each f_{ij} is also computed by a homogeneous formula of size at most s.*

With this, we are ready to prove a more structured depth reduction for homogeneous formulas.

Theorem 3.3.3. *Let f be a homogeneous polynomial of degree d over n variables which is computed by a homogeneous formula of size s. Then for any $0 < t \leq d$, f can also be computed by a homogeneous $\Sigma\Pi^{[a]}\Sigma\Pi^{[t]}$ formula of top fan-in at most $s^{10(d/t)}$ where*

$$a > \frac{1}{10}\frac{d}{t}\log t.$$

The resulting depth four circuit is more structured in the sense that the multiplication gates at the second layer have a much larger fan-in (by a factor of $\log t$). In Theorem 4.1.1,

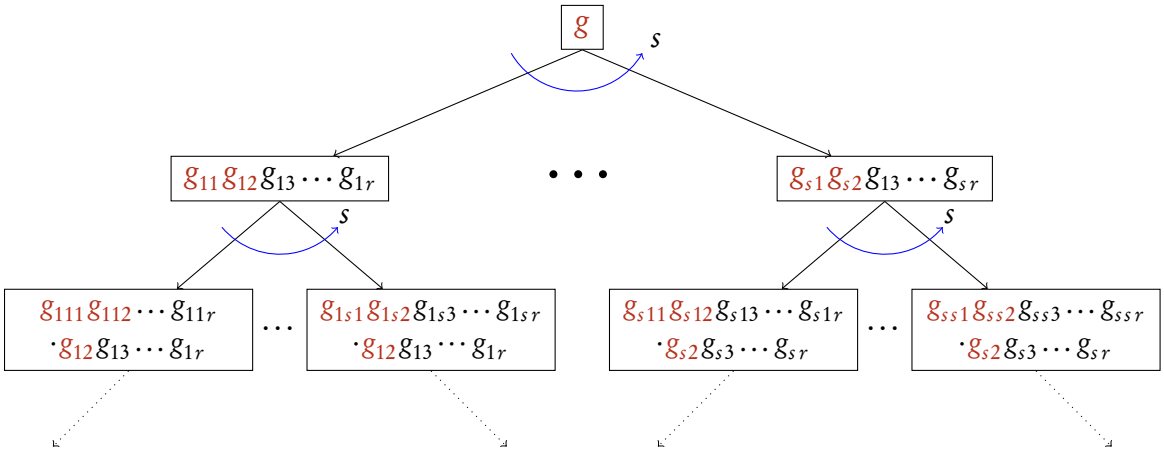

Figure 3.2.: Depth reduction analysis for homogeneous formulas

we only know that the polynomials feeding into these multiplication gates have a degree of at most t. Theorem 3.3.1 states that if we were to begin with a homogeneous formula, then those polynomials of degree at most t can be factorized further to give $\Theta((d/t)\log t)$ *non-trivial* polynomials instead of $\Theta(d/t)$ as obtained in Theorem 4.1.1.

Proof of Theorem 3.3.3. We start with equation (3.3.2) which is easily seen to be a homogeneous $\Sigma\Pi\Sigma\Pi^{[2d/3]}$ circuit with top fan-in s:

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

To obtain a $\Sigma\Pi^{[\Theta((d/t)\log t)]}\Sigma\Pi^{[t]}$ circuit eventually, we shall perform the following steps on the output gate:

1. For each summand $f_{i1}\ldots f_{iu}$ in the RHS, pick the gate f_{ij} with the largest degree (in case of a tie, pick the one with the smaller index j). If f_{ij} has a degree of more than t, expand that f_{ij} in-place using (3.3.2).

2. Repeat this process until all f_{ij}'s on the RHS have a degree of at most t.

Each iteration increases the top fan-in by a factor of s. As long as we expand the factors of degree $k > t$ using (3.3.2), we are guaranteed that each new summand has at least one more factor of degree at least $k/9 > t/9$.

To bound the number of iterations of the above procedure, we use the following potential function: the number of factors of degree strictly greater than $t/9$ in a summand. A factor that is of degree $k > t$ and which is expanded using (3.3.2) contributes at least two factors of degree at least $k/9 > t/9$ per summand. Thus, the net increase in the potential function per iteration is at least 1. Since this is a homogeneous computation, there can be at most $9d/t$ such factors of degree $> t/9$. Thus, the number of iterations must be bounded by $9d/t$ thereby yielding a $\Sigma\Pi\Sigma\Pi^{[t]}$ of top fan-in at most $s^{O(d/t)}$ and size $s^{O(t+d/t)}$. This argument is similar to the argument presented in the proof of Theorem 4.1.1.

We now argue that the fan-in of every product gate at the second level in the $\Sigma\Pi\Sigma\Pi^{[t]}$ circuit obtained is $\Theta(d/t \log t)$. To this end, we shall now show that we require $\Theta(d/t)$ iterations to make sure that all the factors have a degree of at most t. We will say that a factor is *small* if its degree is at most t or *big* otherwise. To prove a lower bound on the number of iterations, we shall use a different potential function: the total degree of all the big factors.

Given the geometric progression of the degrees in Theorem 3.3.1, we can easily see that the total degree of all the small factors in any summand is bounded on the above by $3t$. Hence, the total degree of all the big factors is $d - 3t$. But whenever (3.3.2) is applied on a big factor, we introduce several small degree factors with total degree of at most $3t$. Hence, the potential drops by at most $3t$ per iteration. This implies that we require at least $(d/3t)$ iterations to make the potential function 0, a state in which all the factors are small.

Since every expansion via (3.3.2) introduces at least $(\log_3 t)$ non-trivial factors, it would then follow that every summand at the end has $\frac{1}{(3\log 3)}\frac{d}{t}\log t > \frac{1}{10}\frac{d}{t}\log t$ non-trivial factors. □

Part II.

Constant Depth Arithmetic Circuit Lower Bounds

4

Lower bounds for depth four circuits of bounded fan-in: unified analysis

Tavenas has recently proved that any $n^{O(1)}$-variate and degree n polynomial in VP can be computed by a depth-4 circuit of size $2^{O(\sqrt{n}\log n)}$ [Tav15]. So, to prove VP $\neq$ VNP it is sufficient to show that an explicit polynomial in VNP of degree n requires $2^{\omega(\sqrt{n}\log n)}$ size depth-4 circuits. Soon after Tavenas' result, for two different explicit polynomials, depth-4 circuit size lower bounds of $2^{\Omega(\sqrt{n}\log n)}$ have been proved (see [KSS14] and [FLMS14]). In particular, using combinatorial design Kayal et al. [KSS14] construct an explicit polynomial in VNP that requires depth-4 circuits of size $2^{\Omega(\sqrt{n}\log n)}$ and Fournier et al. [FLMS14] show that the iterated matrix multiplication polynomial (which is in VP) also requires $2^{\Omega(\sqrt{n}\log n)}$ size depth-4 circuits. In this chapter, we identify a simple combinatorial property such that any polynomial f that satisfies this property would achieve a similar depth-4 circuit size lower bound. In particular, it does not matter whether f is in VP or in VNP. As a result, we get a simple unified lower bound analysis for the above mentioned polynomials.

4.1. Introduction

In a surprising result, Agrawal and Vinay [AV08] showed that proving exponential circuit size lower bounds for depth four circuits implies exponential circuit size lower bounds for general arithmetic circuits. In particular, given a polynomial (or sub-

exponential) sized general arithmetic circuit, it can be transformed into a depth four circuit of sub-exponential size. This initiated the effort to prove exponential lower bounds for depth four circuits (cf. the survey by Saptharishi [Sap15] for a series of related results). Koiran [Koi12] and Tavenas [Tav15] carefully analyzed the chasm shown in [AV08] and came up with an *improved* depth reduction.

Theorem 4.1.1 ([AV08, Koi12, Tav15]). *Let f be a polynomial of degree d over n variables and is computed by an arithmetic circuit C of size s. Then, for any $0 < t \leq d$, f can also be computed by a homogeneous $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuit of top fan-in $s^{O(d/t)}$ and size $s^{O(t+d/t)}$.*

The above theorem tells us that proving a size lower bound of $n^{\omega(d/t)}$ against the $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ would imply super polynomial circuit size lower bounds against general circuits. Towards proving such lower bounds Gupta et al. [GKKS14] proved a lower bound of $2^{\Omega(n/t)}$ against $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuits computing the determinant or the permanent polynomial over a $n \times n$ matrix. They used the dimension of the *shifted partial derivate* space as the complexity measure. Kayal [Kay12] introduced this measure to prove exponential circuit size lower bounds against the depth four circuits of the form of sums of powers of constant degree homogeneous polynomials, which compute the monomial $x_1 x_2 \dots x_n$. Kayal, Saha and Saptharishi [KSS14] pushed the bound of Gupta et al. [GKKS14] to $N^{\Omega(d/t)}$ for $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuits computing an explicit polynomial in VNP of degree d over N variables. This candidate polynomial is based on the combinatorial designs of Nisan and Wigderson [NW94].

This gives us two avenues to explore further.

1. Improve the depth reduction analysis of Tavenas [Tav15].

2. Prove better size lower bounds of the order of $n^{\omega(d/t)}$ for $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuits computing explicit polynomials in VNP of degree d over $n^{O(1)}$-variables.

In another surprising result Fournier et al. [FLMS14] proved that a *matching* bound could also be obtained for a polynomial in VP, the iterated matrix multiplication polynomial. This tells us that the bound of Tavenas is tight (up to a constant in the exponent)

and thus *rules out* the approach 1. However, it is important to note that the constant in the exponent of the bound proved by Kayal, Saha and Saptharishi [KSS14] or the one by Fournier et al., is weaker than the constant in the bound of Tavenas [Tav15].

One of the main motivations of our study comes from this tantalizing fact that two seemingly different polynomials $\mathsf{NW}_{n,r} \in \mathsf{VNP}$ and $\mathsf{IMM}_{n,n} \in \mathsf{VP}$ behave very similarly as far as the $2^{\Omega(\sqrt{n}\log n)}$-size lower bound for $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ circuits are concerned. In this chapter we seek a conceptual reason for this behaviour. We identify a simple combinatorial property such that any $n^{O(1)}$-variate polynomial of degree d that satisfies it would require $2^{\Omega(\sqrt{d}\log n)}$ size $\Sigma\Pi^{[O(\sqrt{d})]}\Sigma\Pi^{[\sqrt{d}]}$ circuits. We call this the *Leading Monomial Distance Property*. In particular, it does not matter whether the polynomial is easy (i.e. in VP) or hard (i.e. the polynomial is in VNP but not known to be in VP). As a result of this abstraction we present a simple *unified* analysis of the bounded fan-in depth four circuit size lower bounds for the $\mathsf{NW}_{n,r}$ and $\mathsf{IMM}_{n,d}$ polynomials. Formally, we prove the following.

Theorem 4.1.2. *Let f be a $n^{O(1)}$-variate polynomial of degree n. Let there be $s \geq n^{\delta k}$ (δ is any constant > 0) different polynomials in $\langle \partial^{=k}(f)\rangle$ for $k = \varepsilon\sqrt{n}$ such that any two of their leading monomials have pair-wise distance of at least $\Delta \geq \frac{n}{c}$ for any constant $c > 1$, and $0 < \varepsilon < \frac{1}{40c}$. Then any depth-4 $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ circuit that computes f must be of size $e^{\Omega_{\delta,c}(\sqrt{n}\ln n)}$.*

It is now well understood that the current known techniques can not help with approach 2. However, some very interesting lower bounds were proved in the recent past over some related models (cf. [Sap15]).

4.2. Preliminaries

In this section we shall recall the notion of shifted partial derivatives from [Kay12, GKKS14, KSS14] and define the combinatorial property that is that crux to our arguments.

Shifted partial derivatives

For a monomial $\mathbf{x}^{\mathbf{i}} = x_1^{i_1} x_2^{i_2} \dots x_n^{i_n}$, let $\partial^{\mathbf{i}} f$ be the partial derivative of f with respect to the monomial $\mathbf{x}^{\mathbf{i}}$. The degree of the monomial is denoted by $|\mathbf{i}|$ where $|\mathbf{i}| = (i_1 + i_2 + \cdots + i_n)$. We recall the following definition of shifted partial derivatives from [GKKS14].

Definition 4.2.1. *Let $f(X) \in \mathbb{F}[X]$ be a multivariate polynomial. The span of the ℓ-shifted k-th order derivatives of f, denoted by $\langle \partial^{=k} f \rangle_{\leq \ell}$, is defined as*

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

We denote by $\dim(\langle \partial^{=k} f \rangle_{\leq \ell})$ the dimension of the vector space $\langle \partial^{=k} f \rangle_{\leq \ell}$. ◊

Let $\succ$ be any admissible monomial ordering. The *leading monomial* of a polynomial $f(X) \in \mathbb{F}[X]$, denoted by $\mathsf{LM}(f)$ is the largest monomial $\mathbf{x}^{\mathbf{i}} \in f(X)$ under the order $\succ$. The next lemma follows directly from Proposition 11 and Corollary 12 of [GKKS14].

Lemma 4.2.2. *For any multivariate polynomial $f(X) \in \mathbb{F}[X]$,*

$$\dim(\langle \partial^{=k} f \rangle_{\leq \ell}) \geq \#\{\mathbf{x}^{\mathbf{i}} \cdot \mathsf{LM}(g) : \mathbf{i}, \mathbf{j} \in \mathbb{Z}^n_{\geq 0} \text{ with } |\mathbf{i}| \leq \ell, |\mathbf{j}| = k, \text{ and } g \in \mathbb{F}\text{-}span\{\partial^{\mathbf{j}} f\}\}$$

In [KSS14], the following upper bound on the dimension of the shifted partial derivative space of the polynomials computed by $\Sigma\Pi^{[D]}\Sigma\Pi^{[t]}$ circuits was shown.

Lemma 4.2.3 (Lemma 4, [KSS14])**.** *If $C = \sum_{i=1}^{s'} Q_{i1} Q_{i2} \dots Q_{iD}$ where each $Q_{ij} \in \mathbb{F}[X]$ is a polynomial of degree bounded by t. Then for any $k \leq D$,*

$$\dim(\langle \partial^{=k}(C) \rangle_{\leq \ell}) \leq s' \binom{D}{k} \binom{N + \ell + k(t-1)}{N}$$

Leading monomial distance property

To define the Leading Monomial Distance Property, we first define the notion of distance between two monomials.

Definition 4.2.4. *Let m_1, m_2 be two monomials over a set of variables. Let S_1 and S_2 be*

the (multi)-sets of variables corresponding to the monomials m_1 and m_2 respectively. The distance $\mathrm{dist}(m_1, m_2)$ *between the monomials m_1 and m_2 is the* $\min\{|S_1| - |S_1 \cap S_2|, |S_2| - |S_1 \cap S_2|\}$ *where the cardinalities are the order of the (multi)-sets.* ◊

We say that a $n^{O(1)}$-variate and n-degree polynomial has the Leading Monomial Distance Property, if the leading monomials of a *large subset* ($\approx n^{\delta k}$) of its span of the derivatives (of order $\approx k$) have *good pair-wise distance* for a suitable parameter k.

4.3. Unified analysis

In this section, we first prove a simple combinatorial lemma which we believe is the crux of the best known bounded fan-in depth four circuit size lower bound results. In fact, the lower bounds on the size of $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ circuits computing the polynomials $\mathsf{NW}_{n,r}$ and $\mathsf{IMM}_{n,n}$ follow easily from this lemma by suitably setting the parameters.

Lemma 4.3.1. *Let $m_1, m_2, \ldots, m_s$ be the monomials over N variables such that* $\mathrm{dist}(m_i, m_j) \geq \Delta$ *for all $i \neq j$. Let M be the set of monomials of the form $m_i m'$ where $1 \leq i \leq s$ and m' is a monomial of length at most ℓ over the same set of N variables. Then, the cardinality of M is at least $\left(sB - s^2\binom{N+\ell-\Delta}{N}\right)$ where $B = \binom{N+\ell}{N}$.*

Proof. Let B_i be the set of all monomials $m_i m'$ where m' is a monomial of length at most ℓ. It is easy to see that $|B_i| = \binom{N+\ell}{N}$. We would like to estimate $|\cup_i B_i|$. Using the principle of inclusion and exclusion, we get $|\cup_{i=1}^{s} B_i| \geq \sum_{i\in[s]} |B_i| - \sum_{i,j\in[s], i\neq j} |B_i \cap B_j|$.

Now we estimate the upper bound for $|B_i \cap B_j|$ such that $i \neq j$. Consider the monomials m_i and m_j in B_i and B_j respectively. For m_i and m_j to match, m_i should contain at least Δ variables from m_j and similarly m_j should contain at least Δ variables from m_i. The rest of the at most $(\ell - \Delta)$ degree monomials should be identical in m_i and m_j. The number of such monomials over N variables is at most $\binom{N+\ell-\Delta}{N}$. Thus, $|B_i \cap B_j| \leq \binom{N+\ell-\Delta}{N}$.

Then the total number of monomials of the form $m_i m'$ for all $i \in [s]$ where m' is a

monomial of length at most ℓ is lower bounded as follows.

$$|\cup_{i=1}^{s} B_i| \geq sB - s^2\binom{N+\ell-\Delta}{N} = sB\left(1-\frac{s}{B}\binom{N+\ell-\Delta}{N}\right)$$

□

We use the above lemma to prove Theorem 4.1.2 of this chapter. Even though we prove the bounds against $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ circuits computing $n^{O(1)}$-variate polynomials of degree n, we state the following theorem with some generality in terms of the parameter t.

Theorem 4.3.2. *Let f be a $N = d^{O(1)}$-variate polynomial of degree d. Let there be at least $d^{\delta k}$ (δ is any constant > 0) different polynomials in $\langle \partial^{=k}(f)\rangle$ for $k = \varepsilon\frac{d}{t}$ such that any two of their leading monomials have a distance of at least $\Delta \geq \frac{d}{c}$ for any constant $c > 1$, and $0 < \varepsilon < \frac{1}{40c}$. Then any depth-4 $\Sigma\Pi^{[O(d/t)]}\Sigma\Pi^{[t]}$ circuit that computes f must be of size $e^{\Omega_{\delta,c}(\frac{d}{t}\ln N)}$.*

Proof. Consider a set of $s = d^{\delta k}$ polynomials $f_1, f_2, \ldots, f_s \in \langle \partial^{=k}(f)\rangle$ such that $\mathrm{dist}(\mathsf{LM}(f_i), \mathsf{LM}(f_j)) \geq d/c$ for all $i \neq j$. We denote by m_i, the leading monomial $\mathsf{LM}(f_i)$. We now invoke Lemma 4.3.1 with the parameters $s = d^{\delta k}, \Delta = d/c$. Let N be the number of variables in f. From Lemma 4.3.1, we know that $|\cup_{i=1}^{s} B_i| \geq sB\left(1-\frac{s}{B}\binom{N+\ell-\Delta}{N}\right)$. To get a good lower bound for $|\cup_{i=1}^{s} B_i|$, we need to upper bound $\frac{s}{B}\binom{N+\ell-\Delta}{N}$. Let us bound it by an inverse polynomial in n by suitably choosing ℓ. We set $\frac{s\binom{N+\ell-\Delta}{N}}{\binom{N+\ell}{N}} \leq \frac{1}{p(d)}$ where $p(d)$ is a polynomial in d.

After simplification, we get $s\frac{(N+\ell-\Delta)!}{(N+\ell)!}\frac{\ell!}{(\ell-\Delta)!} \leq \frac{1}{p(d)}$. Using Lemma 2.2.1 we tightly estimate the subsequent computations. In particular, we always choose the parameter ℓ such that $\Delta^2 = o(N+\ell)$. This also shows that the error term given by Lemma 2.2.1 is always asymptotically zero and we need not worry about it.

We now apply Lemma 2.2.1 to derive $s\left(\frac{\ell}{N+\ell}\right)^{\Delta} \leq \frac{1}{p(d)}$ or equivalently $s\left(\frac{1}{1+\frac{N}{\ell}}\right)^{\Delta} \leq \frac{1}{p(d)}$. We use the inequality $1+x > e^{x/2}$ for $0 < x < 1$ to lower bound $\left(1+\frac{N}{\ell}\right)^{\Delta}$ by $e^{\frac{N\Delta}{2\ell}}$. Thus, it is enough to choose ℓ in a way that $s \cdot p(d) \leq e^{\frac{N\Delta}{2\ell}}$ or equivalently $\ell \leq \frac{N\Delta}{2\ln(s\cdot p(d))}$. By

fixing $p(d) = d^2$ and substituting for the parameters k and Δ, we get $\ell \leq \frac{Nt}{4c\delta\varepsilon \ln d}$. From Lemma 4.2.2, we get that the dimension of $\langle \partial^{=k} f \rangle_{\leq \ell} \geq (1 - \frac{1}{d^2})\, s\binom{N+\ell}{N}$.

Combining this with Lemma 4.2.3, we get $s' \geq \frac{(1-\frac{1}{d^2})s\binom{N+l}{N}}{\binom{D}{k}\binom{N+l+k(t-1)}{N}}$. Suppose we choose ℓ such that $(kt-k)^2 = o(\ell)$. Then, by applying Lemma 2.2.1 we can easily show the following.

$$s' \geq \frac{s\left(1-\frac{1}{d^2}\right)}{\binom{D}{k}(1+\frac{N}{l})^{(kt-k)}} \geq \frac{d^{\delta k}\left(1-\frac{1}{d^2}\right)}{\binom{D}{k}e^{\frac{N}{\ell}kt}}$$

Since $D = O(d/t)$ and $k = \varepsilon d/t$, we can estimate $\binom{D}{k}$ to be $e^{O_\varepsilon\left(\frac{d}{t}\right)}$ by Shannon's entropy estimate for binomial coefficients. To get the required lower bound it is sufficient to choose ℓ such that $\frac{Nkt}{\ell} < (0.1)\delta k \ln d$. By comparing the lower and upper bounds of ℓ, we can fix ε such that $\varepsilon < \frac{1}{40c}$. Since ε depends only on c, we can infer that $s' = e^{\Omega_{\delta,c}\left(\frac{d}{t}\ln d\right)} = e^{\Omega_{\delta,c}\left(\frac{d}{t}\ln N\right)}$. □

The above proof clearly goes through even if we set $\frac{Nkt}{\ell} < \mu\delta k \ln d$ for any $0 < \mu < 1$, and choose $\varepsilon < \frac{\mu}{4c}$.

