Electronic Colloquium on Computational Complexity, Report No. 100 (2013)

Lower bounds for depth 4 formulas computing iterated matrix multiplication *

Hervé Fournier† Nutan Limaye‡ Guillaume Malod§ Srikanth Srinivasan¶

July 15, 2013

Abstract

We study the arithmetic complexity of iterated matrix multiplication. We show that any multilinear homogeneous depth 4 arithmetic formula computing the product of d generic matrices of size $n \times n$, $\mathrm{IMM}_{n,d}$, has size $n^{\Omega(\sqrt{d})}$ as long as $d \leqslant n^{1/10}$. This improves the result of Nisan and Wigderson (Computational Complexity, 1997) for depth 4 set-multilinear formulas.

We also study $\Sigma\Pi^{(O(d/t))}\Sigma\Pi^{(t)}$ formulas, which are depth 4 formulas with the stated bounds on the fan-ins of the Π gates. A recent depth reduction result of Tavenas (MFCS, 2013) shows that any n-variate degree $d = n^{O(1)}$ polynomial computable by a circuit of size poly(n) can also be computed by a depth 4 $\Sigma\Pi^{(O(d/t))}\Sigma\Pi^{(t)}$ formula of top fan-in $n^{O(d/t)}$. We show that any such formula computing $\mathrm{IMM}_{n,d}$ has top fan-in $n^{\Omega(d/t)}$, proving the optimality of Tavenas' result. This also strengthens a result of Kayal, Saha, and Saptharishi (ECCC, 2013) which gives a similar lower bound for an explicit polynomial in VNP.

1 Introduction

Arithmetic circuits are a convenient way to model computation when considering objects of an algebraic nature such as the determinant. Thanks to the work of Valiant [Val79, Val82], they are also the basis of a clean theoretical framework to study the complexity of such objects.

In particular, Valiant defined two classes: the class VP of tractable polynomials and the larger class VNP, which contains polynomials thought to be intractable. He then showed the completeness of the permanent polynomial for the class VNP. This contrasts with the determinant polynomial, whose expression is very close to that of the permanent, but which is efficiently computable. Indeed, a slight restriction of tractable computations [Tod92, MP08] yields a class,

*This research was funded by IFCPAR/CEFIPRA Project No 4702-1(A)

†Univ Paris Diderot, Sorbonne Paris Cité, Institut de Mathématiques de Jussieu, UMR 7586 CNRS, F-75205 Paris, France. fournier@math.univ-paris-diderot.fr

‡Indian Institute of Technology, Bombay, Department of Computer Science and Engineering, Mumbai, India. nutan@cse.iitb.ac.in

§Univ Paris Diderot, Sorbonne Paris Cité, Institut de Mathématiques de Jussieu, UMR 7586 CNRS, F-75205 Paris, France. malod@math.univ-paris-diderot.fr

¶Department of Mathematics, Indian Institute of Technology, Bombay, Mumbai, India. srikanth@math.iitb.ac.in

ISSN 1433-8092

$\mathrm{VP_S}$, for which the determinant is complete. As a result, the major open question of the equality of the classes $\mathrm{VP_S}$ and VNP can be stated as the question of whether a permanent can always be expressed as a "not too big" determinant, without mention of a computation model.

Many other questions remain open in arithmetic circuit complexity. One of them is whether the determinant, or other polynomials from the associated class $\mathrm{VP_S}$, can be computed efficiently by weaker models. Among these models are formulas, which define a class $\mathrm{VP_E}$, where partial results cannot be reused, and constant-depth circuits.

In this paper we will focus on iterated matrix multiplication, another fundamental computation which is complete for the class $\mathrm{VP_S}$, and whether it can be computed by depth 4 formulas, with alternating sum and product gates (so-called $\Sigma\Pi\Sigma\Pi$ formulas).

1.1 Motivation and Results

Interest in depth 4 formulas for arithmetic computation was sparked by the result of Agrawal and Vinay [AV08], showing that, for certain lower bound questions, it was enough to consider this depth 4 case. This was later pursued by Koiran [Koi12] and Tavenas [Tav13], the latter showing that a polynomial of degree d over N variables computed by a circuit of size s can also be computed by a formula of depth 4 and size $\exp(O(\sqrt{d\log(ds)\log N}))$. In particular, every polynomial p of degree $d = \mathrm{poly}(N)$ that has a circuit of size $\mathrm{poly}(N)$ has a depth 4 $\Sigma\Pi\Sigma\Pi$ formula C of size $\exp(O(\sqrt{d}\log N))$. The formula C, additionally, has the property that its Π-gates have fan-in $O(\sqrt{d})$; such formulas are called $\Sigma\Pi^{(O(\sqrt{d}))}\Sigma\Pi^{(O(\sqrt{d}))}$ formulas. In the special case that p is homogeneous, then C is also homogeneous.