4.4. Lower bounds for explicit polynomials

In this section we shall apply Theorem 4.3.2 to two explicit polynomials, $\mathsf{NW}_{n,r}$ which is a polynomial in VNP and $\mathsf{IMM}_{n,d}$ which is a polynomial in VP, to derive exponential lower bounds against the depth-4 $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ circuits computing them.

4.4.1. Nisan Wigderson polynomial

Now we derive the depth-4 circuit size lower bound for $\mathsf{NW}_{n,r}$ polynomial by a simple application of Theorem 4.3.2 where $d = n$ and $t = \sqrt{n}$.

Corollary 4.4.1. *For $0 < \varepsilon < 1/80$, any $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ circuit computing the polynomial $\mathsf{NW}_{n,r}(X)$ must be of size $2^{\Omega(\sqrt{n}\log n)}$ where $r = \varepsilon\sqrt{n}$.*

Proof. Recall that $\mathsf{NW}_{n,\varepsilon}(X) = \sum_{a(z)\in\mathbb{F}[z]} x_{1a(1)}x_{2a(2)}\dots x_{na(n)}$ where $\mathbb{F}$ is a finite field of size n and $a(z)$ is a univariate polynomial of degree $\leq r-1$ where $r = \varepsilon\sqrt{n}$. Notice that any two monomials can intersect in at most $r-1$ variables.

We differentiate the polynomial $\mathsf{NW}_{n,r}(X)$ with respect to the first $k = r = \varepsilon\sqrt{n}$ variables of each monomial. After differentiation, we get n^k monomials of length $(n-k)$ each. Since they are constructed from the image of univariate polynomials of degree at most $(k-1)$, the distance Δ between any two monomials $\geq n-2k > n/2$. So to get the required lower bound we invoke Theorem 4.3.2 with $\delta = 1$ and $c = 2$. □

4.4.2. Iterated matrix multiplication polynomial

Next we derive the lower bound on the size of the depth-4 circuit computing $\mathsf{IMM}_{n,n}$.

Corollary 4.4.2. *Any depth-4* $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ *circuit computing the* $\mathsf{IMM}_{n,n}(X)$ *polynomial must be of size* $2^{\Omega(\sqrt{n}\log n)}$.

Proof. Recall that $\mathsf{IMM}_{n,n}(X) = \sum_{i_1,i_2,\dots,i_{n-1}\in[n]} x^{(1)}_{1i_1}x^{(2)}_{i_1i_2}\dots x^{(n-1)}_{i_{(n-2)}i_{(n-1)}}x^{(n)}_{i_{(n-1)}1}$. It is a polynomial over $(n-2)n^2+2n$ variables. We fix the following lexicographic ordering on the variables of the set of matrices $\{X_1, X_2, \dots, X_n\}$ as follows: $X_1 \succ X_2 \succ X_3 \succ \dots \succ X_n$ and in any X_i the ordering is $x^{(i)}_{11} \succ x^{(i)}_{12} \succ \dots \succ x^{(i)}_{1n} \succ \dots \succ x^{(i)}_{n1} \dots \succ x^{(i)}_{nn}$.

Choose a prime p such that $\frac{n}{2} \leq p \leq n$. Consider the set of univariate polynomials $a(z) \in \mathbb{F}_p[z]$ of degree at most $(k-1)$ for $k = \varepsilon\sqrt{n}$ where ε is a small constant to be fixed later in the analysis. Consider a set of $2k$ of the matrices $X_2, X_{3+\frac{n}{4k}}, \dots, X_{2k+1+\frac{(2k-1)n}{4k}}$ such that they are $n/4k$ distance apart. Clearly $2k+1+\frac{(2k-1)n}{4k} < n$. For each univariate polynomial a of degree at most $(k-1)$, define a set $S_a = \{x^{(2)}_{1,a(1)}, x^{(3+\frac{n}{4k})}_{2,a(2)}, \dots, x^{(2k+1+\frac{(2k-1)n}{4k})}_{2k,a(2k)}\}$. Number of such sets is at least $\left(\frac{n}{2}\right)^k$ and $|S_a \cap S_b| < k$ for $a \neq b$. Now we consider a polynomial $f(X)$ which is a restriction of the polynomial $\mathsf{IMM}_{n,n}(X)$. By restriction, we simply mean that a few variables of $\mathsf{IMM}_{n,n}(X)$ are fixed to some elements from the field and the rest of the variables are left untouched. We define the restriction as follows.

$$x^{(q)}_{ij} = 0 \text{ if } r + \frac{(r-2)n}{4k} < q < (r+1) + \frac{(r-1)n}{4k} - 1 \text{ for } 2 \leq r \leq 2k \text{ and } i \neq j.$$

The rest of the variables are left untouched. Next we differentiate the polynomial $f(X)$ with respect to the sets of variables S_a indexed by the polynomials $a(z) \in \mathbb{F}[z]$. Consider the leading monomial of the derivatives with respect to the sets S_a for all $a(z) \in \mathbb{F}[z]$. Since $|S_a \cap S_b| < k$, it is straightforward to observe that the distance between any two leading monomials is at least $k \cdot \frac{n}{4k} = \frac{n}{4}$. The intuitive justification is that whenever there is a difference in S_a and S_b, that difference can be stretched to a distance $\frac{n}{4k}$ because of the restriction that eliminates the non diagonal entries.

Now we prove the lower bound for the polynomial $f(X)$ by applying Theorem 4.3.2. Notice that $f(X)$ is a $n^{O(1)}$-variate polynomial of degree n such that there are at least $(n/2)^k > n^{\frac{1}{4}(2k)}$ different polynomials in $\langle \partial^{=2k}(f) \rangle$ such that any two of their leading monomials have distance $\Delta \geq n/4$. So we set the parameters $\delta = 1/4$ and $c = 4$ in Theorem 4.3.2. A simple calculation shows that the parameter ε can be fixed to something $< 1/320$.

Since $f(X)$ is a restriction of $\mathsf{IMM}_{n,n}(X)$, any lower bound for $f(X)$ is a lower bound for $\mathsf{IMM}_{n,n}(X)$ too. Otherwise, if $\mathsf{IMM}_{n,n}(X)$ has a $2^{o(\sqrt{n}\log n)}$ sized $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ circuit, then we get a $2^{o(\sqrt{n}\log n)}$ sized $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ circuit for $f(X)$ by substituting for the variables according to the restriction. □

5

On the limits of depth reduction to depth three over small finite fields

In this chapter, for an explicit polynomial in VP (over fixed-size finite fields), we prove that any depth-3 circuit computing it must be of size $2^{\Omega(n \log n)}$. The explicit polynomial that we consider is the iterated matrix multiplication polynomial of n generic matrices of size $n \times n$. The importance of this result is that over fixed-size fields there is *no depth reduction technique* that can be used to compute all the $n^{O(1)}$-variate and n-degree polynomials in VP by depth 3 circuits of size $2^{o(n \log n)}$. The result of Grigoriev and Karpinski [GK98] can only rule out such a possibility for depth-3 circuits of size $2^{o(n)}$.

We also give an example of an explicit polynomial ($\mathsf{NW}_{n,\varepsilon}(X)$) in VNP (which is not known to be in VP), for which any depth-3 circuit computing it (over fixed-size fields) must be of size $2^{\Omega(n \log n)}$.

5.1. Introduction

In a recent breakthrough, Gupta et al. [GKKS14] have proved that over $\mathbb{Q}$ (or fields of large characteristic), if an $n^{O(1)}$-variate polynomial of degree d is computable by an arithmetic circuit of size s, then it can also be computed by a depth three $\Sigma\Pi\Sigma$ circuit of size $2^{O\left(\sqrt{d \log(ds) \log n}\right)}$. Through this, they prove the existence of a $\Sigma\Pi\Sigma$ circuit of size $2^{O(\sqrt{n} \log n)}$ computing the determinant polynomial of an $n \times n$ matrix (over $\mathbb{Q}$). Before this result, no depth three circuit for Determinant of size smaller than $2^{O(n \log n)}$ was

known (over any field of characteristic $\neq 2$).

The situation is very different over *fixed-size finite fields*. Grigoriev and Karpinski proved that over fixed-size finite fields, any depth three circuit for the determinant polynomial of a $n \times n$ matrix must be of size $2^{\Omega(n)}$ [GK98]. This does rule out depth reduction to depth three to a circuit of size $2^{O\left(\sqrt{d \log(ds) \log n}\right)}$ (rather $2^{o(n)}$) for a polynomial of degree n over $n^{O(1)}$ variables. Although Grigoriev and Karpinski proved the lower bound result only for the determinant polynomial, it is now a folklore result that a modification of their argument can show a similar depth three circuit size lower bound for the permanent polynomial as well (cf. [Sap15]). Over any field, Ryser's formula for Permanent gives a $\Sigma\Pi\Sigma$ circuit of size $2^{O(n)}$ [Rys63]. Thus, for the permanent polynomial the depth three complexity (over fixed-size finite fields) is essentially $2^{\Theta(n)}$.

The result of [GKKS14] is obtained through an ingenious depth reduction technique but their technique is tailored to the fields of zero characteristic. In particular, the main technical ingredients of their proof are the well-known monomial formula of Fischer [Fis94] and the duality trick of Saxena [Sax08]. These techniques do not work over finite fields. Looking at the contrasting situation over $\mathbb{Q}$ and the fixed-size finite fields, a natural question is to ask whether one can find a new depth reduction technique over fixed-size finite fields such that any $n^{O(1)}$-variate and degree n polynomial in VP can also be computed by a $\Sigma\Pi\Sigma$ circuit of size $2^{o(n \log n)}$.

Question 5.1.1. *Over any fixed-size finite field $\mathbb{F}_q$, is it possible to compute any $n^{O(1)}$-variate and n-degree polynomial in* VP *by a $\Sigma\Pi\Sigma$ circuit of size $2^{o(n \ln n)}$?*

Note that any $n^{O(1)}$-variate and n-degree polynomial can be trivially computed by a $\Sigma\Pi\Sigma$ circuit of size $2^{O(n \log n)}$ by writing it explicitly as a sum of all $n^{O(n)}$ possible monomials. We give a negative answer to the aforementioned question by showing that over fixed-size finite fields, any $\Sigma\Pi\Sigma$ circuit computing the iterated matrix multiplication polynomial (which is in VP for any field) must be of size $2^{\Omega(n \log n)}$. More precisely, we prove that any $\Sigma\Pi\Sigma$ circuit computing the iterated matrix multiplication polynomial of n generic $n \times n$ matrices (denoted by $\mathsf{IMM}_{n,n}(X)$), must be of size $2^{\Omega(n \log n)}$.

Previously, Nisan and Wigderson [NW97] proved a size lower bound of $\Omega(n^{d-1}/d!)$

for any homogeneous ΣΠΣ circuit computing the iterated matrix multiplication polynomial over d generic $n \times n$ matrices. Kumar et al. [KMN13] improved the bound to $\Omega(n^{d-1}/2^d)$. These results work over any field. Over fields of zero characteristic, Shpilka and Wigderson proved a near quadratic lower bound for the size of depth three circuits computing the trace of the iterated matrix multiplication polynomial [SW01]. Recently, Kayal et al. [KST16] and Balaji et al. [BLS16] proved near-cubic circuit size lower bounds for $\Theta(n)$-variate, $\Theta(n)$-degree polynomials in VNP and VP respectively.

Similar to the situation at depth 4, we also give an example of an explicit n^2-variate and n-degree polynomial in VNP (which is not known to be in VP), the Nisan Wigderson polynomial such that over fixed-size finite fields, any depth three ΣΠΣ circuit computing it must be of size $2^{\Omega(n \log n)}$. In fact, from our proof idea it will be clear that the strong depth three size lower bound results that we show for $\mathsf{NW}_{n,\varepsilon n}(X)$ and $\mathsf{IMM}_{n,n}(X)$ polynomials are not really influenced by the fact that the polynomials are either in VNP or VP. Rather, the bounds are determined by a combinatorial property of the subspaces generated by a carefully chosen set of derivatives. Further, this bound holds for any polynomial f such that a subspace of its partial derivative space satisfies the combinatorial property we describe in this chapter. Our main theorem of this chapter is the following.

Theorem 5.1.2. *Over any fixed-size finite field $\mathbb{F}_q$, any depth three ΣΠΣ circuit computing the polynomials $\mathsf{NW}_{n,\varepsilon n}$ or $\mathsf{IMM}_{n,n}$ must be of size at least $2^{\delta n \log n}$, where the parameters δ and $\varepsilon (< 1/2)$ are in $(0,1)$ and depend only on q.*

We shall fix the values of δ and ε suitably later. As an important consequence of the above theorem, we have the following corollary.

Corollary 5.1.3. *Over any fixed-size finite field $\mathbb{F}_q$, there is no depth reduction technique that can be used to compute all the $n^{O(1)}$-variate and n-degree polynomials in VP by depth three circuits of size $2^{o(n \log n)}$.*

The result of [GK98] only says that over fixed-size finite fields, not all the $n^{O(1)}$-variate and n-degree polynomials in VP can be computed by ΣΠΣ circuits of size $2^{o(n)}$. Our

main theorem (Theorem 5.1.2) can also be viewed as the first quantitative improvement over the result of [GK98].

Proof Idea

Our proof technique is quite simple and it borrows ideas mostly from the proof technique of Grigoriev and Karpinski [GK98]. A recurring notion in many papers related to ΣΠΣ circuits is the notion of *rank* of a product gate. Let $T = L_1 L_2 \dots L_d$ be a product gate such that each L_i is an affine linear polynomial over the underlying field. By rank of T, one simply means the maximum rank of the homogeneous linear system corresponding to set of affine linear polynomials $\{L_1, L_2, \dots, L_d\}$.

Over fixed-size finite fields, ΣΠΣ circuits enjoy a nice property that the derivatives of the high rank product gates can be eliminated except for a few erroneous points (denoted by E). This property was first observed by Grigoriev and Karpinski in [GK98]. The intuition is simple. If a product gate has many linearly independent functions, then it is likely that a large number of linear functions will be set to zero if we randomly substitute the variables with elements from the field. Then the derivatives (of relatively low order) of the polynomial obtained from the product gate will disappear on a random point with very high probability.

To quantify the notion of *high rank*, Grigoriev and Karpinski fixed a threshold for the rank of the product gates. Since they were looking for a $2^{\Omega(n)}$ lower bound for the Determinant of a $n \times n$ matrix and the rank of the entire derivative space of of the determinant polynomial is $2^{O(n)}$, it was natural for them to fix the threshold to be $\Theta(n)$. Since the dimension of the derivative spaces of the polynomial families $\{\mathsf{NW}_{n,\varepsilon n}\}_{n>0}$ and $\{\mathsf{IMM}_{n,n}\}_{n>0}$ is $2^{\Omega(n \log n)}$, it is possible for us to choose the threshold for the rank of the product gates to be $\Theta(n \log n)$. This allows us to bound the size of the error set meaningfully.

An overview of the result of Grigoriev and Karpinski

We now give a high level description of the proof technique in [GK98] to motivate our proof strategy. Roughly speaking, they consider the space H spanned by $\Theta(n)$ order derivatives of the determinant. From the rank analysis on the circuit side, they get that the dimension of the space of functions that may not be zero outside the error set E is bounded. Grigoriev and Karpinski then considered the group of invertible matrices G of order $n \times n$ over $\mathbb{F}_q$. For any $g \in G$, they define a $\mathbb{F}_q$-linear operator $T_g : H \to H$ by the formula $(T_g(f))(a) = f(ga)$. The fact that the derivative space of the determinant polynomial of a $n \times n$ matrix is invariant under $\mathbb{GL}_n(\mathbb{F}_q)$ action was crucially used in defining the map. They then prove that the full invertible group G can be covered by taking only a few translates of $G \setminus E$ from G. This was done by appealing to a graph theoretic lemma of Lovász [Lov75]. Now, they observe that we can the bound the dimension of functions that are not zero over all of $\mathbb{GL}_n(\mathbb{F}_q)$ to a quantity smaller than $\dim(H)$. This shows us that there exists a nonzero function in H that evaluates to zero on the entire group G if the determinant polynomial is computed by the depth three circuit. Since the elements in H are only multilinear polynomials, they finally prove that it is impossible to have such a function in H by showing that no nonzero multilinear polynomial can vanish over the entire group G.

An overview of our result

The group symmetry based argument of [GK98] is tailored to the determinant polynomial and it can not be directly applied to the polynomials that we consider. The main technical contribution of this work is to replace the group symmetry based argument by a new argument that makes the proof strategy robust enough to handle the families of polynomials that we consider. We carefully choose a subspace H (of sufficiently large dimension) of the derivative spaces of these polynomials which have an additional structure. The subspace H is spanned by a *downward closed* set of monomials. Let $\mathbb{F}_q$ be the finite field and N be the number of the variables in the polynomial under consideration. The basic idea is to prove that the dimension of the space H is strictly more than the

dimension of the set of functions in H which do not evaluate to zero over the entire space $\mathbb{F}_q^N$ when the polynomial considered is computed by any depth three circuit.

To implement this, we define a linear map $T_u : H \to H$ such that $T_u(f(X)) = f(X-u)$ for any function $f : \mathbb{F}_q^N \to \mathbb{F}_q$ and $u \in \mathbb{F}_q^N$. The map is well-defined by the *downward closed* structure of the generating set for H. Also the map T_u is one to one for any $u \in \mathbb{F}_q^N$. Then from the structure of the depth three circuits, we observe that the dimension of the functions that are non zero over $\mathbb{F}_q \setminus E$ is small. The argument so far for the derivative space of depth three circuits is not over the entire space $\mathbb{F}_q^N$ where as the argument for the polynomial is over $\mathbb{F}_q^N$. Similar to the argument in [GK98], we use the graph theoretic lemma of Lovász [Lov75] to prove that the entire space $\mathbb{F}_q^N$ can be covered by only a few translates of $\mathbb{F}_q^N \setminus E$. Then it is simple to observe that the dimension of the functions that are not zero over all of $\mathbb{F}_q$ is small compared to the dimension of H. As a consequence we get that there is a nonzero multilinear polynomial in H must evaluate to zero over $\mathbb{F}_q^N$, which is not possible by the combinatorial nullstellensatz [Alo99].

5.2. Preliminaries

We will first recall the following theorem from [Alo99].

Theorem 5.2.1. *Let $f(x_1, x_2, \ldots, x_n)$ be a polynomial in n variables over an arbitrary field $\mathbb{F}$. Suppose that the degree of f as a polynomial in x_i is at most t_i, for $1 \leq i \leq n$ and let $S_i \subseteq \mathbb{F}$ such that $|S_i| \geq t_i + 1$. If $f(a_1, a_2, \ldots, a_n) = 0$ for all n-tuples in $S_1 \times S_2 \times \cdots \times S_n$, then $f = 0$.*

We use the following combinatorial property to carefully chose a subspace of the partial derivative space of the polynomials of interest.

Definition 5.2.2. *A set of multilinear monomials M is said to be downward closed if the following property holds. If $m(X) \in M$ and a multilinear monomial $m'(X)$ is such that $\mathrm{var}(m'(X)) \subseteq \mathrm{var}(m(X))$, then $m'(X) \in M$.* ◊

Let us now consider a downward closed set of monomials M over N variables. These monomials can be viewed as functions from $\mathbb{F}_q^N$ to $\mathbb{F}_q$. W.l.o.g, we assume that the con-

stant function is also in M (constant function corresponds to a monomial with an empty set of variables). Let H be the subspace spanned by these functions in M.

Lemma 5.2.3. *Let $m_1(X), m_2(X), \ldots, m_k(X)$ be any set of k distinct multilinear monomials in $\mathbb{F}_q[x_1, x_2, \ldots, x_N]$. For $1 \leq i \leq k$, let $f_i : \mathbb{F}_q^N \to \mathbb{F}_q$ be the function corresponding to the monomial $m_i(X)$, i.e. $f_i(X) = m_i(X)$. Then, f_is are linearly independent in the q^N dimensional vector space over $\mathbb{F}_q$.*

Proof. If f_is are not linearly independent then $\sum_{i=1}^{k} \lambda_i f_i = 0$ for $\bar{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_k) \in \mathbb{F}_q^k \setminus \{\bar{0}\}$. Then, the nonzero multilinear polynomial $\sum_{i=1}^{k} \lambda_i m_i(X)$ evaluates to zero on $\mathbb{F}_q^N$, which contradicts Theorem 5.2.1. □

For any $u \in \mathbb{F}_q^N$, define an operator T_u such that $(T_u(f))(X) = f(X - u)$ for any function $f : \mathbb{F}_q^N \to \mathbb{F}_q$. The following proposition is simple to prove.