There has been recent progress towards proving strong lower bounds for $\Sigma\Pi^{(O(\sqrt{d}))}\Sigma\Pi^{(O(\sqrt{d}))}$ formulas as well: in a breakthrough result, Gupta, Kamath, Kayal and Saptharishi [GKKS13] give $\exp(\Omega(\sqrt{n}))$ lower bounds for $\Sigma\Pi^{(O(\sqrt{n}))}\Sigma\Pi^{(O(\sqrt{n}))}$ formulas computing the $n \times n$ permanent and determinant polynomials. Note that this gives a lower bound of $\exp(\Omega(\sqrt{d}))$, for a polynomial in VP of degree d, which is off by a factor of $\log N$ (here $N = n^2$) in the exponent as compared to the upper bound given by Tavenas. More recently, Kayal, Saha, and Saptharishi [KSS13] give a polynomial of degree d in N variables (for $d = \sqrt{N}$) in VNP such that any $\Sigma\Pi^{(O(\sqrt{d}))}\Sigma\Pi^{(O(\sqrt{d}))}$ formula computing the polynomial has size $\exp(\Omega(\sqrt{d}\log N))$. Thus, improving either the result of Tavenas or the lower bound techniques of [GKKS13, KSS13] a little further could yield the desired separation between VP and VNP.

Here, we take the former approach and consider the question of whether the result of Tavenas [Tav13] can be strengthened. Formally, we ask

> Is it possible to show that any polynomial (respectively homogeneous polynomial) of degree d over N variables that has a $\mathrm{poly}(N)$-sized circuit has a $\Sigma\Pi^{(O(\sqrt{d}))}\Sigma\Pi^{(O(\sqrt{d}))}$ (respectively homogeneous $\Sigma\Pi\Sigma\Pi$) formula of size $\exp(o(\sqrt{d}\log N))$?

We answer this question partially by showing that for all $d \leqslant N^{\varepsilon}$, for some fixed $\varepsilon > 0$, there is an explicit polynomial $f \in \mathrm{VP}$ of degree d on N variables such that any $\Sigma\Pi^{(O(\sqrt{d}))}\Sigma\Pi^{(O(\sqrt{d}))}$ formula computing it has size $\exp(\Omega(\sqrt{d}\log N))$. Thus, in the above regime of parameters, we strengthen the result of [KSS13] by obtaining a similar lower bound for a polynomial in VP.

As a corollary of our technical theorems, we also obtain an optimal lower bound for *regular* *formulas* (see Section 7) for the same polynomial f, answering a question raised in [KSS13].

Moreover, the polynomial f is the Iterated Matrix Multiplication polynomial, which computes a single entry in the product of d many $n \times n$ matrices whose entries are all distinct variables. We will denote this polynomial by $\mathrm{IMM}_{n,d}$. Its complexity is by itself of great interest. It occupies a central position both in algebraic complexity theory — being complete for $\mathrm{VP_S}$ as mentioned above — and in complexity theory in general, since it is closely related to the Boolean and counting versions of the canonical NL-complete problem of deciding reachability in a directed graph. In particular, showing that $\mathrm{IMM}_{n,d}$ does not have polynomial-sized formulas is equivalent to showing a separation between $\mathrm{VP_E}$ and $\mathrm{VP_S}$.

It is easy to see that for any $r \in \mathbb{N}$, the polynomial $\mathrm{IMM}_{n,d}$ has a formula of size $n^{O(rd^{1/r})}$ and product-depth $r \leqslant \log d$ (i.e., at most r Π-gates on any root to leaf path). This formula, constructed using a simple divide-and-conquer technique that requires r levels of recursion, is furthermore a *set-multilinear* formula (see Section 2). In particular, for $r = \log d$, this technique yields a set-multilinear formula of size $n^{O(\log d)}$ for $\mathrm{IMM}_{n,d}$, which is the best known formula upper bound for this polynomial.

In a seminal work, Nisan and Wigderson [NW97] showed lower bounds on the size of product depth-r set multilinear formulas computing $\mathrm{IMM}_{n,d}$. For the case $r = 1$, [NW97] prove an optimal lower bound of n^{d-1}. For $r \geqslant 2$, however, they prove a lower bound of $\exp(\Omega(d^{1/r}))$. Note that the dimension n of the matrices does not feature in the lower bound: indeed, we get the same lower bound for any $n \geqslant 2$. We rectify this situation for $r = 2$ by showing that any set-multilinear $\Sigma\Pi\Sigma\Pi$ formula for $\mathrm{IMM}_{n,d}$ (*with no fan-in restrictions*) must have size at least $n^{\Omega(\sqrt{d})}$. In fact, our lower bound holds in the more general setting of homogeneous multilinear $\Sigma\Pi\Sigma\Pi$ formulas.