Proposition 5.2.4. *Let H be the subspace spanned by a downward closed set of monomials M over the set of variables $\{x_1, x_2, \ldots, x_N\}$. Then for any $u \in \mathbb{F}_q^N$, T_u is a linear map from H to H. Moreover, the map T_u is one-to-one for any $u \in \mathbb{F}_q^N$.*

Proof. Let $g(X)$ be an arbitrary function in H which can be expressed as follows: $g(X) = \sum_{i \geq 1} c_i m_i(X)$ where $m_i(X) \in M$, and $c_i \in \mathbb{F}_q$ for all $i \geq 1$.

$$(T_u(g))(X) = g(X - u) = \sum_{i \geq 1} c_i m_i(X - u).$$

It is sufficient to prove that $m(X - u) \in H$ where $m(X) \in M$. We can express $m(X - u)$ as follows.

$$m(X - u) = \sum_{S \subseteq \text{ var}(m(X))} c_S \prod_{x_r \in S} x_r.$$

where $c_S \in \mathbb{F}_q$. For every $S \subseteq \text{var}(m(X))$, $\prod_{x_r \in S} x_r \in M$ because M is downward closed. Since the choice of S was arbitrary, $m(X - u) \in H$. It is obvious that T_u is a linear map.

To see that T_u is a one-to-one map, it is just enough to observe that $T_u \circ T_{-u} = T_0$ where T_0 is an identity map. □

5.3. The Derivative Space of $\Sigma\Pi\Sigma$ Circuits Over Small Fields

In this section we fix the field $\mathbb{F}$ to be a fixed-size finite field $\mathbb{F}_q$. Let C be a $\Sigma\Pi\Sigma$ circuit of top fan-in s computing a $N = n^{O(1)}$-variate polynomial of degree n. Consider a Π gate $T = L_1 L_2 \dots L_d$ in C. Let r be the rank of the (homogeneous)-linear system over $\mathbb{F}_q$ corresponding to $\{L_1, L_2, \dots, L_d\}$ by viewing each L_i as a vector in $\mathbb{F}_q^{N+1}$. Fix a threshold for the rank of the system of linear functions $r_0 = \beta n \ln n$, where $\beta > 0$ is a constant to be fixed later in the analysis. In our application, the parameter N is at least n^2, so the threshold for the rank is meaningful. W.l.o.g, let $\{L_1, L_2, \dots, L_r\}$ be a set of affine linear forms in $\{L_1, L_2, \dots, L_d\}$ whose homogeneous system forms a maximal independent set of linear functions. The following analysis has been reworked from [GK98] to fix the parameters. It shows that the derivative space of a $\Sigma\Pi\Sigma$ circuit can be approximated by just the derivative space of the low rank product gates of the circuit over a large subset of $\mathbb{F}_q^N$.

Low rank gates : $r \leq r_0$

Over the finite field $\mathbb{F}_q$, we know that $a^q = a$ for any $a \in \mathbb{F}_q$ (due to Fermat). We express $T : \mathbb{F}_q^N \to \mathbb{F}_q$ as a linear combination of $\{L_1^{e_1} L_2^{e_2} \dots L_r^{e_r} : e_i < q \text{ for all } i \in [r]\}$. Since, the derivatives of all orders lie in the same space, the dimension of the set of partial derivatives of T of all orders is bounded by $q^r \leq q^{r_0}$.

High rank gates : $r > r_0$

Let the rank of a high rank gate T be $y\beta n \ln n$ where $y \geq 1$.

We assign values to the variables uniformly at random from $\mathbb{F}_q$ and compute the probability that at most n linearly independent functions evaluate to zero. Let X_a be the event that at most n linearly independent functions evaluate to zero at a.

$$\Pr_{a \in \mathbb{F}_q^N}[X_a] \leq \sum_{i=0}^{n} \binom{r}{i} \left(\frac{1}{q}\right)^i \left(1 - \frac{1}{q}\right)^{r-i} \leq n \binom{r}{n} \left(\frac{1}{q}\right)^n \left(1 - \frac{1}{q}\right)^{r-n}.$$

The above inequality follows from the fact that $r > 2n$ and thus the binomial terms under the summation are monotonically increasing. Hence, if we differentiate T with respect to any set of variables of size at most n and restrict all the variables to values from $\mathbb{F}_q$, the gate T may not vanish over a set of points E_T whose size is estimated below.

$$|E_T| \leq n\binom{r}{n}\left(\frac{1}{q}\right)^n\left(1-\frac{1}{q}\right)^{r-n}q^N.$$

Over all the gates, let E be the set of points over which some of the product gates with large rank may not evaluate to zero. Then by a union bound, we get that $|E| \leq s|E_T|$.

$$\begin{aligned}
|E| &\leq s \cdot n\binom{r}{n}\left(\frac{1}{q}\right)^n\left(1-\frac{1}{q}\right)^{r-n}q^N \\
&\leq s \cdot n\left(\frac{er}{n}\right)^n e^{-\frac{r-n}{q}}q^N \\
&= q^N s \cdot e^{n+n\ln\frac{r}{n}+\ln n-\frac{r-n}{q}} \\
&= q^N s \cdot e^{n+n\ln\frac{\gamma\beta n\ln n}{n}+\ln n-\frac{\gamma\beta n\ln n-n}{q}}.
\end{aligned}$$

To bound the above estimate meaningfully, we need $\frac{\ln s}{n\ln n}$ to be strictly less than $\frac{\gamma\beta n}{q}\ln n - n\ln\gamma$. That is, for some constant $\nu > 0$, we want the following to hold true.

$$\frac{\ln s}{n\ln n} - \frac{\gamma\beta}{q} + \frac{\ln\gamma}{\ln n} + \nu < 0. \tag{5.3.1}$$

Once we satisfy the relation given by the inequality (5.3.1), we can upper bound the size of E as $|E| < q^N\mu^{n\ln n}$ for some suitably fixed constant $\mu = e^{-\nu}$ and μ is between 0 and 1. Now it is clear that over $\mathbb{F}_q^N \setminus E$, the derivative space is spanned just by the derivatives of the low rank gates. We summarize it in the following lemma.

Lemma 5.3.2. *Let $\mathbb{F}_q$ be a fixed-size finite field. Then there exist constants $0 < \beta(q), \nu(q) < 1$ such that the following is true. Let C be a ΣΠΣ circuit of top fan-in s computing a $N = n^{O(1)}$-variate and n-degree polynomial $f(X)$ over the finite field $\mathbb{F}_q$. Further, $s, \beta(q), \nu(q)$*

satisfy the inequality 5.3.1. Then, there exists a set $E \subset \mathbb{F}_q^N$ of size at most $q^N \mu^{n \ln n}$ such that the dimension of the space spanned by the derivatives of order $\leq n$ of C restricted to $\mathbb{F}_q^N \setminus E$ is $\leq s\ q^{\beta n \ln n}$ where $\mu = e^{-\nu}$.

It is worth (re)-emphasizing that, when we consider the derivatives, what we really mean is the formal derivatives of C as polynomials. In the above lemma we view the derivatives as functions from $\mathbb{F}_q^N \to \mathbb{F}_q$. Then it follows from the above analysis that the dimension of the space spanned by the functions corresponding to the derivatives of order $\leq n$ of C restricted to $\mathbb{F}_q^N \setminus E$ is $\leq s\ q^{\beta n \ln n}$. This way of viewing derivatives either as *formal polynomials* or as *functions* is implicit in the work of [GK98]. We shall fix the parameters δ, β, and μ later such that they depend only on the field size q.

A Covering Argument

In this section, we adapt the covering argument of [GK98] where the covering argument was given over the set of invertible matrices. Here we adapt their argument suitably over the entire space $\mathbb{F}_q^N$.

Let H be the derivative space of any polynomial $f(X)$ which is computed by any $\Sigma\Pi\Sigma$ circuit of size s. Define the subspace $H_a := \{f \in H : f(a) = 0\}$ for $a \in \mathbb{F}_q^N$. Let us recall that E is the set of points over which some of the product gates with large rank may not evaluate to zero. Let the set of points $\mathbb{F}_q^N \setminus E$ be denoted by A. Then Lemma 5.3.2 says that in H, we get that the dimension of all the functions that are not all zero over all of A is at most $s\ q^{r_0}$. Formally, $\operatorname{codim}(\bigcap_{a \in A} H_a) < s\ q^{r_0}$. Let us recall that or any $u \in \mathbb{F}_q^N$, we defined an operator T_u such that $(T_u(f))(X) = f(X - u)$ for any function $f : \mathbb{F}_q^N \to \mathbb{F}_q$.

Proposition 5.3.3. *For any $u, a \in \mathbb{F}_q^N$, we have that $T_u(H_a) = H_{u+a}$.*

Proof. Let us recall that the map T_u is one-to-one. That is, $T_{-u} \circ T_u = T_0$ where T_0 is an identity map. It is easy to observe that $T_{-u}(H_{u+a}) = H_a$. □

Let $P = \bigcap_{a \in A} H_a$. Let $S \subset \mathbb{F}_q^N$ be a set such that we can cover the entire space $\mathbb{F}_q^N$ by

the shifts of A with the elements from S.

$$\bigcup_{u\in S} u + A = \mathbb{F}_q^N.$$

Now by applying the map T_u to P which is one-to-one, we get the following.

$$T_u(P) = \bigcap_{a\in A} T_u(H_a) = \bigcap_{b\in u+A} H_b.$$

By a further intersection over S, we get the following.

$$\bigcap_{u\in S} T_u(P) = \bigcap_{u\in S} \bigcap_{b\in u+A} H_b = \bigcap_{b\in\mathbb{F}_q^N} H_b. \tag{5.3.4}$$

From (5.3.4), we get the following estimate.

$$\operatorname{codim}\left(\bigcap_{b\in\mathbb{F}_q^N} H_b\right) = \operatorname{codim}\left(\bigcap_{u\in S} T_u(P)\right) \le |S|\ \operatorname{codim}(P) \le |S|\ s\ q^{r_0}. \tag{5.3.5}$$

The $\operatorname{codim}\left(\bigcap_{b\in\mathbb{F}_q^N} H_b\right)$ refers to the dimension of the set of functions in H which do not evaluate to zero over all the points in $\mathbb{F}_q^N$. Next, we show an upper-bound estimate for the size of the set S. This follows from a simple adaptation of the dominating set based argument given in [GK98].

Upper bound on the size of the set S

Consider the directed graph $G = (V, R)$ defined as follows. The points in $\mathbb{F}_q^N$ are the vertices of the graph. For $u_1, u_2 \in \mathbb{F}_q^N$, the edge $u_1 \to u_2$ is in R iff $u_2 = u_1 + b$ for any $b \in A$. Clearly the in-degree and out-degree of any vertex are equal to $|A|$. Now, we recall Lemma 2 of [GK98] to estimate the size of S.

Lemma 5.3.6 (Lovász, [Lov75]). *Let $G = (V, R)$ be a directed (regular) graph with $|V| = m$ vertices and with the in-degree and the out-degree of each vertex both equal to d. Then there exists a subset $U \subset V$ of size $O(\frac{m}{d}\log(d+1))$ such that for any vertex $v \in V$ there is*

a vertex $u \in U$ forming an edge $(u,v) \in R$.

Let c_0 be the constant fixed by the lemma in its $O()$ notation. By Lemma 5.3.6, we get the following estimate.

$$\begin{aligned}
|S| &\leq c_0 \frac{|\mathbb{F}_q^N|}{|A|} \log(|A|+1) \\
&\leq c_0 \frac{q^N}{q^N - |E|} \log(q^N - |E| + 1) \\
&\leq c_0 (\log q)\, N\, \frac{q^N}{q^N - |E|} \\
&= O(N). \qquad \text{(for fixed } q\text{)}
\end{aligned}$$

The last equation follows from the estimate for $|E|$ from the previous discussions. The following lemma summarizes the content of this section.

Lemma 5.3.7. *Let H be the space of partial derivatives of order at most n, of any N-variate polynomial computed by a $\Sigma\Pi\Sigma$ circuit of size s. Then for a suitable parameter r_0, the dimension of the set of functions in H that do not evaluate to zero over all points in $\mathbb{F}_q^N$ is upper bounded by $O(Nsq^{r_0})$.*

5.4. Derivative Spaces of the Polynomial Families

In this section, we study the derivative spaces of $\mathsf{NW}_{n,\varepsilon n}$ and $\mathsf{IMM}_{n,n}$ polynomials. Instead of considering the full derivative spaces, we focus on a set of carefully chosen derivatives and consider the subspaces spanned by them.

The derivative space of $\{\mathsf{NW}_{n,\varepsilon n}\}_{n>0}$ polynomial family

A set of variables $D = \{x_{i_1 j_1}, x_{i_2 j_2}, \ldots, x_{i_t j_t}\}$ is called an admissible set if the i_ks (for $1 \leq k \leq t$) are all distinct and $t \in [\varepsilon n, n]$. Let H be the subspace spanned by the set of the partial derivatives of the polynomial $\mathsf{NW}_{n,\varepsilon n}(X)$ with respect to the admissible sets of

variables. More formally,

$$H := \mathbb{F}_q\text{-span}\left\{\frac{\partial\, \mathsf{NW}_{n,\varepsilon n}(X)}{\partial D} : D \text{ is an admissible set of variables}\right\}.$$

Since the monomials of the $\mathsf{NW}_{n,\varepsilon n}(X)$ polynomial are defined by the univariate polynomials of degree $< \varepsilon n$, each partial derivative with respect to such a set D yields a multilinear monomial. If we choose ε such that $n - \varepsilon n > \varepsilon n$ (i.e. $\varepsilon < 1/2$), then after the differentiation, all the monomials of length $n - \varepsilon n$ are distinct. This follows from the fact that the monomials are generated from the image of the univariate polynomials of degree $< \varepsilon n$.

Let us treat these monomials as functions from $\mathbb{F}_q^{n^2} \to \mathbb{F}_q$. The Lemma 5.2.3 that the functions corresponding to any set of distinct monomials are linearly independent.

Consider the derivatives of $\mathsf{NW}_{n,\varepsilon n}(X)$ corresponding to the sets $\{x_{1a(1)}, x_{2a(2)}, \ldots, x_{\varepsilon n a(\varepsilon n)}\}$ for all univariate polynomials a of degree $< \varepsilon n$. From Lemma 5.2.3, it follows that $\dim(H) \geq n^{\varepsilon n} = e^{\varepsilon n \ln n}$. We can notice that the constant function $\mathbf{1} : \mathbb{F}_q^{n^2} \to \mathbb{F}_q$ given by $\forall x, \mathbf{1}(x) = 1$ is also in H. This corresponds to the derivatives of order n.

The derivative space of $\{\mathsf{IMM}_{n,n}\}_{n>0}$ polynomial family

For our application, we consider all those n such that $n = 2m$ where m ranges over the positive integers. Consider the set of matrices $X^{(1)}, X^{(3)}, \ldots, X^{(2m-1)}$ corresponding to the odd places. Let S be any set of m variables chosen as follows. Choose any variable from the first row of $X^{(1)}$ and choose any one variable from each of the matrices $X^{(3)}, \ldots, X^{(2m-1)}$. We call such a set S an admissible set.

If we differentiate $\mathsf{IMM}_{n,n}(X)$ with respect to two different admissible sets of variables S and S', then we get two different monomials of length m each. This follows from the structure of the monomials in the $\mathsf{IMM}_{n,n}(X)$ polynomial, whenever we fix two variables from $X^{(i-1)}$ and $X^{(i+1)}$, the variable from $X^{(i)}$ gets fixed. So the number of such monomials after differentiation is exactly $n^{2m-1} = e^{(n-1)\ln n}$.

Let m_S be the monomial obtained after differentiating $\mathsf{IMM}_{n,n}(X)$ by the set of variables in S and $\mathrm{var}(m_S)$ be the set of variables in m_S. Consider the derivatives of $\mathsf{IMM}_{n,n}(X)$ with respect to the following sets of variables $\{S \cup T : T \subseteq \mathrm{var}(m_S)\}$ where S ranges over all admissible sets.

Let H be the subspace spanned by these derivatives. More formally,

$$H := \mathbb{F}_q\text{-span}\left\{\frac{\partial\, \mathsf{IMM}_{n,n}(X)}{\partial D} : D = S \cup T; T \subseteq \mathrm{var}(m_S); S \text{ is an admissible set}\right\}.$$

As before, we can notice that the constant function $\mathbf{1}$ is in H. From Lemma 5.2.3, we know that $\dim(H) \geq e^{(n-1)\ln n}$. Now to unify the arguments for $\mathsf{NW}_{n,\varepsilon n}(X)$ and $\mathsf{IMM}_{n,n}(X)$ polynomials, we introduce the following notion.

We can easily observe that the derivative spaces that we select for $\mathsf{NW}_{n,\varepsilon n}(X)$ and $\mathsf{IMM}_{n,n}(X)$ are spanned by downward closed sets of monomials (Definition 5.2.2).

Lemma 5.4.1. *The generator sets for the derivative subspaces H for $\mathsf{NW}_{n,\varepsilon n}(X)$ and $\mathsf{IMM}_{n,n}(X)$ polynomials are downward closed.*

Proof. Let us consider the $\mathsf{NW}_{n,\varepsilon n}(X)$ polynomial first. Let $m \in H$ be any monomial and D be the admissible set such that $m = \frac{\partial\, \mathsf{NW}_{n,\varepsilon n}(X)}{\partial D}$. Let m' be any monomial such that $\mathrm{var}(m') \subseteq \mathrm{var}(m)$. Then $m' = \frac{\partial\, \mathsf{NW}_{n,\varepsilon n}(X)}{\partial D'}$ where $D' = D \cup (\mathrm{var}(m) \setminus \mathrm{var}(m'))$.

Similarly for the $\mathsf{IMM}_{n,n}(X)$ polynomial, consider any $m \in H$. Then $m = \frac{\partial\, \mathsf{IMM}_{n,n}(X)}{\partial D}$ and $D = S \cup T$ for an admissible set S and $T \subseteq \mathrm{var}(m_S)$. If m' is any monomial such that $\mathrm{var}(m') \subseteq \mathrm{var}(m)$, then $m' = \frac{\partial\, \mathsf{IMM}_{n,n}(X)}{\partial D'}$ where $D' = S \cup (T \cup (\mathrm{var}(m) \setminus \mathrm{var}(m')))$. Clearly $T \cup (\mathrm{var}(m) \setminus \mathrm{var}(m')) \subseteq \mathrm{var}(m_S)$. □

5.4.1. Obtaining the circuit size lower bound

In this section, based on the discussion above, we obtain a lower bound on the size of any $\Sigma\Pi\Sigma$ circuit computing the $\mathsf{NW}_{n,\varepsilon n}(X)$ and $\mathsf{IMM}_{n,n}(X)$ polynomials. We show that the dimension of the set of non zero functions in the derivative space of the polynomial computed by any $\Sigma\Pi\Sigma$ circuit of size at most $2^{\delta n \log n}$, is smaller than dimension of the set of the chosen derivative space of the polynomials we consider. If the depth three circuit

computes the polynomial under consideration, there exists a function f in the derivative space of the polynomial such that it evaluates to zero over all points in $\mathbb{F}_q^N$ which is not possible as per Theorem 5.2.1. Thus, we infer that any $\Sigma\Pi\Sigma$ circuit computing the $\mathsf{NW}_{n,\varepsilon n}(X)$ and $\mathsf{IMM}_{n,d}(X)$ must be of size greater than $2^{\delta n \log n}$, for a suitable parameter δ.

Fixing the parameters

Consider the inequality 5.3.1 which is $\frac{lns}{n\ln n} + \nu < \frac{y\beta}{q} - \frac{\ln y}{\ln n}$. For a parameter δ, fix $s = \exp(\delta n \ln n)$. Fix the values for β, δ, and ν as follows. Set $\beta = \frac{1}{10\ln q}, \delta = \frac{1}{20q\ln q}, \nu = \frac{3\delta}{4}$, and $\mu = e^{-\nu}$. Consider the function $g(y) = y - \frac{10q\ln q}{\ln n}\ln y - 0.75$. Since $g(y)$ is a monotonically increasing function (for n appropriately larger than a threshold value depending on q) which takes the value of 0.25 at $y = 1$, $g(y) > 0$ for $y \geq 1$ and thus for the chosen values of β and δ, $\frac{y\beta}{q} - \frac{\ln y}{\ln n} - \delta > \nu$ and thus $|E| \leq q^N \mu^{n\ln n}$.

Let us consider that the $\Sigma\Pi\Sigma$ circuit computes the $\mathsf{NW}_{n,\varepsilon}(X)$. From Section 5.4, we know that the dimension of the subspace formed by set of chosen derivatives for $\mathsf{NW}_{n,\varepsilon n}(X)$ is at least $e^{\varepsilon n \ln n}$. Consider the upper bound on $\operatorname{codim}\left(\bigcap_{b\in\mathbb{F}_q^N} H_b\right)$ given by the inequality 5.3.5. If we choose ε in such a way that $e^{\varepsilon n\ln n} > |S|\ s\ q^{r_0}$, then there will be a multilinear polynomial f in the chosen derivative space of $\mathsf{NW}_{n,\varepsilon n}(X)$ such that f will evaluate to zero over all points in $\mathbb{F}_q^N$.

$$\exp(\varepsilon n\ln n) > |S|\ s\ q^{r_0} = \exp(\delta n\ln n + (\beta\ln q)n\ln n + \ln N).$$

Considering the terms of the order of $n\ln n$ in the exponent, it is enough to choose $\varepsilon(<1/2)$ such that the following holds.

$$\varepsilon > \delta + \beta\ln q = \frac{1}{20q\ln q} + \frac{1}{10}.$$

Since the dimension of the subspace formed by set of chosen derivatives for $\mathsf{IMM}_{n,n}(X)$ is $\geq e^{(n-1)\ln n}$, the chosen values of β and δ clearly suffice.

Finally, we recall from Theorem 5.2.1 that no non-zero multilinear polynomial can

be zero over $\mathbb{F}_q^N$. That is, f can not be zero over all points in $\mathbb{F}_q^N$. This contradicts our assumption that the top fan-in of the $\Sigma\Pi\Sigma$ circuit is less than $2^{\delta n \log n}$. Thus, we get the main theorem.

Theorem 5.4.2. *For any fixed-size finite field $\mathbb{F}_q$, any depth three $\Sigma\Pi\Sigma$ circuit computing the polynomials $\mathsf{NW}_{n,\varepsilon n}$ or $\mathsf{IMM}_{n,n}$ must be of size at least $2^{\delta n \log n}$ where the parameters δ and $\varepsilon(< 1/2)$ are in $(0,1)$ and depend only on q.*

It is straightforward to observe that the lower bound analysis holds for any polynomial for which we can find a subspace (of sufficiently large dimension) of its derivative space spanned by a downward closed set of monomials.

6

Exponential lower bounds for the depth five powering circuits

Depth five powering circuits are arithmetic circuits of the form $\Sigma \wedge \Sigma \wedge \Sigma$ where 'Σ' and '$\wedge$' represent gates that compute sum and power of their inputs respectively. Such circuits compute polynomials of the form $\sum_{i=1}^{t} Q_i^{\alpha_i}$, where each Q_i is a sum of powers of linear polynomials. These circuits are a natural generalization of the well known class of depth three powering circuits ($\Sigma \wedge \Sigma$ circuits). In this chapter, we study the complexity of the monomial $x_1 x_2 \cdots x_n$ which is computed by some restricted classes of depth five powering circuits.

6.1. Introduction

A power symmetric polynomial of degree d over the variables $\{y_1, \ldots, y_m\}$ is the polynomial $P_m^d(y_1, \ldots, y_m) = y_1^d + \cdots + y_m^d$. For any integers $d, n > 0$, Ellison [Ell69] showed that there exists an integer m for every polynomial of degree d over n variables $X = \{x_1, x_2, \ldots, x_n\}$ such that it can be written as a projection[1] of P_m^d. In other words, there exists an integer m such that every polynomial can be expressed as the sum of d^{th} powers of m linear polynomials. For a polynomial f, the minimal such m is called the Waring rank of the polynomial f and it is denoted by $\text{wrk}(f)$. Fischer [Fis94] showed that $\text{wrk}(x_1 x_2 \ldots x_n)$ is at most 2^{n-1} by giving an explicit set of linear forms (cf. Proposition 6.2.5). Using the technique of partial derivatives, Saxena [Sax08] showed that

[1] A polynomial $f(x_1, \ldots, x_n)$ is said to be a projection of the polynomial $g(y_1, \ldots, y_m)$ if there exist linear polynomials $\{\ell_1, \ldots, \ell_m\}$ over $\mathbb{F}[x_1, \ldots, x_n]$ such that $f = g(\ell_1, \ldots, \ell_m)$.

$\text{wrk}(x_1x_2\dots x_n) \geq 2^{\Omega(n)}$ which is a linear factor away from the upper bound (cf. Chapter 10 of [CKW11]). Ranestad and Schreyer [RS00] proved that the Waring rank of the monomial is exactly 2^{n-1} using algebraic geometry. Recently, Balaji et al. [BLSS17] gave an elegant proof of the same using basic linear algebraic techniques.

Let us now consider the arithmetic circuits that use just the addition gates and the powering gates [2]. The expression in the form of the sum of powers of linear polynomials is a depth three powering circuit, a restriction of the general depth three circuits. Since there exists a depth three powering circuit of size at most $n2^{n-1}$ to compute a monomial, the computational model is *universal* for polynomial computations. In fact, there is a powering circuit of depth $(d+1)$ and size $O\left(n^d \cdot 2^{d \cdot n^{\frac{1}{d}}}\right)$ that computes $x_1x_2\dots x_n$. The afore mentioned lower bounds on the Waring rank imply a size lower bound of $2^{\Omega(n)}$ for any depth three powering circuit computing the monomial $x_1x_2\dots x_n$. In [CKW11] Chen et al. posed the following open question.

Question 6.1.1 ([CKW11])**.** *Can the monomial $x_1x_2\dots x_n$ be efficiently computed by a constant depth powering circuit?*

This is the question that motivates the work presented in this chapter. We show that there are at least two restricted classes of depth five powering circuits can not efficiently compute the monomial.

Saptharishi[3] [Sap15] observed that the monomial $x_1x_2\dots x_n$ has non-trivial $\Sigma\wedge\Sigma\wedge$ and $\Sigma\wedge\Sigma\wedge\Sigma$ circuits of size $2^{O(\sqrt{n})}$ (cf. Lemma 6.4.1). Ideally, we would like to prove matching lower bounds but the current state of affairs is far away from that. But we, along with the work of Engels et al. [ERS16] make partial progress.

Kayal [Kay12] using the technique of shifted partial derivatives proved an exponential bound of $2^{\Omega\left(\frac{n}{d}\right)}$ against any $\Sigma\wedge\Sigma^{[\text{hom}]}\Pi^{[d]}$ computing the monomial $x_1x_2\dots x_n$. If there is a $\Sigma\wedge\Sigma\Pi^{[d]}$ circuit of size s then there is a $\Sigma\wedge\Sigma^{[m]}\wedge^{[d]}\Sigma^{[\text{hom}]}$ circuit of size $m \cdot s$ where $m = 2^d \cdot \binom{n+d}{n}$. Thus, Kayal's bound implies an exponential size lower bound for the $\Sigma\wedge\Sigma^{[\text{hom},\ m]}\wedge^{[d]}\Sigma^{[\text{hom}]}$ computing the monomial when $d \leq \sqrt{\frac{n}{\log n}}$ and $m = 2^d \cdot \binom{n+d}{n}$.

[2] A powering gate takes in the tuple (f, d) as the input and output the polynomial f^d. It is denoted by $\wedge$.

[3] Saptharishi attributes the observation to Forbes.

Our results

Engels et al. [ERS16] consider the depth five powering circuits which compute the polynomials of the form $g = \sum_{i=1}^{s} f_i^{\alpha_i}$ where $f_i = (\ell_1^d + \cdots + \ell_n^d + c_i)$, ℓ_is are homogeneous linear polynomials, c_is are non-zero elements, and d is at least 21. They use the dimension of *projected multilinear derivatives* as the complexity measure to prove such a result.

Theorem 6.1.2 ([ERS16]). *Let $g = \sum_{i=1}^{s} f_i^{\alpha_i}$ where $f_i = P_n^{d_i}(\ell_{i1}, \ldots, \ell_{in}) + \beta_i$, either $d_i = 1$ or $d_i \geq 21$ and $\ell_{i1}, \ldots, \ell_{in}$ are homogeneous linear forms for every i. If $g = x_1 x_2 \ldots x_n$ then $s = 2^{\Omega(n)}$.*

They get the following corollary.

Corollary 6.1.3 ([ERS16]). *Let $g = \sum_{i=1}^{s} f^{\alpha_i}$ where $f_i = P_n^{d_i}(\ell_{i1}, \ldots, \ell_{iN}) + \beta_i$, $\sqrt{n} \leq d_i \leq n$, $N < 2^{0.001\sqrt{n}}$ and $\ell_{i1}, \ldots, \ell_{in}$ are homogeneous linear forms for every i. If $g = x_1 x_2 \ldots x_n$ then $s = 2^{\Omega(n)}$.*

When $d = 1$, it reduces to the case of depth three powering circuits. Otherwise, we improve upon the result of [ERS16] by using a simpler complexity measure and do a tighter analysis. We further extend the model to accomodate the case where the degrees to which the linear forms are raised to, are not necessarily uniform. Formally, we consider the polynomials of the form $g = \sum_{i=1}^{s} f_i^{\alpha_i}$ where $f_i = (\ell_1^{d_{i1}} + \cdots + \ell_m^{d_{im}} + c_i)$, ℓ_is are homogeneous linear polynomials over $\mathbb{F}[X]$ and $c_i \in \mathbb{F}$. The analysis of Engels et al. holds only when $\max\left\{d_{ij} : (i,j) \in [s] \times [m]\right\} \leq \frac{1.02}{d} 2^{0.489d}$ where $d = \min\left\{d_{ij} : (i,j) \in [s] \times [m]\right\}$. We get rid of that constraint on the $\max\left\{d_{ij} : (i,j) \in [s] \times [m]\right\}$ and we just need the minimum of those degrees to be at least 8. We summarize this result in the following theorem.

Theorem 6.1.4. *Let $g = \sum_{i=1}^{s} f_i^{\alpha_i}$ where $f_i = (\ell_1^{d_{i1}} + \cdots + \ell_m^{d_{im}} + c_i)$, ℓ_is are homogeneous linear polynomials over $\mathbb{F}[X]$ and $c_i \in \mathbb{F}^*$. Let d be the minimum of $\left\{d_{ij} : (i,j) \in [s] \times [m]\right\}$. If $g \equiv x_1 x_2 \ldots x_n$ then for $m = n$ and $d \geq 8$, $s = 2^{\Omega(n)}$.*

Further, we also observe that such a bound also holds for larger values of m. That helps us arrive at the following corollary.

Corollary 6.1.5. *For any integer n, the monomial $x_1x_2\dots x_n$ can be computed by a $\Sigma\wedge\Sigma^{[\text{hom},\ 2^{\sqrt{n}}]}\wedge^{[=\sqrt{n}]}\Sigma^{[\text{hom}]}$ circuit of size $2^{O(\sqrt{n})}$ but any $\Sigma\wedge\Sigma^{[2^{0.955\sqrt{n}}]}\wedge^{[\geq\sqrt{n}]}\Sigma^{[\text{hom}]}$ computing it must be of size at least $2^{\Omega(n)}$.*

We also consider the depth five powering circuits of the form $g=\sum_{i=1}^{s}f_i^{\alpha_i}$ where $f_i=(\ell_1^d+\dots+\ell_m^d+c_i)$ where ℓ_is are homogeneous linear polynomials, c_is are non-zero field elements, and ℓ_is form a low rank subspace. Formally, we prove the following theorem.

Theorem 6.1.6. *Let $g=\sum_{i=1}^{s}f_i^{\alpha_i}$ be such that $f_i=(\ell_1^d+\dots+\ell_m^d+c_i)$, ℓ_is are homogeneous linear polynomials over $\mathbb{F}[X]$ and $c_i\in\mathbb{F}^*$. Let $r<\varepsilon d$ be the rank of the linear forms $\{\ell_1,\ell_2,\dots,\ell_m\}$ for a parameter $\varepsilon<1$. If $g\equiv x_1x_2\dots x_n$ then there exists a suitable value for the parameter ε such that $s=2^{\Omega(n)}$.*

This gives us the following insight.

Corollary 6.1.7. *For any integer n, the monomial $x_1x_2\dots x_n$ can be computed by a $\Sigma\wedge\Sigma\wedge^{[\text{hom},\ =\sqrt{n}]}\Sigma^{[\text{hom},\{=\sqrt{n}\}]}$ circuit of size $2^{O(\sqrt{n})}$ but any $\Sigma\wedge\Sigma\wedge^{[=\sqrt{n}]}\Sigma^{[\text{hom},\ \{\leq\varepsilon\sqrt{n}\}]}$ computing it must be of size at least $2^{\Omega(n)}$.*

6.2. Preliminaries

First, we shall present the definition of the powering circuits again, for the sake of completeness.

Definition 6.2.1 (Powering circuits)**.** *A powering circuit is an arithmetic circuit where the internal nodes are either addition $(+)$ or the powering gates $(\wedge)$.* ◊

In this work, we shall work with two specific restrictions of powering circuits of depth five.

Definition 6.2.2 (Depth five powering circuits)**.** *A depth five powering circuit computes the sum of powers of sums of powers of linear forms. Formally, a $\Sigma\wedge\Sigma\wedge\Sigma$ circuit computes*

the polynomials of the form

$$\sum_{i=1}^{s}\left(\sum_{j=1}^{m}\ell_{ij}^{d_{ij}}+c_i\right)^{\alpha_i}$$

where ℓ_is are linear forms over $\mathbb{F}[x_1,x_2,\ldots,x_n]$. ◊

Here are the following two restrictions of the depth five powering circuits that we consider.

- A $\Sigma\wedge\Sigma^{[m]}\wedge^{[\geq d]}\Sigma^{[\mathrm{hom}]}$ circuit computes the polynomials of the form

  $$\sum_{i=1}^{s}\left(\sum_{j=1}^{m}\ell_{ij}^{d_{ij}}+c_i\right)^{\alpha_i}$$

  such that, for any $i\in[s]$, the linear forms $\left\{\ell_{ij}:(i,j)\in[s]\times[m]\right\}$ are homogeneous linear polynomials and $d=\min\left\{d_{ij}:(i,j)\in[s]\times[m]\right\}$.

- A $\Sigma\wedge\Sigma\wedge^{[=d]}\Sigma^{[\mathrm{hom},\{r\}]}$ circuit computes the polynomials of the form

  $$\sum_{i=1}^{s}\left(\sum_{j=1}^{m}\ell_{ij}^{d}+c_i\right)^{\alpha_i}$$

  such that, m is unbounded and for any $i\in[s]$, the linear forms $\left\{\ell_{ij}:(i,j)\in[s]\times[m]\right\}$ are homogeneous linear polynomials and for every i, the rank of the linear system $\left\{\ell_{ij}:j\in[m]\right\}$ is at most r.

Complexity measure: Multilinear derivatives

Let us define our complexity measure as follows.

Definition 6.2.3. *For an integer $k>0$, we define the dimension of the multilinear derivatives of order k (denoted by Γ_k) as follows.*

$$\Gamma_k(f)=\dim(\mathbb{F}\text{-}span\{\pi_m(\partial_{\mathrm{ML}}^{=k}f)\})$$

where $\partial^{=k}_{\mathrm{ML}} f$ *is the space of partial derivatives of order* k *with respect to the multilinear monomials and* π_m *projects the polynomial to its multilinear component.* ◊

Since the polynomial we consider is a multilinear polynomial, it is sufficient to consider the derivatives with respect to multilinear monomials and multilinear projections. This measure is a restriction of the measure introduced by Nisan and Wigderson [NW97] and is a stripped down version of *Projected multilinear derivatives* [4] which was introduced by Engels et al. [ERS16]. It is easy to see that the measure is sub-additive. We crucially use that property in our proof.

Proposition 6.2.4. *The measure, the dimension of the multilinear derivatives is sub-additive. Formally,* $\Gamma_k(f_1+f_2) \leq \Gamma_k(f_1)+\Gamma_k(f_2)$.

We shall now mention the result of Fischer [Fis94] where he showed that $\mathrm{wrk}(x_1x_2\dots x_n)$ is at most 2^{n-1} by giving an explicit set of linear forms.

Proposition 6.2.5 ([Fis94])**.** *For any* n*, the monomial* $x_1\dots x_n$ *can be expressed as a linear combination of* 2^{n-1} *powers of linear forms as following.*

$$2^{n-1}\cdot n!\cdot x_1\dots x_n = \sum_{(r_2,r_3,\dots,r_n)\in\{\pm1\}^{n-1}} (-1)^{\mathrm{wt}(\mathbf{r})}\cdot\left(x_1+\sum_{j=1}^{n} r_i x_i\right)^n$$

where $\mathrm{wt}(\mathbf{r}) = |\{i : r_i = -1\}|$.

The following property is true about the binary entropy function.

Claim 6.2.6. *For positive real numbers* $\{a_1,a_2,\dots,a_p,u_1,u_2,\dots,u_p\}$*, each of which is in* $[0,1]$*, such that* $u_1+u_2+\cdots+u_p=1$*,*

$$\sum_{i=1}^{p} u_i H(a_i) \leq H\left(\sum_{i=1}^{p} u_i a_i\right)$$

where $H(q)$ *is the binary entropy function.*

[4] The dimension of the projected multilinear derivatives in [ERS16] is defined as $\dim(\mathbb{F}\text{-span}\{\sigma_S(\pi_m(\partial_{\mathrm{ML}} f))\})$ where σ_S sets the variables in S to zero. They need the extra projection to take care of the depth four powering circuits that they consider. We observe that it is not necessary in the case of depth five powering circuits.

The proof of this claim follows easily from the fact that the binary entropy function is a concave function over $(0,1)$.

6.3. Depth five powering circuit for the monomial

In this section we shall construct a depth five powering circuit of size $2^{\Omega(\sqrt{n})}$ that computes $x_1 x_2 \dots x_n$.

Proposition 6.3.1. *There is a* $\Sigma\wedge\Sigma^{\text{hom},\left[2^{\sqrt{n}-1}\right]}\wedge\left[\sqrt{n}\right]\Sigma\left[\sqrt{n}\right]$ *formula of size* $2^{\sqrt{n}}$ *computing the monomial* $x_1 x_2 \dots x_n$.

Proof. Let the monomial $x_1 x_2 \dots x_n$ be expressed as $m_1 m_2 \dots m_{\sqrt{n}}$ where

$$m_i = x_{((i-1)\sqrt{n}+1)} x_{((i-1)\sqrt{n}+2)} \dots x_{(i\sqrt{n})} \;\forall\; i \in [\sqrt{n}].$$

Invoking Proposition 6.2.5 with $m_1, m_2, \dots, m_{\sqrt{n}}$ as the variables, we get a depth three powering circuit of size $\sqrt{n}\cdot 2^{\sqrt{n}-1}$ over $\mathbb{F}\left[m_1, m_2, \dots, m_{\sqrt{n}}\right]$. Furthermore, using Proposition 6.2.5, each of these m_is can be expressed as depth three powering circuits of size $\sqrt{n}\cdot 2^{\sqrt{n}-1}$ over $\mathbb{F}\left[x_{(i-1)\sqrt{n}+1}, x_{(i-1)\sqrt{n}+2}, \dots, x_{i\sqrt{n}}\right]$. The $\Sigma\wedge\Sigma^{\text{hom},\left[2^{\sqrt{n}-1}\right]}\wedge\left[\sqrt{n}\right]\Sigma\left[\sqrt{n}\right]$ circuit thus obtained is of size at most $n\cdot 2^{2\sqrt{n}-2}$. □

Saptharishi [Sap15] gives an elegant construction of a depth four powering circuit of size $2^{O(\sqrt{n})}$ using Ellison's lemma/Newton identities in the first step and the Fisher's identity in the second step.

6.4. Hardness of the monomial under this measure

Lemma 6.4.1. *For any integer* $k < n$,

$$\Gamma_k(x_1 x_2 \dots x_n) = \binom{n}{k}.$$

Proof. The derivative space of the monomial $x_1x_2\dots x_n$, of order k is spanned by the multilinear monomials of degree exactly $n-k$ over $\mathbb{F}[X]$. Thus,

$$\Gamma_k(x_1x_2\dots x_n)=\binom{n}{n-k}=\binom{n}{n-(n-k)}=\binom{n}{k}.$$

□

6.5. Weakness of the $\Sigma\wedge\Sigma^{[m]}\wedge^{[\geq d]}\Sigma^{[\hom]}$ circuits under this measure

In this section, we show that $\Sigma\wedge\Sigma^{[m]}\wedge^{[\geq d]}\Sigma^{[\hom]}$ circuits of polynomial size cannot compute the monomial $x_1x_2\cdots x_n$. We start by showing that the dimension of the multilinear derivative space for any polynomial computed by this model is low.

Lemma 6.5.1. *Let k and $t>\frac{n}{2}$ be some parameters. Let $f=(\ell_1^{d_i}+\cdots+\ell_m^{d_m}+c)$ where ℓ_is are homogeneous linear polynomials over $\mathbb{F}[X]$ and c is a non-zero field element. Then for any positive integer α,*

$$\Gamma_k(f^\alpha)\leq k\left[\binom{m+p}{p}\binom{k}{p}+\frac{n}{2}\binom{n}{t}\right]$$

where $d=\min\{d_1,d_2,\dots,d_m\}$ and $p<\frac{t+k}{d}$.

Proof. Let us note that the space of partial derivatives of order k, of f lies in $\mathbb{F}\text{-span}\{\ell_i^{d_i-k}:i\in[m]\}$. Extending this, we obtain the following.

$$\begin{aligned}\partial^k f^\alpha &\subseteq \mathbb{F}\text{-span}\{f^{\alpha-p}\cdot\partial^{k_1}f\cdot\ldots\partial^{k_p}f : p\in[k]\ \&\ k_1+\cdots+k_p=k\}\\ &\subseteq \mathbb{F}\text{-span}\{f^{\alpha-p}:p\in[k]\}\otimes\mathbb{F}\text{-span}\{\partial^{k_1}f\cdot\ldots\partial^{k_p}f : p\in[k]\ \&\ k_1+\cdots+k_p=k\}.\end{aligned}$$

The dimension of $\mathbb{F}\text{-span}\{f^{\alpha-p}:p\in[k]\}$ is trivially upper bounded by k. Let W be the

vector space $\mathbb{F}$-span$\{\partial^{k_1}f\cdot\cdots\cdot\partial^{k_p}f : k_1+\cdots+k_p=k\}$. Then, $\Gamma_k(f^\alpha)\leq k\cdot\Gamma_k(W)$.

$$\begin{aligned}W &= \mathbb{F}\text{-span}\{\partial^{k_1}f\cdot\cdots\cdot\partial^{k_p}f : k_1+\cdots+k_p=k\}\\ &\subseteq \mathbb{F}\text{-span}\{\ell_{i_1}^{d_{i_1}-k_1}\cdot\cdots\cdot\ell_{i_p}^{d_{i_p}-k_p} : \bar{i}=(i_1,\ldots,i_p)\in[m]^p \;\&\; k_1+\cdots+k_p=k\}.\end{aligned}$$

If the degree of any term $T_{\bar{i}}=\ell_{i_1}^{d_{i_1}-k_1}\cdot\cdots\cdot\ell_{i_p}^{d_{i_p}-k_p}$ is greater than n then its contribution to $\Gamma_k(W)$ is zero. Let us now consider all the other terms in W whose degree is at most n. Let t be a degree threshold such that $t>\frac{n}{2}$.