1.2 Related work

As mentioned above, the Iterated Matrix Multiplication polynomial has been considered before in a work of Nisan and Wigderson [NW97], which also introduced the important technique of using *partial derivatives* to prove lower bounds in arithmetic complexity. We use a recent strengthening of this technique due to Kayal [Kay12] and Gupta et al. [GKKS13], which uses *shifted partial derivatives*. We briefly survey some results that use this technique, but refer the reader to [GKKS13] for a more thorough account.

Kayal [Kay12] used the shifted partial derivative technique to show a lower bound for expressing the monomial $x_1 x_2 \cdots x_n$ as a sum of powers of bounded degree polynomials in $x_1, \ldots, x_n$. Gupta et al. [GKKS13] showed lower bounds for $\Sigma\Pi\Sigma\Pi$ formulas (with fan-in bounds on the Π-gates) computing the permanent and determinant polynomials. More recently, the shifted partial derivative method has been used by Kumar and Saraf [KS13] to prove lower bounds for homogeneous $\Sigma\Pi\Sigma\Pi$ formulas (see Section 2) with bounded fan-in at the top Σ gate computing the permanent and by Kayal, Saha, and Saptharishi [KSS13] to prove stronger lower bounds for bounded Π-gate fan-in $\Sigma\Pi\Sigma\Pi$ formulas computing a certain explicit polynomial in VNP.

It is interesting to note that the result of [GKKS13] itself implies a lower bound for $\Sigma\Pi^{(O(\sqrt{d})}\Sigma\Pi^{(O(\sqrt{d}))}$ formulas computing the iterated matrix multiplication polynomial. This is because of the well known fact (see for instance [MV97]) that an $m \times m$ determinant is a projection of $\mathrm{IMM}_{n,d}$, where $n = O(m^2)$ and $d = m$. Thus, for this setting of parameters, the lower bound of [GKKS13] for the determinant gives a lower bound of $\exp(\Omega(\sqrt{d}))$ for $\mathrm{IMM}_{n,d}$.

There has also been a considerable amount of research into lower bounds for set-multilinear and more generally, multilinear formulas. Nisan and Wigderson [NW97] proved lower bounds on the size of small-depth set-multilinear formulas for the Iterated Matrix Multiplication polynomial. Building on their techniques, the breakthrough work of Raz [Raz09] proved superpolynomial lower bounds for multilinear formulas computing the determinant and permanent polynomials. Follow-up work of Raz [Raz06] (see also Raz and Yehudayoff [RY08]) showed a superpolynomial separation between $\mathrm{VP_E}$ and VP in the multilinear setting. This was recently strengthened by Dvir, Malod, Perifel, and Yehudayoff [DMPY12] to a superpolynomial separation between $\mathrm{VP_E}$ and $\mathrm{VP_S}$ in the multilinear setting.

A result that is closely related to ours is the work of Raz and Yehudayoff [RY09], who also prove strong exponential lower bounds for constant-depth multilinear formulas. More precisely, they give an explicit multilinear polynomial of degree N over N variables that has no multilinear $\Sigma\Pi\Sigma\Pi$ formulas of size less than $\exp(\Omega(\sqrt{N \log N}))$. Their results are somewhat incomparable to ours, since

- Our lower bound is stronger in that it matches Tavenas' upper bound [Tav13] for $\Sigma\Pi\Sigma\Pi$ formulas for any degree-d polynomial with $\mathrm{poly}(N)$-sized circuits. The above lower bound is slightly weaker.
- The results of Raz and Yehudayoff apply not just to $\Sigma\Pi\Sigma\Pi$ formulas, but to all *small-depth* (up to $o(\log N/\log\log N)$)) multilinear formulas, without homogenity restrictions. (The bounds get weaker with larger depth.)
- As far as we are aware, their techniques — or indeed, any of the general techniques used to prove multilinear formula lower bounds — are not applicable to the Iterated Matrix Multiplication polynomial.

2 Definitions and notations

Let X be a set of variables and let $\mathbb{F}[X]$ denote the set of polynomials over variables X and field $\mathbb{F}$.

2.1 Arithmetic circuits and branching programs

An *arithmetic circuit* is a finite simple directed acyclic graph. The vertices of in-degree 0 are called *input gates* and are labeled by constants from $\mathbb{F}$ or variables from X. The vertices of in-degree at least 2 are labeled by $+$ or $\times$. The *output gate* is a vertex with out-degree 0. The polynomial computed by a node is defined in an obvious inductive way. The polynomial computed by the arithmetic circuit is the polynomial computed by the output gate.

The size of the circuit is the number of nodes in the graph. The depth of the circuit is the length of the longest input gate to output gate path. The in-degree (out-degree) of a node/gate is often called its *fan-in* (*fan-out*, respectively). We do not assume any bound on the fan-ins or fan-outs of the nodes unless stated otherwise. A circuit is called layered if the underlying graph is layered.

An *algebraic branching program*, or ABP, over the set of variables X and field $\mathbb{F}$ is a tuple (G, s, t) where G is a weighted simple directed acyclic graph and s and t are special vertices in G. The *weight* of an edge in a branching program is a linear form in $\mathbb{F}[X]$. The *weight* of a path is the product of the weights of its edges. The polynomial computed by a branching program G is the sum of the weights of all the paths from s to t in G. The *size* of a branching program is the number of its vertices. The *length* of a branching program is the length of the longest s to t path. If we can partition the vertices of a branching program in levels so that there are only edges between vertices in successive levels, we say that the branching program is *layered*.

2.2 Arithmetic formulas and variants

An *arithmetic formula* is an arithmetic circuit which is a simple directed tree. The size, depth, fan-ins, fan-outs and layers for formulas are defined similarly to that of circuits. We fix the convention that in a layered circuit/formula, the layers are numbered in increasing order with input gates getting the smallest number (0) and output gates getting the largest number.

A $\Sigma\Pi\Sigma\Pi$ formula is a layered formula in which gates at layer 1 and 3 are labeled $\times$ and gates at layer 2 and 4 are labeled $+$. We will also use the notation $\Sigma\Pi^{(\alpha)}\Sigma\Pi^{(\beta)}$ to indicate that the fan-in of gates on the first and third layers is bounded by β and α respectively.

Recall that a polynomial is called *homogeneous* if each monomial in it has the same degree. A formula is called homogeneous if each of its gates computes a homogeneous polynomial.

Fix a partition $X_1, X_2, \ldots, X_d$ of X. For a subset $T \subseteq [d]$ we say that a monomial over the variables in X is T-*multilinear* if it is a product of variables such that exactly one variable comes from each X_i $(i \in T)$. A polynomial is called T-multilinear if it is a linear combination of T-multilinear monomials. We say that a polynomial is *set-multilinear* if it is T-multilinear for some $T \subseteq [d]$.

A formula is called set-multilinear if every node in the formula computes a set-multilinear polynomial. Note that a set-multilinear formula is by definition homogeneous.

We also consider *multilinear polynomials*, which are a slight generalization of set-multilinear polynomials. A monomial over a set of variables X is called multilinear if each variable in X has degree at most one in the monomial. A polynomial is called multilinear if it is a linear combination of multilinear monomials. A formula is called multilinear if each node in the formula computes a multilinear polynomial. It is called *homogeneous multilinear* if it is simultaneously homogeneous and multilinear.

For any node g in the formula, let X_g denote the set of variables in the polynomial computed by g. A formula is called *syntactic multilinear* if, for each $\times$ node g in the formula, the sets $X_{g_1}, X_{g_2}, \ldots, X_{g_k}$ are mutually disjoint, where $g_1, g_2, \ldots, g_k$ are the children of g.

It is known from [SY10] that if there is a multilinear formula F of size s computing a multilinear polynomial $p \in \mathbb{F}[X]$, then there exists a syntactic multilinear formula of size at most s computing p; similar statements are also true for $\Sigma\Pi\Sigma\Pi$ and homogeneous $\Sigma\Pi\Sigma\Pi$ formulas. Therefore, we assume without loss of generality that the formulas computing multilinear polynomials are syntactic multilinear.

It will be convenient for us to blur the distinction between multilinear monomials over the set of variables X and subsets of X. Thus, we freely apply reasonable set-theoretic operations to multilinear monomials. For example, for multilinear monomials m_1 and m_2, $m_1 \cup m_2$ is the *multilinear* monomial that contains exactly the variables that occur in either m_1 or m_2; we can similarly define $m_1 \cap m_2$ and $m_1 \backslash m_2$; $|m|$ will denote the degree of a multilinear monomial m.

2.3 The Iterated Matrix Multiplication polynomial

Throughout, let $n, d \geqslant 2$ be fixed parameters.

We consider polynomials defined on variable sets $X_1, \ldots, X_d$. For $i \in [d] \backslash \{1, d\}$, let X_i be the set of variables $x_{j,k}^{(i)}$ for $j, k \in [n]$; for $i \in \{1, d\}$, let X_i be the set of variables $x_j^{(i)}$ for $j \in [n]$. Let $X = \bigcup_{i \in [d]} X_i$. We will use N to denote $|X| = (d-2)n^2 + 2n$.

The Iterated Matrix Multiplication polynomial on X, denoted $\mathrm{IMM}_{n,d}$, is defined to be

$$\mathrm{IMM}_{n,d} = \sum_{j_1, \cdots, j_{d-1}} x_{j_1}^{(1)} x_{j_1,j_2}^{(2)} x_{j_2,j_3}^{(3)} \cdots x_{j_{d-2},j_{d-1}}^{(d-1)} x_{j_{d-1}}^{(d)}.$$

Note that the polynomial $\mathrm{IMM}_{n,d}$ is the sum of the entries of the product of d generic matrices (of dimensions $1 \times n$, $n \times n$ ($d-2$ times), and $n \times 1$), the ith matrix having entries from the variable set X_i. Hence in the remainder of the paper we refer to "the matrix X_i".