- Let us consider the terms $T_{\bar{i}}$ such that their degree is in $[t,n]$. There are at most $\sum_{j\in[t,n]}\binom{n}{j}\leq\binom{n}{t}\cdot(n-t+1)$ many multilinear monomials over $\mathbb{F}[X]$ of degree at least t and at most n. Their contribution to $\Gamma_k(W)$ is at most $\frac{n}{2}\binom{n}{t}$.
- Otherwise, the degree of any other term $T_{\bar{i}}$ is at most $t-1$.

$$\begin{aligned}d_{i_1}-k_1+d_{i_2}-k_2+\cdots+d_{i_p}-k_p &< t\\ p\cdot\min\left\{d_{i_1},d_{i_2},\ldots,d_{i_p}\right\}-(k_1+k_2+\cdots+k_p) &< t\\ \implies p &< \frac{t+k}{d}\end{aligned}$$

since $d\leq\min\left\{d_{i_1},d_{i_2},\ldots,d_{i_p}\right\}$. The number of terms $\{T_{\bar{i}}\}_{\bar{i}\in[m]^p}$ of degree at most $t-1$ can be counted as follows. We can choose the indices $(i_1,i_2,\ldots,i_p)$ in $\binom{m+p}{m}$ ways, and choose $k_1,k_2,\ldots,k_p$ in at most $\binom{k}{p}$ ways such that $k_1+k_2+\cdots+k_p=k$.

Thus,

$$\Gamma_k(f^\alpha)\leq k\cdot\Gamma_k(W)\leq k\left[\binom{k}{p}\binom{m+p}{p}+\frac{n}{2}\binom{n}{t}\right].$$

□

Putting it all together

Theorem 6.5.2. *Let $g=\sum_{i=1}^{s}f_i^{\alpha_i}$ be such that $f_i=(\ell_1^{d_{i1}}+\cdots+\ell_m^{d_{im}}+c_i)$ such that ℓ_is are homogeneous linear polynomials over $\mathbb{F}[X]$ and $c_i\in\mathbb{F}^*$. Let d be the minimum of $\left\{d_{ij}:(i,j)\in[s]\times[m]\right\}$. If $g\equiv x_1x_2\ldots x_n$ then for $m=n$ and $d\geq 8$, $s=2^{\Omega(n)}$.*

Proof. Since $g = \sum_{i=1}^{s} f_i^{\alpha_i}$ and from Proposition 6.2.4 we can infer the following.

$$\Gamma_k(g) \leq s \cdot \max_{i \in [s]}(\Gamma_k(f_i^{\alpha_i})).$$

From Lemma 6.4.1 and the fact that $g \equiv x_1 x_2 \ldots x_n$, we can infer that $\Gamma_k(g) = \binom{n}{k}$. By invoking Lemma 6.5.1 we can get an upper bound on $\max_{i \in [s]}(\Gamma_k(f_i^{\alpha_i}))$. Thus,

$$s \cdot k \cdot \left[\binom{k}{p}\binom{m+p}{p} + \binom{n}{t} \right] \geq \binom{n}{k}$$

$$\implies s \geq \frac{\binom{n}{k}}{k\left[\binom{k}{p}\binom{m+p}{p} + \frac{n}{2}\binom{n}{t}\right]}.$$

Let us fix the value of k to $0.5n$ so as to maximize the numerator.

- $\binom{n}{k} = 2^n$.
- $\binom{k}{p}\binom{m+p}{p} = 2^{k \cdot H\left(\frac{p}{k}\right) + (m+p) \cdot H\left(\frac{p}{m+p}\right)} = 2^{n \cdot \left(0.5H\left(\frac{p}{k}\right) + \frac{m+p}{n} H\left(\frac{p}{m+p}\right)\right)}$.
- $\binom{n}{t} = 2^{n \cdot H\left(\frac{t}{n}\right)}$.

$$\log(sk) \geq n \cdot \left(1 - \max\left\{0.5H\left(\frac{p}{k}\right) + \frac{m+p}{n} H\left(\frac{p}{m+p}\right), H\left(\frac{t}{n}\right)\right\}\right).$$

We will set the value of t to a value which is away from $0.5n$ on the greater side so that $\binom{n}{t}$ is the non dominant term. Let us fix the parameters as follows: $k = 0.5n$, $t = 0.545n$ and $m = n$. This setting of parameters forces the criterion that d must at least be 8 and thus $p < 0.13n$. By substituting the values, we get $\binom{k}{p}\binom{m+p}{p} = 2^{0.99815n}$ and $\binom{n}{t} = 2^{0.99415n}$. Thus, $s \geq 2^{0.00184n}$. □

Theorem 6.5.3. *Any $\Sigma\wedge\Sigma^{[m]}\wedge^{[\geq d]}\Sigma^{[\hom]}$ circuit computing the monomial $x_1 x_2 \ldots x_n$ must be of size $2^{\Omega(n)}$ where $m \leq \frac{0.29n}{d^2} \cdot 2^{0.955d}$.*

Proof. From the proof of Theorem 6.5.2, we get that

$$s \geq \frac{\binom{n}{k}}{k\left[\binom{k}{p}\binom{m+p}{p}+\binom{n}{t}\right]}$$

and the value of p is at most $\frac{t+k}{d}$. Let us fix the value of k to $0.5n$ again so as to maximize the numerator and we will set the value of t to a value of $0.545n$ which is away from $0.5n$ on the greater side so that $\binom{n}{t}$ is the non dominant term.

Then we would want the other term in the denominator to be exponentially smaller than $2^{nH\left(\frac{k}{n}\right)}$ (say by $2^{\gamma n}$ where $\gamma = 0.998$) at its maximum.

$$\max\left(\binom{k}{p}\binom{m+p}{p}\right) \leq 2^{\gamma n} \implies \frac{e^2k(m+p)}{p^2} \leq 2^{\frac{\gamma n}{p}}$$

$$(m+p) \leq \left(\frac{p^2}{e^2k}\right)\cdot 2^{\frac{\gamma n}{p}} \implies m \leq \left(\frac{p^2}{e^2k}\right)\cdot 2^{\frac{\gamma n}{p}} - p < \left(\frac{(t+k)^2}{e^2d^2k}\right)\cdot 2^{\frac{\gamma n}{p}}$$

$$m < \frac{(1.045)^2 n}{0.5e^2d^2}\cdot 2^{\frac{\gamma d}{1.045}} = \frac{0.29n}{d^2}\cdot 2^{0.955d}.$$

This completes the proof. □

In particular, we infer the following by setting d to $\sqrt{n}$.

Corollary 6.5.4. *For any integer n, the monomial $x_1x_2\ldots x_n$ can be computed by a $\Sigma\wedge\Sigma^{[\text{hom},\ 2^{\sqrt{n}}]}\wedge^{[=\sqrt{n}]}\Sigma^{[\text{hom}]}$ circuit of size $2^{O(\sqrt{n})}$ but any $\Sigma\wedge\Sigma^{[2^{0.955\sqrt{n}}]}\wedge^{[\geq\sqrt{n}]}\Sigma^{[\text{hom}]}$ computing it must be of size at least $2^{\Omega(n)}$.*

6.6. Weakness of the $\Sigma\wedge\Sigma\wedge^{[=d]}\Sigma^{[\text{hom},\ \{r\}]}$ circuits under this measure

In this section, we show that $\Sigma\wedge\Sigma\wedge^{[=d]}\Sigma^{[\text{hom},\ \{r\}]}$ circuits of polynomial size cannot compute the monomial $x_1x_2\cdots x_n$. We start by showing that the dimension of the multilinear

derivative space for any polynomial computed by this model is low.

Lemma 6.6.1. *Let k and $t > \frac{n}{2}$ be some parameters. Let $f = (\ell_1^d + \cdots + \ell_m^d + c)$ where ℓ_is are homogeneous linear polynomials over $\mathbb{F}[X]$ such that the rank of $\{\ell_1, \ell_2, \ldots, \ell_m\}$ is at most r and c is a non-zero field element. Then for any positive integer α,*

$$\Gamma_k(f^\alpha) \leq k\left[\binom{k}{p} \cdot 2^{(p(d+r)-k)\cdot H\left(\frac{rp}{p(d+r)-k}\right)} + \frac{n}{2}\binom{n}{t}\right]$$

where $p < \frac{t+k}{d}$.

Proof. Similar to the proof of Lemma 6.5.1, we have the following.

$$\begin{aligned}\partial^k f^\alpha &\subseteq \mathbb{F}\text{-span}\{f^{\alpha-p} \cdot \partial^{k_1} f \cdot \ldots \partial^{k_p} f : p \in [k] \;\&\; k_1 + \cdots + k_p = k\} \\ &\subseteq \mathbb{F}\text{-span}\{f^{\alpha-p} : p \in [k]\} \otimes \mathbb{F}\text{-span}\{\partial^{k_1} f \cdot \ldots \partial^{k_p} f : p \in [k] \;\&\; k_1 + \cdots + k_p = k\}.\end{aligned}$$

The dimension of $\mathbb{F}$-span$\{f^{\alpha-p} : p \in [k]\}$ is trivially upper bounded by k. Let W be the vector space $\mathbb{F}$-span$\{\partial^{k_1} f \cdot \cdots \cdot \partial^{k_p} f : k_1 + \cdots + k_p = k\}$. Thus, $\Gamma_k(f^\alpha) \leq k \cdot \Gamma_k(W)$.

$$\begin{aligned}W &= \mathbb{F}\text{-span}\{\partial^{k_1} f \cdot \cdots \cdot \partial^{k_p} f : k_1 + \cdots + k_p = k\} \\ &\subseteq \mathbb{F}\text{-span}\{\ell_{i_1}^{d-k_1} \cdot \cdots \cdot \ell_{i_p}^{d-k_p} : \bar{i} = (i_1, \ldots, i_p) \in [m]^p \;\&\; k_1 + \cdots + k_p = k\} \\ &\subseteq \mathbb{F}\text{-span}\left\{\ell_{i_1}^{d-k_1} : i_1 \in [m]\right\} \otimes \cdots \otimes \mathbb{F}\text{-span}\left\{\ell_{i_p}^{d-k_p} : i_p \in [m]\right\};\; k_1 + \cdots + k_p = k.\end{aligned}$$

If the degree of any term $T_{\bar{i}} = \ell_{i_1}^{d-k_1} \cdot \cdots \cdot \ell_{i_p}^{d-k_p}$ is greater than n then its contribution to $\Gamma_k(W)$ is zero. Let us now consider all the other terms in W whose degree is at most n. Let $t > \frac{n}{2}$ be a degree threshold that we shall fix later. Let us consider the terms $T_{\bar{i}}$ whose degree lies in $[t, n]$. There are at most $\binom{n}{t} \cdot (n - t + 1)$ many multilinear monomials over $\mathbb{F}[X]$ of degree at least t and at most n. Otherwise, the degree of $T_{\bar{i}}$ is at most $t - 1$.

$$\begin{aligned}d - k_1 + d - k_2 + \cdots + d - k_p &< t \\ p \cdot d - (k_1 + k_2 + \cdots + k_p) &< t \\ \implies p &< \frac{t+k}{d}\end{aligned}$$

Without loss of generality, let us suppose that the set $\{\ell_1,\ell_2,\ldots,\ell_r\}$ forms the linear basis for $\{\ell_1,\ell_2,\ldots,\ell_m\}$. Thus, ℓ_{i_1} can be written as a linear combination of the basis elements. Also,

$$\begin{aligned}\partial^{k_i} f &\subseteq \mathbb{F}\text{-span}\left\{\ell_j^{d-k_i} : j\in[m]\right\}\\ &\subseteq \mathbb{F}\text{-span}\left\{\ell_1^{a_1}\ell_2^{a_2}\ldots\ell_r^{a_r} : a_1+a_2+\cdots+a_r = d-k_i\right\}.\end{aligned}$$

The number of integral solutions to the equation $a_1+a_2+\cdots+a_r = d-k_i$ is at most $\binom{d-k_i+r}{r}$. Thus for a fixed $(k_1,k_2,\ldots,k_p)$ such that $k_1+k_2+\cdots+k_p = k$,

$$\dim\left(\otimes_{j\in[p]}\mathbb{F}\text{-span}\left\{\ell_{i_j}^{d-k_j} : i_j\in[m]\right\}\right)\leq\prod_{i=1}^{p}\binom{d-k_i+r}{r}.$$

The number of ways of choosing $(k_1,k_2,\ldots,k_p)$ is at most $\binom{k}{p}$.

$$\begin{aligned}\implies \dim(W) &\leq \binom{k}{p}\cdot\prod_{i=1}^{p}\binom{d-k_i+r}{r}\\ &\leq \binom{k}{p}\cdot 2^{\sum_{i=1}^{p}(d-k_i+r)\cdot H\left(\frac{r}{d-k_i+r}\right)}\\ &= \binom{k}{p}\cdot 2^{(p(d+r)-k)\sum_{i=1}^{p}\frac{(d-k_i+r)}{p(d+r)-k}\cdot H\left(\frac{r}{d-k_i+r}\right)}\end{aligned}$$

and by using Claim 6.2.6 we get that

$$\begin{aligned}2^{(dm+rm-k)\sum_{i=1}^{p}\frac{(d-k_i+r)}{p(d+r)-k}\cdot H\left(\frac{r}{d-k_i+r}\right)} &\leq 2^{(p(d+r)-k)\cdot H\left(\sum_{i=1}^{p}\frac{(d-k_i+r)}{p(d+r)-k}\cdot\frac{r}{d-k_i+r}\right)}\\ &= 2^{(p(d+r)-k)\cdot H\left(\frac{pr}{p(d+r)-k}\right)}.\end{aligned}$$

Thus,

$$\Gamma_k(f^\alpha)\leq k\cdot\Gamma_k(W)\leq k\left[\binom{k}{p}2^{(p(d+r)-k)\cdot H\left(\frac{pr}{p(d+r)-k}\right)}+\frac{n}{2}\binom{n}{t}\right].$$

□

Putting it all together

Theorem 6.6.2. *Let $g = \sum_{i=1}^{s} f_i^{\alpha_i}$ be such that $f_i = (\ell_1^d + \cdots + \ell_m^d + c_i)$, ℓ_is are homogeneous linear polynomials over $\mathbb{F}[X]$ and $c_i \in \mathbb{F}$. Let $r < \varepsilon d$ be the rank of the linear forms $\{\ell_i : i \in [m]\}$ for a parameter $\varepsilon < 1$. If $g \equiv x_1 x_2 \ldots x_n$ then there exists a setting for the parameter ε such that $s = 2^{\Omega(n)}$.*

Proof. From Proposition 6.2.4,

$$\Gamma_k(g) \leq s \cdot \max_{i \in [s]}(\Gamma_k(f_i^{\alpha_i})).$$

From Lemma 6.4.1 and the fact that $g \equiv x_1 x_2 \ldots x_n$, we can infer that $\Gamma_k(g) = \binom{n}{k}$. By invoking Lemma 6.6.1 we can get an upper bound on $\max_{i \in [s]}(\Gamma_k(f_i^{\alpha_i}))$. Thus,

$$s \cdot k \cdot \left[\binom{k}{p} \cdot 2^{(p(d+r)-k) \cdot H\left(\frac{rp}{p(d+r)-k}\right)} + \frac{n}{2}\binom{n}{t} \right] \geq \binom{n}{k}$$

$$\implies s \geq \frac{\binom{n}{k}}{k \cdot \left[\binom{k}{p} \cdot 2^{(p(d+r)-k) \cdot H\left(\frac{rp}{p(d+r)-k}\right)} + \frac{n}{2}\binom{n}{t} \right]}$$

As before, let us fix the value of k to $0.5n$ so as to maximize the numerator and we will set the value of t to $0.545n$, a value which is away from $0.5n$ on the greater side so that $\binom{n}{t}$ is the non dominant term.

$$\begin{aligned}
&\log\left[\binom{k}{p} \cdot 2^{(p(d+r)-k) \cdot H\left(\frac{rp}{p(d+r)-k}\right)} \right] \\
&= p \log\left(\frac{ek}{p}\right) + (p(d+r)-k) \cdot H\left(\frac{rp}{p(d+r)-k}\right) \\
&\leq \frac{(t+k)}{d} \log\left(\frac{ek}{p}\right) + \frac{(t+k)(d+r)-kd}{d} \cdot H\left(\frac{\varepsilon(t+k)}{(d+r)(t+k)-kd}\right) \\
&= \frac{1.045n}{d} \log\left(\frac{ed}{1.045}\right) + (0.545 + 1.045\varepsilon)n \cdot H\left(\frac{1.045\varepsilon}{(0.545+1.045\varepsilon)}\right).
\end{aligned}$$

For a constant value of $d \geq 8$, the value of ε can be set to 0.1. Also, the value of ε can be raised to 0.4 when $d \geq 200$. For a reasonably large value of $d = \omega(1)$, the second summand dominates the first. For a value of $\varepsilon = 0.44001$, the second summand would compute to a value of $0.9996n$. This gives us a size lower bound of at least $2^{0.0004n}$. □

Again, by setting the value of d to $\sqrt{n}$, we get the following corollary.

Corollary 6.6.3. *For any integer n, the monomial $x_1x_2\ldots x_n$ can be computed by a $\Sigma\wedge\Sigma^{[\text{hom}]}\wedge^{[=\sqrt{n}]}\Sigma^{[\text{hom},\ \{=\sqrt{n}\}]}$ circuit of size $2^{O(\sqrt{n})}$ but any $\Sigma\wedge\Sigma\wedge^{[=\sqrt{n}]}\Sigma^{[\text{hom},\ \{\leq\varepsilon\sqrt{n}\}]}$ computing it must be of size at least $2^{\Omega(n)}$.*

Part III.

Determinantal Complexity

7

Determinantal complexity of Iterated matrix multiplication polynomial

One of the goals in this study is to compare our current knowledge of the depth-4 circuit size lower bounds and the determinantal complexity lower bounds. Currently the best known determinantal complexity lower bound is $\Omega(n^2)$ for Permanent of a $n \times n$ matrix (which is a n^2-variate and degree n polynomial) [MR04, CCL08, Yab15]. We prove that the determinantal complexity of the iterated matrix multiplication polynomial is $\Omega(dn)$ where d is the number of matrices and n is the dimension of the matrices. So for $d = n$, we get that the iterated matrix multiplication polynomial achieves the current best known lower bounds in both fronts: depth-4 circuit size and determinantal complexity. Our result also settles the determinantal complexity of the iterated matrix multiplication polynomial to $\Theta(dn)$.

To the best of our knowledge, a $\Theta(n)$ bound for the determinantal complexity for the iterated matrix multiplication polynomial was known only for any constant $d > 1$ [Jan11].

7.1. Introduction

Let us recall that a multivariate polynomial family $\{f_n(X) \in \mathbb{F}[x_1, x_2, \ldots, x_n] : n \geq 1\}$ is in the class VP if f_n has degree of at most poly(n) and can be computed by an arithmetic

circuit of size poly(n). It is in VNP if it can be expressed as

$$f_n(X) = \sum_{Y \in \{0,1\}^m} g_{n+m}(X,Y)$$

where $m = |Y| = \text{poly}(n)$ and g_{n+m} is a polynomial family in VP. Permanent polynomial characterizes the class VNP over the fields of all characteristics except 2 and the determinant polynomial characterizes the class VP with respect to the quasi-polynomial projections.

Definition 7.1.1. *The determinantal complexity of a polynomial f, over n variables, is the minimum m such that there are affine linear polynomials $A_{k,\ell}$, $1 \leq k,\ell \leq m$ defined over the same set of variables and $f = \det((A_{k,\ell})_{1\leq k,\ell\leq m})$. It is denoted by* $\text{DetComp}(f)$. ◊

To resolve Valiant's hypothesis, proving $\text{DetComp}(\text{Perm}_n) = n^{\omega(\log n)}$ is sufficient. Von zur Gathen [vzG86] proved that $\text{DetComp}(\text{Perm}_n) \geq \sqrt{\frac{8}{7}}n$. Later Cai [Cai90], Babai and Seress [vzG87], and Meshulam [Mes89] independently improved the lower bound to $\sqrt{2}n$. In 2004, Mignon and Ressayre [MR04] proved that $\text{DetComp}(\text{Perm}_n) \geq \frac{n^2}{2}$ over the fields of characteristic zero, using algebraic geometry. Subsequently, Cai et al. [CCL08] extended the result of Mignon and Ressayre [MR04] to all fields of characteristic $\neq 2$. They also provided a simpler analysis.

For any polynomial f, Valiant [Val79] proved that $\text{DetComp}(f) \leq 2(F(f)+1)$ where $F(f)$ is the arithmetic formula complexity of f. Later, Nisan [Nis91] proved that $\text{DetComp}(f) = O(B(f))$ where $B(f)$ is the arithmetic branching program complexity of f.

The main result of this chapter is a lower bound on the determinantal complexity of the iterated matrix multiplication polynomial.

Theorem 7.1.2. *For any integers n and $d > 1$, the determinantal complexity of the iterated matrix multiplication polynomial* $\text{IMM}_{n,d}$ *is* $0.5dn$.

Since $\text{IMM}_{n,d}(X)$ has an algebraic branching program of size $O(dn)$ [Nis91], from the above theorem it follows that $\text{DetComp}(\text{IMM}_{n,d}(X)) = \Theta(dn)$. This improves upon the

earlier bound of $\Theta(n)$ for the determinantal complexity of the iterated matrix multiplication polynomial for any constant $d > 1$ [Jan11]. Similar to the approach of [CCL08] and [MR04], we also use the the rank of Hessian matrix as our main technical tool.

As mentioned before, the current best known determinantal complexity (DetComp) lower bound for an explicit polynomial in VNP is only quadratic, for the permanent polynomial [MR04]. Before the result of Mignon and Ressayre, the best known determinantal complexity lower bound for the $n \times n$ permanent polynomial was $\sqrt{2}n$ [Cai90, vzG87]. These results were proved using nontrivial algebraic-geometric concepts. Of course, one can prove that VP $\neq$ VNP by proving a super-quasi-polynomial determinantal complexity lower bound for any other explicit polynomial in VNP. One such polynomial that we consider is the Nisan Wigderson polynomial.

Here we first show that $\mathrm{DetComp}(\mathsf{NW}_{n,\varepsilon n}(X)) \geq \Omega(n^{1.5})$ using elementary ideas. This is in contrast to the results of Cai and von zur Gathen [Cai90, vzG87]. We will also prove a lower bound on the determinantal complexity of $\mathsf{IMM}_{n,d}$ using the partial derivatives in Section 7.2.

7.2. Lower bounds via the partial derivatives

Let us recall the following definition for the sake of completeness.

Definition 7.2.1. *The dimension of the space of partial derivatives of a polynomial f with respect to a parameter k is defined as* $\Gamma_k(f) := \dim\left(\partial^{=k} f\right)$. ◊

If a polynomial $f = \mathsf{Det}_m(A(X))$ then we need $\Gamma_k(\mathsf{Det}(A(X)))$ must be at least $\Gamma_k(f)$. Let us first obtain a lower bound on the derivative space of $\mathsf{Det}_m(A(X))$.

Derivative space of Det_m polynomial

We will now lower bound the derivative space of $\mathsf{Det}_m(A(X))$ polynomial where $A(X)$ is a $m \times m$ matrix whose entries are linear polynomials over $\mathbb{F}[X]$. Now consider the polynomial $\mathsf{Det}_m(Y)$ over $\mathbb{F}[Y]$ where $Y = \{y_{11}, \ldots, y_{mm}\}$. By the chain rule of deriva-

tives,

$$\frac{\partial \operatorname{Det}_m(A(X))}{\partial x_{i,j}} = \sum_{s,t\in[m]} \frac{\partial \operatorname{Det}_m(Y)}{\partial y_{s,t}}\Big|_{Y\leftarrow A(X)} \cdot \frac{\partial (A(X))_{s,t}}{\partial x_{i,j}}.$$

Since the entries of $A(X)$ are linear polynomials, $\frac{\partial (A(X))_{s,t}}{\partial x_{i,j}}$ is a constant. Generalizing this, we get that

$$\frac{\partial^{=k} \operatorname{Det}_m(A(X))}{\partial x_{i_1,j_1}\dots\partial x_{i_k,j_k}} = \sum_{s_p,t_p\in[m]:p\in[k]} \frac{\partial^{=k} \operatorname{Det}_m(Y)}{\partial y_{s_1,t_1}\dots\partial y_{s_k,t_k}}\Big|_{Y\leftarrow A(X)} \cdot \frac{\partial (A(X))_{s_1,t_1}}{\partial x_{i_1,j_1}} \cdots \frac{\partial (A(X))_{s_k,t_k}}{\partial x_{i_k,j_k}}.$$

This implies that the span of the partial derivative space of $\operatorname{Det}_m(A(X))$, of order k, is a subset of the span of the kth order partial derivative space (with respect to Y) of $\operatorname{Det}_m(Y)$. More formally,

$$\mathbb{F}\text{-span}\left\{\partial^{=k} \operatorname{Det}_m(A(X))\right\} \subseteq \mathbb{F}\text{-span}\left\{(\partial_S \operatorname{Det}_m(Y))|_{y_{ij}=A_{ij};i,j\in[m]} : S\subseteq Y \;\&\; |S|=k\right\}.$$

We note the following simple property of the derivative space of the determinant polynomial. This follows from the fact that a kth order derivative corresponds to a minor of the order $(n-k)$ and any two distinct minors do not share a monomial in common[1].

Proposition 7.2.2. *For any k, $\Gamma_k(\operatorname{Det}_m(Y)) = \binom{m}{k}^2$.*

Invoking the Proposition 7.2.2 and from the discussion above, we get that

$$\Gamma_k\left(\operatorname{Det}_m(A(X))\right) \le \Gamma_k\left(\operatorname{Det}_m(Y)\right) = \binom{m}{k}^2.$$

Derivative space of Nisan-Wigderson polynomial

Let us recall that $\operatorname{NW}_{n,\varepsilon n}(X) = \sum_{a(z)\in\mathbb{F}[z]} x_{1a(1)}x_{2a(2)}\dots x_{na(n)}$ where $\mathbb{F}$ is a finite field of size n and $a(z)$ is a univariate polynomial of degree $< \varepsilon n$ where $\varepsilon \in (0, 0.5)$. Notice that

[1] A much stronger statement about the determinantal ideal can be found in (Theorem 22) [GKKS14] and the references therein.

any two of its monomials can intersect in at most $\varepsilon n - 1$ variables. We now differentiate the polynomial $\mathsf{NW}_{n,\varepsilon n}(X)$ with respect to the first $k = \varepsilon n$ variables of every monomial. After the differentiation, we get $n^{\varepsilon n}$ distinct monomials each of which is of length $(1-\varepsilon)n$. Thus, $\Gamma_k\left(\mathsf{NW}_{n,\varepsilon n}\right) \geq n^{\varepsilon n}$.

Theorem 7.2.3. *For any $\varepsilon \in (0, 0.5)$, it is true that* $\mathrm{DetComp}(\mathsf{NW}_{n,\varepsilon n}) \geq \Omega(n^{1.5})$. *This holds over any field.*

Proof. If the dimension of the partial derivative space of the $\mathsf{Det}_m(A(X))$ is less than the dimension of the partial derivative space of the $\mathsf{NW}_{n,\varepsilon n}(X)$ polynomial, then $\mathsf{NW}_{n,\varepsilon n}(X) = \mathsf{Det}_m(A(X))$ can not hold true. Thus for $k = \varepsilon n$,

$$\begin{aligned} \binom{m}{k}^2 &\leq n^{\varepsilon n} \\ \left(\frac{e \cdot m}{k}\right)^{2k} &\geq n^{\varepsilon n} \\ m &\geq \frac{\varepsilon n \cdot \sqrt{n}}{e} = \Omega(n^{1.5}) \end{aligned}$$

Thus m has to be at least $\Omega(n^{1.5})$ for the $\mathsf{NW}_{n,\varepsilon n}(X)$ polynomial to be written as the affine projection of the Det_m polynomial, that is as $\mathsf{Det}_m(A_{k,\ell})$ where $A_{k,\ell}$, $1 \leq k, \ell \leq m$ are linear polynomials in $\mathbb{F}[X]$. □

Derivative space of the iterated matrix multiplication polynomial

The iterated matrix multiplication polynomial is defined over the disjoint sets of variables $X_1, X_2, \ldots, X_d$.

$$\mathsf{IMM}_{n,d}(X) = \sum_{i_1, i_2, \ldots, i_{n-1} \in [n]} x^{(1)}_{1 i_1} x^{(2)}_{i_1 i_2} \ldots x^{(d-1)}_{i_{(d-2)} i_{(d-1)}} x^{(d)}_{i_{(d-1)} 1}.$$

We will lower bound $\Gamma_k(\mathsf{IMM}_{n,d}(X))$ by the dimension of a specific subspace of the derivative space of $\mathsf{IMM}_{n,d}(X)$. That is dimension of the entire derivative space is lower bounded by the dimension of the subspace that we will now consider. For some distinct elements $J = \{j_1, j_2, \ldots, j_k\}$ such that $|j_s - j_t| > 2$ for any distinct $s, t \in [k]$, consider

the sets of variables $X_{j_1}, X_{j_2}, \ldots, X_{j_k}$. Let us consider the set of monomials M of degree k such that for any monomial $m \in M$, $\left|\mathrm{var}(m) \cap X_{j_t}\right| = 1$ for all $t \in [k]$, and all suitable sets J. It is easy to see that the partial derivatives with respect to the monomials in M are pairwise distinct. The number of ways of picking such a suitable set J is $\binom{d-k+1}{k}$. The number of monomials of degree k in M corresponding to a particular set J is n^{2k}. Thus,

$$\Gamma_k(\mathsf{IMM}_{n,d}(X)) \geq \binom{d-k+1}{k} \cdot n^{2k}.$$

We need that $\Gamma_k(\mathsf{IMM}_{n,d}(X)) \leq \Gamma_k(\mathsf{Det}_m(A(X)))$. Thus for $k = \delta d$ for a suitable $\delta \in (0,1)$,

$$\begin{aligned}
\binom{m}{k}^2 &\geq \binom{d-k+1}{k} \cdot n^{2k} \\
\left(\frac{e \cdot m}{k}\right)^{2k} &\geq \left(\frac{d-k+1}{k}\right)^{k} \cdot n^{2k} \\
\frac{e \cdot m}{k} &\geq \sqrt{\frac{d-k+1}{k}} \cdot n \\
m &\geq e^{-1} n \cdot \sqrt{k(d-k+1)} = \Omega(dn).
\end{aligned}$$

We will improve on this result by a constant factor in Section 7.3.

7.3. Lower bounds via the Hessian

Approach of Mignon and Ressayre

We start by recalling a few facts from [CCL08]. Let f be the target polynomial over N variables. Let $A_{k,\ell}(X)$, $1 \leq k, \ell \leq m$ be the affine linear polynomials over $\mathbb{F}[X]$ such that $f(X) = \det((A_{k,\ell}(X))_{1 \leq k,\ell \leq m})$. Consider a point $X_0 \in \mathbb{F}^N$ such that $f(X_0) = 0$. The affine linear functions $A_{k,\ell}(X)$ can be expressed as $L_{k,\ell}(X - X_0) + y_{k,\ell}$ where $L_{k,\ell}$ is a linear form and $y_{k,\ell}$ is a constant from the field. Thus, $(A_{k,\ell}(X))_{1 \leq k,\ell \leq m} = (L_{k,\ell}(X - X_0))_{1 \leq k,\ell \leq m} + \mathrm{Y}_0$. If $f(X_0) = 0$ then $\det(\mathrm{Y}_0) = 0$. Let C and D be two non-singular matrices such that $\mathrm{C}\mathrm{Y}_0\mathrm{D}$ is a diagonal matrix.

$$\mathrm{CY_0D} = \begin{pmatrix} 0 & 0 \\ 0 & \mathrm{I}_s \end{pmatrix}$$

Since $\det(\mathrm{Y_0}) = 0$, $s < m$. It is enough to assume that $s = m-1$. Since the first row and the first column of $\mathrm{CY_0D}$ are zero, we may multiply $\mathrm{CY_0D}$ by $\mathrm{diag}(\det(\mathrm{C})^{-1}, 1, \ldots, 1)$ and $\mathrm{diag}(\det(\mathrm{D})^{-1}, 1, \ldots, 1)$ on the left and the right side. Without loss of generality, we may assume that $\det(\mathrm{C}) = \det(\mathrm{D}) = 1$. By multiplying with C and D on the left and the right and by suitably renaming $(L_{k,\ell}(X - X_0))_{1 \le k,\ell \le m}$ and $\mathrm{Y_0}$ we get that

$$f(X) = \det((L_{k,\ell}(X - X_0)_{1 \le k,\ell \le m} + \mathrm{Y_0}))$$

where $\mathrm{Y_0} = \mathrm{diag}(0, 1, \ldots, 1)$.

We use $\mathrm{Hess}_f(X)$ to denote the Hessian matrix of the polynomial f and is defined as follows.

$$\mathrm{Hess}_f(X) = (H_{s;ij,t;k\ell}(X))_{1 \le i,j \le n, 1 \le s,t \le d} \text{ such that } H_{s;ij,t;k\ell}(X) = \frac{\partial^2 f(X)}{\partial x_{ij}^{(s)} \partial x_{k\ell}^{(t)}}$$

where $x_{ij}^{(s)}$ and $x_{k\ell}^{(t)}$ denote the (i,j)th and (k,ℓ)th entries of the variable sets X_s and X_t respectively.

By taking second order derivatives and evaluating the Hessian matrices of $f(X)$ and $\det((A_{k,\ell}(X))_{1 \le k,\ell \le m})$ at X_0, we obtain $\mathrm{Hess}_f(X_0) = \mathrm{L}\,\mathrm{Hess}_{\det}(\mathrm{Y_0})\mathrm{L}^T$ where L is a $N \times m^2$ matrix with entries from the field. It follows that $\mathrm{rank}(\mathrm{Hess}_f(X_0)) \le \mathrm{rank}(\mathrm{Hess}_{\det}(\mathrm{Y_0}))$. It was observed in the earlier work of [MR04] and [CCL08] that it is relatively easy to get an upper bound for $\mathrm{rank}(\mathrm{Hess}_{\det}(\mathrm{Y_0}))$. The main task is to construct a point X_0 such that $f(X_0) = 0$, yet the rank of $\mathrm{Hess}_f(X_0)$ is high.

Determinantal complexity of $\mathsf{IMM}_{n,d}$

We shall fix our target polynomial to be $\mathsf{IMM}_{n,d}$ where $N = n^2d$. We give an explicit construction of a point $X_0 \in \mathbb{F}^{n^2d}$ such that $\mathsf{IMM}_{n,d}(X_0) = 0$ and $\operatorname{rank}(\operatorname{Hess}_{\mathsf{IMM}_{n,d}}(X_0)) \geq d(n-1)$. First for the sake of completeness, we briefly recall the upper bound argument for the rank of $\operatorname{Hess}_{\det}(Y_0)$ from Section 2.1 of [CCL08].

Upper bound for the rank of $\operatorname{Hess}_{\det}(Y_0)$

When we take a partial derivative of the determinant polynomial with respect to the variable x_{ij}, the result is a minor that is obtained by striking out the row i and column j. The second order derivative of $\det(Y)$ with respect to the variables y_{ij} and $y_{k\ell}$ eliminates the rows $\{i,k\}$ and the columns $\{j,\ell\}$. Considering the form of Y_0, the non-zero entries in $\operatorname{Hess}_{\mathsf{Det}}(Y_0)$ are obtained only if $1 \in \{i,k\}$ and $1 \in \{j,\ell\}$ and thus $(ij,k\ell)$ are of the form $(11,tt)$ or $(t1,1t)$ or $(1t,t1)$ for any $t > 1$. Thus, $\operatorname{rank}(\operatorname{Hess}_{\det}(Y_0)) = 2m$.

Lower bound for the rank of $\operatorname{Hess}_{\mathsf{IMM}_{n,d}}(X_0)$

In this section, we shall prove Theorem 7.1.2. In particular, we give a polynomial time algorithm to construct a point X_0 explicitly such that $\mathsf{IMM}_{n,d}(X_0) = 0$ and $\operatorname{rank}(\operatorname{Hess}_{\mathsf{IMM}_{n,d}}(X_0)) \geq d(n-1)$. Since $\operatorname{rank}(\operatorname{Hess}_{\det}(Y_0)) = 2m$ and $\operatorname{rank}(\operatorname{Hess}_{\mathsf{IMM}_{n,d}}(X_0)) \leq \operatorname{rank}(\operatorname{Hess}_{\det}(Y_0))$, we get that $m = d(n-1)/2$. As mentioned in Section 7.1, the determinantal complexity of $\mathsf{IMM}_{n,d}(X)$ is $O(dn)$. Together, it implies that $m = \Theta(dn)$.

Theorem 7.3.1. *For any integers $n,d > 1$, there is a point $X_0 \in \mathbb{F}^{n^2d}$ such that $\mathsf{IMM}_{n,d}(X_0) = 0$ and $\operatorname{rank}(\operatorname{Hess}_{\mathsf{IMM}_{n,d}}(X_0)) \geq d(n-1)$. Moreover, the point X_0 can be constructed explicitly in polynomial time.*

Proof. We prove the theorem by induction on d, the degree of the polynomial. For the purpose of induction, we maintain that the entries indexed by the indices $(1,2),(1,3),\ldots,(1,n)$ of the matrix obtained after multiplying the first $(d-1)$ matrices are not all zero at X_0.

We first prove the base case for $d = 2$. The corresponding polynomial is $\mathsf{IMM}_{n,2}(X) = \sum_{i=1}^{n} x_{1i}^{(1)} x_{i1}^{(2)}$. It is easy to observe that the rank of the corresponding Hessian matrix is $2n > 2(n-1)$ at any point since each non-zero entry of the Hessian matrix is 1 and the structure of the Hessian matrix is the following:

$$\mathrm{Hess}_{\mathsf{IMM}_{n,2}}(X) = \begin{bmatrix} 0 & B_{12} \\ B_{21} & 0 \end{bmatrix}$$

where $B_{21} = B_{12}^T$ and the matrix B_{12} is formally described as

$$(B_{12})_{x_{ij}^{(1)} x_{kl}^{(2)}} = \begin{cases} 1 & \text{if } i = l = 1 \text{ and } j = k \\ 0 & \text{otherwise.} \end{cases}$$

We set the values of the variables as follows: $x_{11}^{(1)} = 0$, $x_{11}^{(2)} = 1$, $x_{21}^{(2)} = x_{31}^{(2)} = \cdots = x_{n1}^{(2)} = 0$ and $x_{12}^{(1)}, x_{13}^{(1)}, \ldots, x_{1n}^{(1)}$ to arbitrary values but not all to zero. The point thus obtained (say X_0) is clearly a zero of the polynomial $\mathsf{IMM}_{n,2}(X)$.

For induction hypothesis, assume that the statement of the theorem is true for the case where the number of matrices being multiplied is $\leq d$. Consider the polynomial $\mathsf{IMM}_{n,(d+1)}(X)$.

$$\mathsf{IMM}_{n,(d+1)}(X) = \sum_{i_1, i_2, \ldots, i_{d-1}, i_d \in [n]} x_{1 i_1}^{(1)} x_{i_1 i_2}^{(2)} \ldots x_{i_{(d-2)} i_{(d-1)}}^{(d-1)} x_{i_{(d-1)} i_d}^{(d)} x_{i_d 1}^{(d+1)}$$

Let the matrix obtained after multiplying the first d matrices be $(P_{k\ell})_{(k,\ell) \in [n] \times [n]}$ where

$$P_{k\ell}(X) = \sum_{i_1, i_2, \ldots, i_{d-1} \in [n]} x_{k i_1}^{(1)} x_{i_1 i_2}^{(2)} \ldots x_{i_{(d-2)} i_{(d-1)}}^{(d-1)} x_{i_{(d-1)} \ell}^{(d)} \text{ for } 1 \leq k, l \leq n.$$

Thus, we have the following expression.

$$\mathsf{IMM}_{n,(d+1)}(X) = P_{11}(X) x_{11}^{(d+1)} + P_{12}(X) x_{21}^{(d+1)} + \cdots + P_{1n}(X) x_{n1}^{(d+1)}$$

Now consider the Hessian matrix $\mathrm{Hess}_{\mathsf{IMM}_{n,d+1}}(X)$ which is a $(d+1)n^2 \times (d+1)n^2$ sized matrix.

$$\mathrm{Hess}_{\mathsf{IMM}_{n,d+1}}(X) = \begin{bmatrix} 0 & B_{1,2} & B_{1,3} & B_{1,4} & \cdots & B_{1,(d+1)} \\ B_{2,1} & 0 & B_{2,3} & B_{2,4} & \cdots & B_{2,(d+1)} \\ B_{3,1} & B_{3,2} & 0 & B_{3,4} & \cdots & B_{3,(d+1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ B_{(d+1),1} & B_{(d+1),2} & \cdots & \cdots & B_{(d+1),d} & 0 \end{bmatrix}$$

Each $B_{i,j}$ is a block of size $n^2 \times n^2$ and is indexed by the variables sets X_i and X_j respectively. Consider the block $B_{(d+1),d}$ which is indexed by the variable sets X_{d+1} and X_d. The only non-zero rows in $B_{(d+1),d}$ are indexed by the variables $x_{11}^{(d+1)}, x_{21}^{(d+1)}, \ldots, x_{n1}^{(d+1)}$. The potential non-zero entries for the row $x_{11}^{(d+1)}$ are indexed by the columns $x_{11}^{(d)}, x_{21}^{(d)}, \ldots, x_{n1}^{(d)}$. Similarly the potential non-zero entries for the row $x_{21}^{(d+1)}$ are indexed by the columns $x_{12}^{(d)}, x_{22}^{(d)}, \ldots, x_{n2}^{(d)}$ and so on.

Consider the entries corresponding to the indices $(x_{11}^{(d+1)}, x_{11}^{(d)}), (x_{11}^{(d+1)}, x_{21}^{(d)}), \ldots, (x_{11}^{(d+1)}, x_{n1}^{(d)})$, say $Q_1, Q_2, \ldots, Q_n$ respectively where

$$Q_j = \sum_{i_1, i_2, \ldots, i_{d-2} \in [n]} x_{1i_1}^{(1)} x_{i_1 i_2}^{(2)} \ldots x_{i_{(d-2)} j}^{(d-1)} \text{ for } 1 \leq j \leq n.$$

For the other rows indexed by the variables $x_{21}^{(d+1)}, x_{31}^{(d+1)}, \ldots, x_{n1}^{(d+1)}$, the sequence of potential non-zero entries is the same $(Q_1, Q_2, \ldots, Q_n)$ but their positions are shifted by a column compared to the previous non-zero row. Formally,

$$(B_{(d+1),d})_{x_{ij}^{(d+1)} x_{kl}^{(d)}} = \begin{cases} Q_k & \text{if } j = 1,\ l = i, \text{ and } i, k \in [n] \\ 0 & \text{otherwise.} \end{cases}$$

$Q_1, Q_2, \ldots, Q_n$ are also the entries indexed by the indices $(1,1), (1,2), \ldots, (1,n)$ of the

matrix obtained after multiplying the first $(d-1)$ matrices. By induction hypothesis, we know that the entries indexed by the indices $(1,2),\ldots,(1,n)$ are not all zero at the point X_0, the zero of the polynomial $\mathsf{IMM}_{n,d}(X)$. This also makes the rows indexed by the variables $x_{11}^{(d+1)}, x_{21}^{(d+1)},\ldots,x_{n1}^{(d+1)}$ linearly independent. It is important to note that $P_{11}(X) = \mathsf{IMM}_{n,d}(X)$.