Another way to define this polynomial is to see it as a generic layered algebraic branching program with $d+1$ layers $V_0, \ldots, V_d$ where $V_i = \{v_1^{(i)}, \ldots, v_n^{(i)}\}$ for $0 < i < d$ and $V_i = \{v^{(i)}\}$ for $i \in \{0, d\}$. The graph contains all possible edges from V_i to V_{i+1} for $i \in \{0, \ldots, d-1\}$. The edge from $v_j^{(i-1)}$ to $v_k^{(i)}$ is labeled with the variable $x_{j,k}^{(i)}$ for $0 < i < d-1$ and the edges from $v^{(0)}$ to $v_j^{(1)}$ and $v_j^{(d-1)}$ to $v^{(d)}$ are labeled $x_j^{(1)}$ and $x_j^{(d)}$ respectively. Then, $\mathrm{IMM}_{n,d}$ is the polynomial computed by the branching program, i.e., the sum of the weights of all the paths from the vertex $v^{(0)}$ to the vertex $v^{(d)}$.

We denote by $\mathcal{A}$ the canonical ABP defined above. Given a path ρ in the ABP $\mathcal{A}$, we will also denote by ρ the product of all the variables that occur along the edges in the path ρ.

2.4 The dimension of the shifted partial derivatives

As in [Kay12, GKKS13], we will use the dimension of shifted partial derivatives as our complexity measure.

For $k, \ell \in \mathbb{N}$ and a multivariate polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$, we define

$$\langle \partial_k f \rangle_{\leqslant \ell} = \operatorname{span} \left\{ x_1^{j_1} \ldots x_n^{j_n} \cdot \frac{\partial^k f}{\partial x_1^{i_1} \ldots \partial x_n^{i_n}} \;\middle|\; i_1 + \ldots + i_n = k,\ j_1 + \ldots + j_n \leqslant \ell \right\}.$$

The complexity measure we use is $\dim(\langle \partial_k f \rangle_{\leqslant \ell})$.

3 Preliminaries

In this section we give a few technical lemmas and definitions which will be used in the subsequent sections.

3.1 The derivatives of $\mathrm{IMM}_{n,d}$

The derivatives of $\mathrm{IMM}_{n,d}$ have a simple form that is easily described. Since we will be interested in lower bounding the size of the partial derivative space of this polynomial, we only choose a subset of all partial derivatives available to us. Let k denote a parameter which we will choose later. Let r denote $\left\lfloor \frac{d}{k+1} \right\rfloor - 1$. We fix k matrices among $X_1, \ldots, X_d$ that are placed evenly apart. Formally, choose k matrices $X_{p_1}, \ldots, X_{p_k}$ such that $p_q - (p_{q-1} + 1) \geqslant r$ for all $1 \leqslant q \leqslant k+1$, where $p_0 = 0$ and $p_{k+1} = d+1$. We then choose one variable from each of these chosen matrices, say $x^{(p_1)}_{i_1,j_1}, \ldots, x^{(p_k)}_{i_k,j_k}$ and take derivatives with respect to these variables. We denote this derivative by $\partial_{\mathcal{I}} \mathrm{IMM}_{n,d}$, where $\mathcal{I}$ denotes $(i_1, j_1, \ldots, i_k, j_k) \in [n]^{2k}$.

Note that $\partial_{\mathcal{I}} \mathrm{IMM}_{n,d}$ can be written as a sum of monomials m such that $m = \rho_1 \rho_2 \ldots \rho_{k+1}$, where ρ_q is a path from $v^{(p_{q-1})}_{j_{q-1}}$ to $v^{(p_q - 1)}_{i_q}$ in $\mathcal{A}$ for all $2 \leqslant q \leqslant k$, ρ_1 is a path from vertex $v^{(0)}$ to $v^{(p_1 - 1)}_{i_1}$ in $\mathcal{A}$, and ρ_{k+1} is a path from $v^{(p_k)}_{j_k}$ to vertex $v^{(d)}$ in $\mathcal{A}$. Clearly, $\partial_{\mathcal{I}} \mathrm{IMM}_{n,d}$ is a homogeneous polynomial of degree $d - k$.