Now, let us define a point such that it is a zero of the polynomial $\mathsf{IMM}_{n,(d+1)}(X)$. Let X_0 be the zero of the polynomial $P_{11}(X) = \mathsf{IMM}_{n,d}(X)$. Now to construct the new point, we inductively fix the variables appearing in $P_{11}(X)$ by the values assigned by X_0. We set $x_{11}^{(d+1)} = 1$ and $x_{21}^{(d+1)} = x_{31}^{(d+1)} = \cdots = x_{n1}^{(d+1)} = 0$. We will fix the rest of the variables later. We call the new point which is a zero of the polynomial $\mathsf{IMM}_{n,(d+1)}(X)$, as X_0 as well.

Now, consider the first $d \times d$ blocks of the Hessian matrix $\mathrm{Hess}_{\mathsf{IMM}_{n,(d+1)}}(X_0)$. It precisely represents the Hessian matrix of $P_{11}(X)$ which is also the Hessian matrix of the polynomial $\mathsf{IMM}_{n,d}(X)$ at the point X_0. This can be easily seen from the setting of the variables $x_{11}^{(d+1)} = 1$ and $x_{21}^{(d+1)} = x_{31}^{(d+1)} = \cdots = x_{n1}^{(d+1)} = 0$. By induction hypothesis, the rank of this minor of $\mathrm{Hess}_{\mathsf{IMM}_{n,(d+1)}}(X_0)$ is at least $d(n-1)$. The only non-zero entries in the columns indexed by the variable set $X^{(d)}$ are indexed by the variables $x_{11}^{(d)}, x_{21}^{(d)},\ldots,x_{n1}^{(d)}$. This is because the other variables of X_d do not appear in $\mathsf{IMM}_{n,d}(X)$. The row in $B_{(d+1)d}$ indexed by $x_{11}^{(d+1)}$ is the only row that interferes with any of the rows of $B_{1d}, B_{2d},\ldots,B_{dd}$. The rows indexed by the variables $x_{21}^{(d+1)}, x_{31}^{(d+1)},\ldots,x_{n1}^{(d+1)}$ in $B_{(d+1)d}$ are linearly independent of the rows of $B_{1d}, B_{2d},\ldots,B_{dd}$. Hence the rank of $\mathrm{Hess}_{\mathsf{IMM}_{n,(d+1)}}$ at the point described is $\geq (d+1)(n-1)$.

For the purpose of induction, we must verify that the entries indexed by the indices $(1,2),(1,3),\ldots,(1,n)$ of the matrix obtained after multiplying the first d matrices are not all zero at X_0. These entries are the polynomials $P_{12}, P_{13},\ldots,P_{1n}$. We shall express each of the polynomials in terms of $Q_1, Q_2,\ldots,Q_n$ as follows.

$$P_{1j} = Q_1 x_{1j}^{(d)} + Q_2 x_{2j}^{(d)} + \cdots + Q_n x_{nj}^{(d)} \text{ for } 2 \leq j \leq n.$$

By induction hypothesis, we already know that $Q_2, Q_3, \ldots, Q_n$ are not all zero at X_0. Notice that the variables in $X^{(d)} \setminus \{x_{11}^{(d)}, x_{21}^{(d)}, \ldots, x_{n1}^{(d)}\}$ were never set in the previous steps of induction. This is because of the fact that they do not appear in the polynomial P_{11}. Therefore, we can fix these variables suitably such that $P_{12}, P_{13}, \ldots, P_{1n}$ are not all zero when evaluated at the point X_0 (in fact, we can make all of them non-zero). It is clear that we construct the point X_0 in polynomial time. This completes the proof. □

7.4. Formula size lower bound for $\mathsf{IMM}_{n,d}$

In general, a *strong* enough lower bound on the determinantal complexity of a polynomial also implies a lower bound on the formula complexity. But here, in the case of the iterated matrix multiplication, the best bound on the determinantal complexity that we can get is $O(dn)$ for it has an algebraic branching program of that size. This does not imply any non trivial bound on the formula complexity of the polynomial.

In this section, we shall prove a super-linear but subquadratic lower bound on the size of any formula that computes the $\mathsf{IMM}_{n,d}$ polynomial. The following proof is an adaptation of the proof strategy of Kalorkoti [Kal85]. Let us first recall the notion of algebraic independence and transcendence degree.

Definition 7.4.1. *A set of polynomials $f_1, f_2, \ldots, f_m \in \mathbb{F}[X]$ are said to be* algebraically independent *if the only polynomial $F \in \mathbb{F}[y_1, y_2, \ldots, y_m]$ satisfying $F(f_1, f_2, \ldots, f_m) \equiv 0$ is the zero polynomial.*

The transcendental degree *of the polynomials $f_1, f_2, \ldots, f_m \in \mathbb{F}[X]$, denoted by* $\mathrm{trdeg}(f_1, f_2, \ldots, f_m)$*, is the maximal size of the subset S of $[m]$ such that $\{f_i\}_{i \in S}$ are algebraically independent.* ◊

We shall now define the notion of transcendental degree of a polynomial with respect to a subset of its variables.

Definition 7.4.2. *Let $f \in \mathbb{F}[X]$ be a polynomial and $X' \subset X$ a set of variables. Let f be expressed as $\sum_{m \in M} f_m \cdot m$ where M is set of all monomials over the variables in X' and degree at most* $\deg(f)$*. The complexity measure* $\mathrm{trdeg}_{X'}(f)$ *is defined as the transcendental*

degree of $\{f_m\}_{m\in M}$. ◊

The following lemma is the key to the formula size lower bound in [Kal85] (cf. [Sap15]).

Lemma 7.4.3. *Let* $f \in \mathbb{F}[X]$ *and* $X_1, X_2, \ldots, X_t$ *be a partition of* X. *Then every arithmetic formula for* f *must be of size* $\Omega(\sum_{i\in[t]} \mathrm{trdeg}_{X_i}(f))$.

Theorem 7.4.4. *For all integers* $n, d > 0$, *any arithmetic formula computing the* $\mathsf{IMM}_{n,d}(X)$ *polynomial must be of size* $\Omega(d\,n^3)$.

Proof. The main idea is to find a suitable partition of the input variables. For simplicity we assume that d is a multiple of four. Let $M_1, M_2, \ldots, M_d$ be the generic $n \times n$ matrices being multiplied in $\mathsf{IMM}_{n,d}(X)$ polynomial. For all i such that i is of the form $4t+1$ or $4t+2$, $t \in [0, d/4-1]$, partition the variables in the matrices M_i and M_{i+2} by grouping jth row of M_i and jth column of M_{i+2} together, for all $j \in [n]$. We shall denote such a set by $X_{ij} = \{x^{(i)}_{j,1}, \ldots, x^{(i)}_{j,n}, x^{(i+2)}_{1,j}, \ldots, x^{(i+2)}_{n,j}\}$. The final partition of the variables is as follows.

$$X \quad = \quad \sqcup_{i\in\{4t+1,4t+2 : t\in[0,d/4-1]\}} \sqcup_{1\le j\le n} X_{ij}.$$

Now we express the polynomial $\mathsf{IMM}_{n,d}(X)$ w.r.t the set of variables X_{ij} as explained in the definition 7.4.2.

$$\mathsf{IMM}_{n,d}(X) \quad = \quad \sum_{k,\ell\in[n]} (x^{(i+1)}_{k,l} P_1) \cdot x^{(i)}_{j,k} x^{(i+2)}_{\ell,j} + P_2.$$

The first summand in the above expression is the summation of all monomials that contain the variables $x^{(i)}_{j,k}$ and $x^{(i+2)}_{\ell,j}$ for all $k, \ell \in [n]$ and P_2 is the summation of the rest of the monomials. Formally,

$$P_1(X) = \sum_{a_t\in[n]} x^{(1)}_{1,a_1} \ldots x^{(i-1)}_{a_{i-2},j} x^{(i+3)}_{j,a_{i+3}} \ldots x^{(d)}_{a_{d-1},1}.$$

Now, $\mathrm{trdeg}_{X_{ij}}(\mathsf{IMM}_{n,d}(X))$ is at least the transcendental degree of the set of polyno-

mial $\mathscr{P} = \{P_1 \cdot x_{k,\ell}^{(i+1)}\}_{k,\ell\in[n]}$. Notice that $|\mathscr{P}| = n^2$. Let us introduce new variables $Y = \{y_1, \ldots, y_{n^2}\}$ to lexicographically correspond to polynomials in $\mathscr{P} = \{P_1 \cdot x_{k,\ell}^{(i+1)}\}_{k,\ell\in[n]}$. To prove their algebraic independence, we need to prove that there is no non-zero polynomial over $\mathbb{F}[Y]$ such that substitution for y_i with the corresponding polynomials in $\mathscr{P}$ makes it a zero polynomial over $\mathbb{F}[X]$.

For the sake of contradiction, let us assume that there is a polynomial $g \in \mathbb{F}[Y]$ that annihilates the polynomials in $\mathscr{P}$. Consider two distinct monomials $m_1 = y_1^{\alpha_1} y_2^{\alpha_2} \ldots y_{n^2}^{\alpha_{n^2}}$ and $m_2 = y_1^{\beta_1} y_2^{\beta_2} \ldots y_{n^2}^{\beta_{n^2}}$ in g such that $\bar{\alpha} \neq \bar{\beta}$. Consider $m_1' = m_1|_{y_i \leftarrow P_i \in \mathscr{P}}$ and $m_2' = m_2|_{y_i \leftarrow P_i \in \mathscr{P}}$. We can see that $m_1' = P_1^{\left(\sum_{r\in[n^2]}\alpha_r\right)} \left(x_{1,1}^{(i+1)}\right)^{\alpha_1} \left(x_{1,2}^{(i+1)}\right)^{\alpha_2} \ldots \left(x_{n,n}^{(i+1)}\right)^{\alpha_{n^2}}$ and $m_2' = P_1^{\left(\sum_{r\in[n^2]}\beta_r\right)} \left(x_{1,1}^{(i+1)}\right)^{\beta_1} \left(x_{1,2}^{(i+1)}\right)^{\beta_2} \ldots \left(x_{n,n}^{(i+1)}\right)^{\beta_{n^2}}$.

W.l.o.g, let us assume that $\alpha_1 > \beta_1$. The overall degree of $x_{1,1}^{(i)}$ in m_1' is equal to α_1 and similarly the overall degree of the variable $x_{1,1}^{(i)}$ in m_2' is equal to β_1, and hence the monomials in m_1' and m_2' are distinct. So, one can conclude that no two distinct monomials in g can share a monomial after the substitution. Hence, the polynomial g can not annihilate the polynomials in $\mathscr{P}$. From Lemma 7.4.3, we get that the size of any arithmetic formula computing $\mathsf{IMM}_{n,d}(X)$ is of size at least $\sum_{i,j} \mathrm{trdeg}_{X_{i,j}}(\mathsf{IMM}_{n,d}(X)) = \Omega(dn^3)$.

□

Part IV.

Tensor rank

8

Arithmetic formula size lower bounds from the tensor rank

Raz [Raz10] showed that for any n and d, such that $\omega(1) \leq d \leq O\left(\frac{\log n}{\log\log n}\right)$, constructing explicit tensors $T : [n]^d \to \mathbb{F}$ of high enough rank would imply superpolynomial lower bounds for arithmetic formulas over the field $\mathbb{F}$. Using the additional structure we obtain from the depth reduction for arithmetic formulas (Chapter 3), we give a new and arguably simpler proof of this connection. We also extend this result for homogeneous formulas to show that, in fact, the connection holds for any d such that $\omega(1) \leq d \leq n^{o(1)}$.

8.1. Introduction

Proving size lower bounds for arithmetic formulas computing explicit polynomials has been a tough task. Kalorkoti [Kal85] proved a quadratic lower bound using *transcendence degree* as a complexity measure. This measure does not yield lower bounds better than quadratic.

In an apriori surprising result, Raz [Raz10] showed that for any n and d, such that $\omega(1) \leq d \leq O\left(\frac{\log n}{\log\log n}\right)$, constructing explicit tensors $T : [n]^d \to \mathbb{F}$ of high enough rank would imply super-polynomial lower bounds for arithmetic formulas over the field $\mathbb{F}$. Using the additional structure we obtain from our proof of the depth reduction for arithmetic formulas (cf. Chapter 3), we give a new and arguably a simpler proof of this connection. We also extend the result to show that, in fact, such connection holds in the case of homogeneous formulas for any d such that $\omega(1) \leq d \leq n^{o(1)}$.

In [Raz10] Raz showed us that an arithmetic formula can be homogenized efficiently

when $d = O(\log n)$. Combining this fact with our extended result we show that, for any n and d, such that $\omega(1) \leq d \leq O(\log n)$, constructing explicit tensors $T : [n]^d \to \mathbb{F}$ of high enough rank would imply super-polynomial lower bounds for arithmetic formulas over the field $\mathbb{F}$. This improves upon the range of parameters of Raz [Raz10].

8.2. Background

Tensors

Given two vector spaces U and V over $\mathbb{F}$, we can define a linear map $\phi : U \to V$. This map can be represented as a matrix. Similarly for the vector spaces $\{V_1, V_2, \ldots, V_d\}$ of dimension $\{m_1, m_2, \ldots, m_d\}$ respectively, we can define a map $\phi' : V_1 \times V_2 \times \cdots \times V_{d-1} \to V_d$. This map can be represented as a natural *higher dimensional* analogue of a matrix. We refer to that as a tensor. Formally, we have the following definition.

Definition 8.2.1 (Tensor). *A tensor T is a map of the form*

$$T : V_1 \times \cdots \times V_d \longrightarrow \mathbb{F}$$

where each V_i is a vector space over $\mathbb{F}$, of dimension say m_i. The parameter d is called the order *of the tensor, and we say that the* shape of T *is* $[m_1] \times \cdots \times [m_d]$. ◊

A tensor T is linear in every coordinate. That is,

$$T(\mathbf{v}_1, \ldots, \alpha\mathbf{v}_i + \beta\mathbf{v}_i', \ldots, \mathbf{v}_d) = \alpha T(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_d) + \beta T(\mathbf{v}_1, \ldots, \mathbf{v}_i', \ldots, \mathbf{v}_d).$$

So a tensor can indeed be thought of as filling up a d-dimensional array of shape $[m_1] \times \cdots \times [m_d]$ by field elements, the same way an $m \times n$ matrix is specified by an $m \times n$ array filled up with field elements. Indeed, a matrix is nothing but an order-2 tensor.

It would sometimes be useful to switch between the two notions of thinking of a tensor as a multilinear map from $V_1 \times \cdots \times V_d$ to $\mathbb{F}$ and thinking of a tensor as just a map from $[m_1] \times \cdots \times [m_d]$ to $\mathbb{F}$.

Tensors as polynomials

In this setting, it would be useful to think of tensors as a restricted form of multilinear polynomials that are called *set-multilinear polynomials*.

Definition 8.2.2 (Set-multilinear polynomials)**.** *Let $X = X_1 \sqcup \cdots \sqcup X_d$ be a partition of variables and let $|X_i| = m_i$. A polynomial $f(X)$ is said to be* set-multilinear *with respect to the above partition if every monomial M in f satisfies $|\mathrm{var}(M) \cap X_i| \leq 1$ for all $i \in [d]$. A set-multilinear formula is an arithmetic formula where the polynomial computed at every node is a set-multilinear polynomial.* ◊

In other words, each monomial in f picks up at most one variable from each part in the partition. It is easy to see that many natural polynomials that we consider in this thesis such as Det, IMM, NW and Perm are all set-multilinear for an appropriate partition of variables.

Observation 8.2.3. *For any tensor T of shape $[m_1] \times \cdots \times [m_d]$, we can associate a set-multilinear polynomial $f(X)$ where $X = X_1 \sqcup \cdots \sqcup X_d$ and $X_i = \left\{x_{i1}, \ldots, x_{im_i}\right\}$ as follows.*

$$f(X) \quad = \quad \sum_{\substack{1 \leq i_j \leq m_j \\ \forall j \in [d]}} T(i_1, \ldots, i_d) \cdot x_{1i_1} \cdots x_{di_d}. \tag{8.2.4}$$

The same also holds in the other direction where we can interpret any set-multilinear polynomial as an appropriate tensor.

Rank of a tensor

The notion of the *rank of a tensor* is a generalization of the notion of the rank of a matrix.

Definition 8.2.5 (Elementary tensors, and tensor rank)**.** *For any vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d$ in $V_1, V_2, \cdots, V_d$ respectively, we define the tensor $\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \cdots \otimes \mathbf{v}_d$ to be the tensor E given by $E[j_1, \ldots, j_d] = (\mathbf{v}_1)_{j_1} (\mathbf{v}_2)_{j_2} \cdots (\mathbf{v}_d)_{j_d}$.* ◊

We shall call such tensors as *elementary tensors* or *rank*-1 *tensors*. For an arbitrary tensor

T, the *tensor rank of* T, denoted by $\mathrm{TensorRank}(T)$, is the smallest r such that T can be expressed as a sum of r elementary tensors.

Let us consider the *set-multilinear* polynomial setting as in (8.2.4). It is easy to see that a rank-1 tensor is precisely a *set-multilinear* product of linear forms such as $\ell_1(X_1) \cdot \ell_2(X_2) \cdot \ldots \cdot \ell_d(X_d)$ where each $\ell_i(X_i)$ is a linear polynomial over the variables in X_i.

Properties of tensor rank

The following are a couple of basic properties that follow almost immediately from the definitions.

Lemma 8.2.6 (Sub-additivity of tensor rank)**.** *Let T_1 and T_2 be two tensors of the same shape and order. Then, if $T = T_1 + T_2$, then* $\mathrm{TensorRank}(T) \leq \mathrm{TensorRank}(T_1) + \mathrm{TensorRank}(T_2)$.

Lemma 8.2.7 (Sub-multiplicativity of tensor rank)**.** *Let $T_1 : V_1 \times \cdots \times V_{d_1} \to \mathbb{F}$ and $T_2 : W_1 \times \cdots \times W_{d_2} \to \mathbb{F}$ be two tensors. Then if $T = T_1 \otimes T_2$ given by $T[i_1, \ldots, i_{d_1}, j_1, \ldots, j_{d_2}] = T_1[i_1, \ldots, i_{d_1}] \cdot T_2[j_1, \ldots, j_{d_2}]$, then* $\mathrm{TensorRank}(T) \leq \mathrm{TensorRank}(T_1) \cdot \mathrm{TensorRank}(T_2)$.

The following is a trivial upper bound on the tensor rank of set-multilinear polynomial f of degree d over $X = X_1 \sqcup \cdots \sqcup X_d$.

Lemma 8.2.8. *Let f be a set-multilinear polynomial with respect to the partition $X = X_1 \sqcup \cdots \sqcup X_d$ and say $n_i = |X_i|$. Then,* $\mathrm{TensorRank}(f) \leq \frac{\prod_{i=1}^{d} n_i}{\max_i n_i}$. *In particular, if $n_i = n$ for all $i \in [d]$, then* $\mathrm{TensorRank}(f) \leq n^{d-1}$.

A counting argument would imply that there do exist tensors of rank at least n^{d-1}/d. This is due to the fact that each elementary tensor has nd *degrees of freedom* and an arbitrary tensor has n^d *degrees of freedom*.[1]

So, it is a natural question to understand if we can construct explicit tensors of high rank? Raz [Raz10] showed that in certain regimes of parameters involved, an answer to the above question would yield size lower bounds for arithmetic formulas. We shall now elaborate on this.

[1] One might think that the above upper bound of n^{d-1} should be tight. Bizarrely, it is not! For example (cf. [Pam85]), the maximum rank of any tensor of shape $2 \times 2 \times 2$ is 3 and not 4 as one might expect!

8.3. Tensor rank of small formulas

Henceforth, the variables in X are partitioned as $X = X_1 \sqcup \cdots \sqcup X_d$ with $|X_i| = n$ for all $i \in [d]$.

8.3.1. Overview of Raz's proof

The main motivating question of Raz [Raz10] was the following:

> If f is a set-multilinear polynomial that is computed by a small formula, then what can one say about its tensor rank?

Raz gave a partial[2] answer to this question by showing the following result.

Theorem 8.3.1. *Let Φ be a formula of size $s \leq n^c$ computing a set-multilinear polynomial $f(X)$ with respect to the partition $X = X_1 \sqcup \cdots \sqcup X_d$. If $d = O(\log n / \log\log n)$, then* $\mathrm{TensorRank}(f) \leq n^{d(1-1/\exp(O(c)))}$.

To prove Theorem 8.3.1, Raz [Raz10] first showed that when d is small compared to n (specifically, $d = O(\log n / \log\log n)$), any small formula can be *homogenized* and then be converted to a *set-multilinear* formula with just a polynomial over-head. This is interesting and surprising in its own right [3].