3.2 Restrictions

Definition 1. *By a* restriction *of the variable set X, we will mean a function $\sigma : X \to \{0, *\}$. Given $f \in \mathbb{F}[X]$ and a restriction σ on X, we denote by $f|_\sigma$ the polynomial $g \in \mathbb{F}[X]$ obtained by setting all the variables $x \in \sigma^{-1}(0)$ to 0 (the other variables remain as they are).*

Given polynomials $f, g \in \mathbb{F}[X]$, we say that g is a restriction of f *if there exists a restriction σ on X such that $g = f|_\sigma$.*

Given a formula C over the variables X and a restriction σ, we define $C|_\sigma$ to be the circuit obtained by replacing all the variables $x \in \sigma^{-1}(0)$ with 0 then simplifying the formula accordingly, by suppressing any Π gate receiving a variable set to 0. Clearly, if C computes the polynomial $f \in \mathbb{F}[X]$, then $C|_\sigma$ computes the polynomial $f|_\sigma$.

We will mostly be interested in restrictions of $\mathrm{IMM}_{n,d}$. In this setting, the following basic observation helps simplify many arguments.

Remark 2. *In Section 2.3, we defined the $\mathrm{IMM}_{n,d}$ polynomial to be the polynomial computed by an ABP $\mathcal{A}$ such that for each edge of $\mathcal{A}$, the linear form labeling that edge was a distinct variable from X. Restrictions F of $\mathrm{IMM}_{n,d}$ are polynomials obtained when we set certain variables of X to 0 in $\mathrm{IMM}_{n,d}$; equivalently, we may see F as the polynomial computed by the ABP $\mathcal{A}_F$ obtained when we delete the edges corresponding to the variables that are set to 0 by the restriction.*

3.3 The shifted partial derivative space of $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas

We need an upper bound on the dimension of the shifted partial derivative space of polynomials computed by small $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas. The following is implicit in the work of Gupta et al. [GKKS13] and is stated explicitly in [KSS13].

Lemma 3 ([KSS13], Lemma 4)**.** *Let $D, t, k, \ell \in \mathbb{N}$ be arbitrary parameters. Let C be a $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formula with at most s Π gates at layer 3 computing a polynomial in N variables. Then, we have*

$$\dim(\langle \partial_k C \rangle_{\leqslant \ell}) \leqslant s \cdot \binom{D}{k} \cdot \binom{N + \ell + (t-1)k}{\ell + (t-1)k}.$$

3.4 Technical lemmas

Fact 4. *For any integers N, ℓ, r such that $r < \ell$, we have*

$$\left(\frac{N+\ell}{\ell}\right)^r \leqslant \frac{\binom{N+\ell}{\ell}}{\binom{N+\ell-r}{\ell-r}} \leqslant \left(\frac{N+\ell-r}{\ell-r}\right)^r.$$

Claim 5. *For any integers $n, d \geqslant 2$, $N = (d-2)n^2 + 2dn$ and $t \geqslant 1$, there exists an integer $\ell > d$ such that $n^{1/16} \leqslant \left(\frac{N+\ell}{\ell}\right)^t \leqslant n^{1/4}$.*

Proof. We choose ℓ to be the least positive integer such that $f(\ell) := \left(\frac{N+\ell}{\ell}\right)^t \leqslant n^{1/4}$. Note that such an ℓ exists since $f(1) = (N+1)^t > n^{1/4}$ and $\lim_{\ell \to \infty} f(\ell) = 1$. We must also have $\ell > d$ since for $\ell \leqslant d$, we have $f(\ell) \geqslant ((N/d) + 1)^t \geqslant (n+1)^t > n^{1/4}$.

The only thing left to show is that for this choice of ℓ, we have $\left(\frac{N+\ell}{\ell}\right)^t \geqslant n^{1/16}$. To prove this, we claim that it suffices to prove the following inequality for any $\ell' \geqslant 1$

$$\sqrt{f(\ell')} \leqslant f(\ell' + 1). \tag{1}$$

To see this, note that assuming the above inequality, we have $f(\ell) \geqslant \sqrt{f(\ell - 1)} \geqslant n^{1/16}$, where the last inequality follows from the fact that $f(\ell - 1) \geqslant n^{1/4}$.

The proof of Inequality (1) is elementary. We need to show that

$$\begin{aligned} \left(\frac{N+\ell'}{\ell'}\right)^{t/2} &\leqslant \left(\frac{N+\ell'+1}{\ell'+1}\right)^t \\ \Leftrightarrow \quad \left(\frac{N+\ell'}{\ell'}\right)^{1/2} &\leqslant \frac{N+\ell'+1}{\ell'+1}. \end{aligned}$$

Squaring both sides and cross multiplying we see that (1) is equivalent to

$$\begin{aligned} \frac{(\ell'+1)^2}{\ell'} &\leqslant \frac{(N+\ell'+1)^2}{N+\ell'} \\ \Leftrightarrow \quad \ell' + \frac{1}{\ell'} + 2 &\leqslant N + \ell' + \frac{1}{N+\ell'} + 2 \end{aligned}$$

which is easily verified for $N \geqslant 1$.

□

4 Proof overview

In this section, we briefly describe the outline of the proof of the main theorems.