Theorem 8.3.2 ([Raz10])**.** *Let Φ be a formula of size s computing an n-variate homogeneous polynomial f of degree d. Then, there is a* homogeneous formula Φ' *that also computes f of size at most* $\mathrm{poly}\left(s, \binom{d+\log s}{d}\right)$. *In particular, if $d = O(\log n)$ and $n = \mathrm{poly}(n)$ then we have* $\mathrm{size}(\Phi') = \mathrm{poly}(n)$ *as well.*

Theorem 8.3.3 ([Raz10])**.** *Suppose $d = O\left(\frac{\log n}{\log\log n}\right)$. If Φ is a formula of size $s = \mathrm{poly}(n)$ that computes a set-multilinear polynomial $f(X_1, \cdots, X_d)$, then there is a* set-multilinear *formula of* $\mathrm{poly}(s)$ *size that computes f as well.*

[2] Partial in the sense that we do not know if the bound is tight.

[3] It was believed that transforming a formula into a homogeneous formula would cause a super-polynomial blow up in its size if the degree of the polynomial computed by the formula is growing with n [NW97].

He then proceeds to show that set-multilinear formulas of polynomial size can only compute polynomials with tensor rank non-trivially far from the upper bound of n^{d-1}. More formally, he shows the following theorem.

Theorem 8.3.4 ([Raz10]). *Let Φ be a set-multilinear formula of size $s \leq n^c$ computing a polynomial $f(X_1, \cdots, X_d)$. Then* $\text{TensorRank}(f) < \frac{n^d}{n^{d/\exp(O(c))}}$.

It is immediately clear that Theorem 8.3.3 and Theorem 8.3.4 imply Theorem 8.3.1[4]. We shall now present a simpler proof of Theorem 8.3.4 using Theorem 3.3.3.

8.3.2. Our proof

Crux of our arguments:

Let f be a set-multilinear polynomial over $X_1 \sqcup \cdots \sqcup X_d$ such that $|X_i| = n$ for all i. The tensor rank is at most n^{d-1}. But if also know that $f = f_1 \times f_2$ where the product respects set-multilinearity, then

$$\text{TensorRank}(f) \leq \text{TensorRank}(f_1) \cdot \text{TensorRank}(f_2) \leq n^{d_1-1} \cdot n^{d-d_1-1} = n^{d-2}$$

where d_1 is the degree of f_1.

If a set multilinear polynomial that corresponds to an explicit tensor can be expressed as a summation over a *few* summands each of which has *large* number of factors, then we can get non-trivial bounds on the tensor rank.

Proof of Theorem 8.3.4. We shall start with the set-multilinear formula Φ of size n^c and reduce it to a depth-4 circuit via Theorem 3.3.3 for a bottom degree parameter t that we shall fix shortly. It is fairly straightforward to observe that the depth reduction preserves multilinearity and set-multilinearity as well. Therefore we now have a set-multilinear expression of the form

$$f \quad = \quad T_1 + \cdots + T_{s'}$$

[4] We refer the reader to Raz's paper [Raz10] or a survey by Saptharishi [Sap15] for a full proof of Theorem 8.3.2 and Theorem 8.3.3.

where $s' \leq s^{10(d/t)} = n^{10c(d/t)}$ and each $T_i = Q_{i1} \cdots Q_{ia_i}$ is a set-multilinear product. Let us fix one such term $T = Q_1 \cdots Q_a$ and we know that this is a set-multilinear product with $a \geq \frac{d \log t}{10t}$ non-trivial factors (by Theorem 3.3.3). Let $d_i = \deg(Q_i)$. By the sub-multiplicativity of tensor rank (Lemma 8.2.7) and the trivial upper bound (Lemma 8.2.8) we have

$$\begin{aligned} & \text{TensorRank}(T) \leq n^{d_1 - 1} \cdots n^{d_a - 1} \\ \implies & \text{TensorRank}(f) \leq s' \cdot n^{d-a} && \text{(Lemma 8.2.6)} \\ & \leq \frac{n^d}{n^{a - 10c(d/t)}}. \end{aligned}$$

Let us focus on the exponent of n in the denominator. Using the lower bound on a from Theorem 3.3.3, we get

$$a - 10c(d/t) \quad > \quad \frac{d \log t}{10t} - 10c\frac{d}{t} \quad = \quad \frac{d}{t}\left(\frac{\log t}{10} - 10c\right).$$

If we set $\frac{\log t}{10} = 11c$, then we get $a - 10c(d/t) > cd/t = d/\exp(O(c))$. Hence,

$$\text{TensorRank}(f) \quad < \quad \frac{n^d}{n^{d/\exp(O(c))}}.$$

□

We would like to remark that, in spirit, a tensor rank upper bound for formulas is essentially a form of non-trivial reduction to set-multilinear depth three circuits. In this sense, this connection between tensor rank upper bound and reduction to depth four is perhaps not too un-natural.

Remark 8.3.5. *If instead of a general set-multilinear formula, had we started with a constant depth set-multilinear formula, we would have obtained a slightly better upper bound (better dependence on c) on the tensor rank of f. (cf. [CKSV16, Sap15]).* ◊

8.4. Tensor rank upper bound for homogeneous formulas

The result of Raz [Raz10] required d to be $O(\log n/\log\log n)$ to be able to *set-multilinearize* the formula without much overhead. However, we show that via the improved depth reduction, we can delay the set-multilinearization until a later stage and thus get the same upper bound on the tensor rank for much larger d, provided that the formula we started with was homogeneous.

Theorem 8.4.1. *Let f be a set-multilinear polynomial with respect to $X = X_1 \sqcup \cdots \sqcup X_d$ that is computed by a homogeneous formula (not necessarily set-multilinear) Φ of size $s = n^c$. If d is* sub-polynomial *in n, that is* $\log d = o(\log n)$, *then* $\mathrm{TensorRank}(f) < \frac{n^d}{n^{d/\exp(O(c))}}$.

Proof. As earlier, we shall start with the formula Φ of size n^c and reduce it to a $\Sigma\Pi\Sigma\Pi^{[t]}$ formula Φ' of size $n^{10c(d/t)}$ (using Theorem 3.3.3) for a t that shall be chosen shortly. Again, Φ' is a sum of terms of the form $T = Q_1 \cdots Q_a$, a product of $a \geq \frac{d \log t}{10t}$ non-trivial factors. The difference here is that this is not necessarily a set-multilinear product. Let $d_i = \deg(Q_i)$. Among the monomials in Q_i, there may be some that are divisible by two or more variables from some part X_j and others that are products of variables from distinct parts. For any $S \subset [d]$ let $Q_{i,S}$ be the sum of monomials of Q_i that is a product of exactly only variable from each X_j for $j \in S$. Note that no monomials of Q_i that is a product of two or more variables from some X_j can contribute to a set-multilinear monomial of f. Thus, if $\mathrm{SML}(T)$ is the restriction of T to just the set-multilinear monomials of T, then

$$\mathrm{SML}(T) \quad = \sum_{\substack{S_1 \sqcup \cdots \sqcup S_a = [d] \\ |S_i| = d_i}} Q_{1,S_1} \cdots Q_{a,S_a}.$$

We can observe that the tensor rank of each summand is upper bounded by $n^{d_1-1} n^{d_2-1} \cdots n^{d_a-1}$ and the number of summands is at most $\binom{d}{d_1}\binom{d-d_1}{d_2}\cdots\binom{d-\sum_{i=1}^{a-1} d_i}{d_a}$. Us-

ing Lemma 8.2.6 and Lemma 8.2.7, we get the following.

$$\begin{aligned} \text{TensorRank}(\text{SML}(T)) &\leq \frac{n^d}{n^a} \cdot \binom{d}{d_1\, d_2\, \cdots\, d_a} \\ &\leq n^{d-a} \cdot d^d \\ &= n^{d-a} \cdot n^{d \log d / \log n} \\ \implies \quad \text{TensorRank}(f) &\leq n^d / n^{a - 10c(d/t) - d \log d / \log n}. \end{aligned}$$

Again, let us focus on the exponent in the denominator

$$a - \frac{10c \cdot d}{t} - \frac{d \log d}{\log n} \;>\; \frac{d}{t}\left(\frac{\log t}{10} - 10c - \frac{t \log d}{\log n}\right).$$

Once again we shall set $t = 2^{O(c)}$ so that $\frac{\log t}{10} - 10c = c$ and since $\log d = o(\log n)$ it follows that

$$\frac{d}{t}\left(\frac{\log t}{10} - 10c - \frac{t \log d}{\log n}\right) > \frac{d}{\exp(O(c))}.$$

Hence,

$$\text{TensorRank}(f) \;<\; \frac{n^d}{n^{d/\exp(O(c))}}.$$

□

An improvement

Further, we observe that we can improve upon the range of parameters in Theorem 8.3.1.

Corollary 8.4.2. *Let Φ be a formula of size $s \leq n^c$ computing a set-multilinear polynomial $f(X)$ with respect to $X = X_1 \sqcup \cdots \sqcup X_d$. If $d = O(\log n)$, then* $\text{TensorRank}(f) \leq n^{d(1-1/\exp(O(c)))}$.

The proof of this follows directly from Theorem 8.3.2 and Theorem 8.4.1.

9 Conclusion

In Chapter 4 we have proved exponential lower bounds against $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ circuits computing explicit polynomials ($\mathsf{NW}_{n,r}$ and $\mathsf{IMM}_{n,n}$) using the method of shifted partial derivatives. But it is now well understood that this technique will not help us cross the chasm and obtain the lower bounds of the order of $n^{\omega(d/t)}$ to prove Valiant's hypothesis [GKKS14]. We do need a new set of techniques to cross the chasm. On the other hand, we still do not have better bounds for the determinant and the permanent polynomials. A natural question to ask here is to see if we can improve the lower bound against the $\Sigma\Pi^{[O(\sqrt{n})]}\Sigma\Pi^{[\sqrt{n}]}$ circuits computing either the determinant or the permanent polynomial to $2^{\omega(\sqrt{n})}$.

In Chapter 5, over fixed size finite fields, we showed *tight* lower bounds against $\Sigma\Pi\Sigma$ circuits computing two explicit polynomials, $\mathsf{NW}_{n,r}$ and $\mathsf{IMM}_{n,n}$. This was done by improving upon and adapting the work of Grigoriev and Karpinski [GK98]. It was indeed very surprising to know that a polynomial in VP, the iterated matrix multiplication polynomial, is harder than the permanent polynomial at depth three over fixed size finite fields. Keeping this in mind and because of the non-existence of the chasm over small finite fields over depth three, it is conceivable to expect a $2^{\omega(n)}$ lower bound against $\Sigma\Pi\Sigma$ circuits computing the Det_n polynomial.

In Chapter 6, we showed exponential size lower bounds against two classes of depth five powering circuits. It is now understood that extending the exponential size lower bounds to other classes of powering circuits would also need new techniques. However, we believe that slight modification of our current measure would yield results with

some circuit models that are very closely related to those that we consider. Otherwise, because of the range of parameters involved, neither the method of *projected shifted partial derivatives* nor *shifted projected partial derivatives* could be helpful [KLSS14].

In Chapter 7, we proved a lower bound on the determinantal complexity of the iterated matrix multiplication polynomial. The lower bound is away from the upper bound by a factor of 0.5. It is important to note that this result is true over all fields. However, it is conceivable that an adaptation of the argument of Yabe [Yab15] could yield a better constant factor, over $\mathbb{R}$. The question of proving a quadratic lower bound for the Nisan Wigderson polynomial still lies open. We ask if the methods of projected shifted partials or the shifted projected partials be of some help with this problem.

In Chapter 8, we extended the work of Raz [Raz10] and showed that the existence of a *stronger* connection between the size of the homogeneous formulas computing set multiliner polynomials and the rank of the corresponding tensor. Thus, for a range of parameters, strong enough lower bounds on the tensor rank for explicit tensors would imply (homogeneous) formula size lower bounds. It is also conceivable that such a connection can be established for any circuit model that exhibits a combinatorial property (as discussed in Chapter 8).

Apart from the questions and directions mentioned above, there are a lot of interesting problems within the relevance of the topics mentioned in this thesis. We firmly believe that the answer to some of these questions could be through a combination of techniques from the existing knowledge base. For others, we need to discover newer techniques.

Bibliography

[AJMV98] Eric Allender, Jia Jiao, Meena Mahajan, and V. Vinay. Non-Commutative Arithmetic Circuits: Depth Reduction and Size Lower Bounds. *Theoretical Computer Science*, 209(1-2):47–86, 1998. Pre-print available at eccc:TR95-043.

[Alo99] Noga Alon. Combinatorial Nullstellensatz. *Combinatorics, Probability and Computing*, 8, 1999.

[AV08] Manindra Agrawal and V. Vinay. Arithmetic Circuits: A Chasm at Depth Four. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2008)*, pages 67–75, 2008. Pre-print available at eccc:TR08-062.

[BLS16] Nikhil Balaji, Nutan Limaye, and Srikanth Srinivasan. An Almost Cubic Lower Bound for ΣΠΣ Circuits Computing a Polynomial in VP. 2016.

[BLSS17] Nikhil Balaji, Nutan Limaye, Sai Sandeep, and Srikanth Srinivasan. Personal communication, 2017.

[Bre74] Richard P. Brent. The Parallel Evaluation of General Arithmetic Expressions. *Journal of the ACM*, 21(2):201–206, April 1974.

[BS83] Walter Baur and Volker Strassen. The Complexity of Partial Derivatives. *Theoretical Computer Science*, 22:317–330, 1983.

[Bür00] Peter Bürgisser. *Completeness and Reduction in Algebraic Complexity Theory*. Algorithms and Computation in Mathematics. Springer, 2000. Online version: http://math-www.uni-paderborn.de/agpb/work/habil.ps.

[Cai90] Jin-Yi Cai. A note on the determinant and permanent problem. *Information and Computation*, 84(1):119–127, 1990.

[CCL08] Jin-Yi Cai, Xi Chen, and Dong Li. A quadratic lower bound for the permanent and determinant problem over any characteristic$\neq 2$. In *Symposium on Theory of Computing (STOC)*, pages 491–498. ACM, 2008.

[CKSV16] Suryajith Chillara, Mrinal Kumar, Ramprasad Saptharishi, and V. Vinay. The Chasm at Depth Four, and Tensor Rank : Old results, new insights. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:96, 2016.

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

[CM14a] Suryajith Chillara and Partha Mukhopadhyay. Depth-4 Lower Bounds, Determinantal Complexity : A Unified Approach. *Proceedings of the 31st Symposium on Theoretical Aspects of Computer Science (STACS 2014)*, 2014. Pre-print available at arXiv:1308.1640.

[CM14b] Suryajith Chillara and Partha Mukhopadhyay. On the Limits of Depth Reduction at Depth 3 Over Small Finite Fields. In *Mathematical Foundations of Computer Science (MFCS)*, pages 177–188, 2014. Pre-print available at arXiv:1401.0189.

[CS17] Suryajith Chillara and Ramprasad Saptharishi. Exponential lower bounds for some depth five powering circuits. Manuscript, 2017.

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

[ERS16] Christian Engels, B. V. Raghavendra Rao, and Karteek Sreenivasaiah. Lower Bounds and Identity Testing for Projections of Power Symmetric Polynomials. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:153, 2016.

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

[FLMS14] Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth 4 formulas computing iterated matrix multiplication. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC 2014)*, pages 128–135, 2014. Pre-print available at `eccc:TR13-100`.

[GK98] Dima Grigoriev and Marek Karpinski. An Exponential Lower Bound for Depth 3 Arithmetic Circuits. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC 1998)*, pages 577–582, 1998.

[GKKS13] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic Circuits: A Chasm at Depth Three. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2013)*, pages 578–587, 2013. Pre-print available at `eccc:TR13-026`.

[GKKS14] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Approaching the Chasm at Depth Four. *Journal of the ACM*, 61(6):33:1–33:16, 2014. Preliminary version in the *28th Annual IEEE Conference on Computational Complexity (CCC 2013)*. Pre-print available at `eccc:TR12-098`.

[HY11] Pavel Hrubeš and Amir Yehudayoff. Homogeneous Formulas and Symmetric Polynomials. *Computational Complexity*, 20(3):559–578, 2011.

[Hya79] Laurent Hyafil. On the Parallel Evaluation of Multivariate Polynomials.

SIAM Journal of Computing, 8(2):120–123, 1979. Preliminary version in the *10th Annual ACM Symposium on Theory of Computing (STOC 1978)*.

[Jan11] Maurice Jansen. Lower Bounds for the Determinantal Complexity of Explicit Low Degree Polynomials. *Theory of Computing Systems*, 49(2):343–354, 2011.

[Kal85] Kyriakos Kalorkoti. A Lower Bound for the Formula Size of Rational Functions. *SIAM Journal of Computing*, 14(3):678–687, 1985.

[Kay12] Neeraj Kayal. An exponential lower bound for the sum of powers of bounded degree polynomials. In *Electronic Colloquium on Computational Complexity (ECCC)TR12-081*, 2012.

[KLSS14] Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An Exponential Lower Bound for Homogeneous Depth Four Arithmetic Circuits. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2014)*, 2014. Pre-print available at `eccc:TR14-005`.

[KMN13] Mrinal Kumar, Gaurav Maheshwari, and Jayalal Sarma M. N. Arithmetic Circuit Lower Bounds via MaxRank. In *International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 661–672, 2013.

[Koi12] Pascal Koiran. Arithmetic Circuits: The Chasm at Depth Four Gets Wider. *Theoretical Computer Science*, 448:56–65, 2012. Pre-print available at `arXiv:1006.4700`.

[KSS14] Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC 2014)*, pages 146–153, 2014. Pre-print available at `eccc:TR13-091`.

[KST16] Neeraj Kayal, Chandan Saha, and Sébastien Tavenas. An almost Cubic Lower Bound for Depth Three Arithmetic Circuits. Technical re-

port, Electronic Colloquium on Computational Complexity (ECCC), eccc:TR16-006, 2016.

[Lov75] László Lovász. On the ratio of optimal integral and fractional covers. *Discrete mathematics*, 13(4):383–390, 1975.

[Mes89] Roy Meshulam. On two extremal matrix problems. *Linear Algebra and its Applications*, 114:261–271, 1989.

[MR04] Thierry Mignon and Nicolas Ressayre. A quadratic bound for the determinant and permanent problem. *International Mathematics Research Notes*, 2004(79):4241–4253, 2004. Available on citeseer:10.1.1.106.4910.

[MV97] Meena Mahajan and V. Vinay. A Combinatorial Algorithm for the Determinant. In *Proceedings of the 8th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 1997)*, pages 730–738, 1997. Available on citeseer:10.1.1.31.1673.

[Nis91] Noam Nisan. Lower bounds for non-commutative computation. In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing (STOC 1991)*, pages 410–418, 1991. Available on citeseer:10.1.1.17.5067.

[NW94] Noam Nisan and Avi Wigderson. Hardness vs Randomness. *Journal of Computer and System Sciences*, 49(2):149–167, 1994. Available on citeseer:10.1.1.83.8416.

[NW97] Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational Complexity*, 6(3):217–234, 1997. Available on citeseer:10.1.1.90.2644.

[Pam85] P. Pamfilos. On the maximum rank of a tensor product. *Acta Mathematica Hungarica*, 45(1):95–97, 1985.

[Raz10] Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC 2010)*, pages 659–666, 2010. Pre-print available at eccc:TR10-002.

[RS00] Kristian Ranestad and Frank-Olaf Schreyer. Varieties of sums of powers. *JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK*, pages 147–182, 2000.

[Rys63] Herbert J. Ryser. Combinatorial Mathematics. *Mathematical Association of America*, 14, 1963.

[Sap15] Ramprasad Saptharishi. A survey of lower bounds in arithmetic circuit complexity. Github survey, 2015.

[Sax08] Nitin Saxena. Diagonal Circuit Identity Testing and Lower Bounds. In *Proceedings of the 35th International Colloquium on Automata, Languages and Programming (ICALP 2008)*, pages 60–71, 2008. Pre-print available at eccc:TR07-124.

[Sha49] Claude Shannon. The synthesis of two-terminal switching circuits. *Bell Labs Technical Journal*, 28(1):59–98, 1949.

[SW01] Amir Shpilka and Avi Wigderson. Depth-3 arithmetic circuits over fields of characteristic zero. *Computational Complexity*, 10(1):1–27, 2001. Preliminary version in the *14th Annual IEEE Conference on Computational Complexity (CCC 1999)*.

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

[Tav15] Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. *Inf. Comput.*, 240:2–11, 2015. Preliminary version in the *38th Interna-*

tionl Symposium on the Mathematical Foundations of Computer Science (MFCS 2013).

[Tur37] Alan Mathison Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London mathematical society*, 2(1):230–265, 1937.

[Val79] Leslie G. Valiant. Completeness Classes in Algebra. In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing (STOC 1979)*, pages 249–261, 1979.

[Val92] Leslie G Valiant. Why is Boolean complexity theory difficult. *Boolean Function Complexity*, 169:84–94, 1992.

[VSBR83] Leslie G. Valiant, Sven Skyum, S. Berkowitz, and Charles Rackoff. Fast Parallel Computation of Polynomials Using Few Processors. *SIAM Journal of Computing*, 12(4):641–644, 1983. Preliminary version in the *6th Internationl Symposium on the Mathematical Foundations of Computer Science (MFCS 1981)*.

[vzG86] Joachim von zur Gathen. Permanent and Determinant. In *Foundations of Computer Science (FOCS)*, pages 398–401. IEEE Computer Society, 1986.

[vzG87] Joachim von zur Gathen. Permanent and determinant. *Linear Algebra and its Applications*, 96:87–100, 1987.

[Yab15] Akihiro Yabe. Bi-polynomial rank and determinantal complexity. *CoRR*, abs/1504.00151, 2015. Pre-print available at `arXiv:1504.00151`.