Our theorems prove strong lower bounds on variants of $\Sigma\Pi\Sigma\Pi$ formulas computing $\mathrm{IMM}_{n,d}$. Recall that we already have tight lower bounds on $\Sigma\Pi\Sigma$ set-multilinear formulas computing $\mathrm{IMM}_{n,d}$ due to Nisan and Wigderson [NW97]. A natural first step for us, therefore, would be to prove an optimal lower bound for set-multilinear formulas which are sums of products of quadratics, i.e., set-multilinear $\Sigma\Pi\Sigma\Pi^{(2)}$, or more generally, sums of products of low degree polynomials. To do this, we use the *shifted partial derivative* method of Gupta et al. [GKKS13], who introduced this technique to prove that any $\Sigma\Pi^{(O(n/t))}\Sigma\Pi^{(t)}$ formula (not necessarily homogeneous) computing the permanent or the determinant polynomial (on n^2 variables) must have size $\exp(\Omega(n/t))$. Their proof was made up of two steps.

- First, they observed that the shifted partial derivative space of $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas, for suitable D and t, has *small* dimension.
- Then, they showed that the dimension of $\langle\partial_k F\rangle_{\leqslant\ell}$ is quite large for suitable k and ℓ, where F is any one of the determinant or permanent polynomials.

We prove a strong lower bound on dimension of the shifted partial derivative space of $\mathrm{IMM}_{n,d}$, thereby proving a lower bound of $n^{\Omega(d/t)}$ for $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas computing $\mathrm{IMM}_{n,d}$, as long as D is small enough compared to n. In fact, we manage to prove something slightly stronger. We prove that some carefully chosen *restrictions* (see Section 3.2) of the $\mathrm{IMM}_{n,d}$ polynomial have shifted partial derivative spaces of large dimension. Putting this together with Lemma 3 implies strong lower bounds for $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas computing even these restrictions of $\mathrm{IMM}_{n,d}$.

In order to prove our next result, a lower bound for set-multilinear $\Sigma\Pi\Sigma\Pi$ formulas and homogeneous multilinear $\Sigma\Pi\Sigma\Pi$ formulas of possibly unbounded bottom fan-in computing $\mathrm{IMM}_{n,d}$, we reduce to the case of formulas with bounded bottom fan-in using the idea of *random restrictions*. This is motivated by, and reminiscent of some arguments in [FSS84, Hås87, NW97]; our restrictions themselves, however, look quite different.

We force the fan-in of the bottom Π gates to less than some threshold t by using random restrictions. This is quite intuitive, since a random restriction that sets any variable to 0 with good probability should set any high degree (multilinear) monomial to 0 with probability close to 1. Importantly for us, though, we can devise such a set of restrictions with the additional property that these restrictions remain hard to compute for homogeneous $\Sigma\Pi\Sigma\Pi^{(t)}$ formulas, by the ideas used to prove the lower bound for $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas.

We consider two different sets of restrictions. The first set of restrictions is simpler, but only works to reduce the fan-in of *set-multilinear* $\Sigma\Pi\Sigma\Pi$ formulas. By considering a second, slightly more involved, family of restrictions, we prove a lower bound for homogeneous multilinear $\Sigma\Pi\Sigma\Pi$ formulas as well. Note that the lower bound result for $\Sigma\Pi\Sigma\Pi$ homogeneous multilinear formulas subsumes the lower bound result for set-multilinear $\Sigma\Pi\Sigma\Pi$ formulas and indeed, there is a good amount of overlap between the two proofs. However, for the sake of clarity of exposition, we give detailed proofs for both.

5 Lower bounds for set-multilinear formulas

We start by defining a set of restrictions of $\mathrm{IMM}_{n,d}$, then we prove a lower bound for $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas computing them, and finally we show that there exists a restriction in the set which changes a set-multilinear formula to a $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formula.

5.1 Nice restrictions of $\mathrm{IMM}_{n,d}$

Our restrictions are related to the evenly-spaced matrices chosen before: recall that we have chosen indices $p_1, \ldots, p_k$ and also set $p_0 = 0$ and $p_{k+1} = d+1$. We will now choose new indices p'_q, where p'_q is roughly in the middle between p_{q-1} and p_q: for each $q \in [k+1]$, we choose a p'_q such that $p_{q-1} \leqslant p'_q \leqslant p_q$ and $\min\{p'_q - (p_{q-1}+1), p_q - (p'_q+1)\} \geqslant \left\lfloor \frac{r-1}{2} \right\rfloor$. We define P' to be $\{p_q \mid q \in [k]\} \cup \left\{p'_q \,\middle|\, q \in [k+1]\right\}$.

We then consider the set $\mathcal{R}$ of restrictions which:

- keep only one variable in the first (row) matrix,
- for each index $p \notin P' \cup \{1, d\}$, keep only the variables in X_p of the form $x_{i,\pi_p(i)}$ for some permutation π_p of $[n]$.
- for each index $p \in P'$, leave the variables of X_p untouched,
- keep only one variable in the last (column) matrix.

More formally, let $\mathcal{R}$ be the set of restrictions τ defined below, for any choice of integers j_1 and j_d in $[n]$ and any set $\{\pi_p \mid p \notin P' \cup \{1, d\}\}$ of permutations of $[n]$:

$$\tau(x) = \begin{cases} 0 & \text{if } x = x_j^{(1)} \text{ for some } j \neq j_1, \text{ or} \\ & \text{if } x = x_j^{(d)} \text{ for some } j \neq j_d, \text{ or} \\ & \text{if } x = x_{i,j}^{(p)} \text{ for } j \neq \pi_p(i) \text{ and } p \notin P' \cup \{1, d\}, \\ * & \text{otherwise.} \end{cases}$$

For example, we can choose to keep the first variable of X_1 and X_d and in each matrix X_p for $p \in P'$ to keep only variables on the diagonal, thus defining a restriction σ:

$$\sigma(x) = \begin{cases} 0 & \text{if } x = x_j^{(1)} \text{ for some } j \neq 1, \text{ or} \\ & \text{if } x = x_j^{(d)} \text{ for some } j \neq 1, \text{ or} \\ & \text{if } x = x_{i,j}^{(p)} \text{ for } i \neq j \text{ and } p \notin P' \cup \{1, d\}, \\ * & \text{otherwise.} \end{cases}$$

Let F be the polynomial $\mathrm{IMM}_{n,d}|_\sigma$. As we saw in Remark 2, we can also define the polynomial F in the language of ABPs. Consider the ABP $\mathcal{A}$ defined in Section 2.3 above. Construct a new ABP $\mathcal{A}'$ by removing edges from $\mathcal{A}$ as follows:

- Remove all edges from $v^{(0)}$ to $v_j^{(1)}$ for $j \neq 1$.

- For $p \notin P' \cup \{1, d\}$, remove all edges between V_{p-1} to V_p except for those of the form $(v_j^{(p-1)}, v_j^{(p)})$ for $j \in [n]$.
- Remove all edges from $v_j^{(d-1)}$ to $v^{(d)}$ for $j \neq 1$.

The ABP $\mathcal{A}'$ computes exactly the polynomial F.

5.2 A lower bound for nice restrictions of $\mathrm{IMM}_{n,d}$

We will work with F for most of the lower bound proof, and then show that the lower bound holds for all the polynomials obtained from $\mathrm{IMM}_{n,d}$ by a restriction in $\mathcal{R}$.

5.2.1 The dimension of the space of shifted partial derivatives of F

As with $\mathrm{IMM}_{n,d}$ in Section 3.1, we will consider the derivatives of F with respect to a tuple of variables $x_{i_1,j_1}^{(p_1)}, \ldots, x_{i_k,j_k}^{(p_k)}$ and denote this derivative by $\partial_{\mathcal{I}} F$, where $\mathcal{I}$ denotes $(i_1, j_1, \cdots, i_k, j_k)$ as before. It can be observed from the restriction defining F that $\partial_{\mathcal{I}} F$ is now a *single monomial* of degree $d - k$, which we denote $p(\mathcal{I})$. In fact, we can write $p(\mathcal{I}) = \rho_1 \rho_2 \ldots \rho_{k+1}$, where

$$\rho_1 = \underbrace{\left(x_1^{(1)} \cdot \prod_{1<p<p_1'} x_{1,1}^{(p)} \right)}_{g_1^{\mathcal{I}}} \cdot x_{1,i_1}^{(p_1')} \cdot \underbrace{\left(\prod_{p_1'<p<p_1} x_{i_1,i_1}^{(p)} \right)}_{h_1^{\mathcal{I}}}$$

$$\rho_q = \underbrace{\left(\prod_{p_{q-1}<p<p_q'} x_{j_{q-1},j_{q-1}}^{(p)} \right)}_{g_q^{\mathcal{I}}} \cdot x_{j_{q-1},i_q}^{(p_q')} \cdot \underbrace{\left(\prod_{p_q'<p<p_q} x_{i_q,i_q}^{(p)} \right)}_{h_q^{\mathcal{I}}} \qquad \text{(for } 1 < q < k+1\text{)}$$

$$\rho_{k+1} = \underbrace{\left(\prod_{p_k<p<p_{k+1}'} x_{j_k,j_k}^{(p)} \right)}_{g_{k+1}^{\mathcal{I}}} \cdot x_{j_k,1}^{(p_{k+1}')} \cdot \underbrace{\left(\left(\prod_{p_{k+1}'<p<d} x_{1,1}^{(p)} \right) \cdot x_1^{(d)} \right)}_{h_{k+1}^{\mathcal{I}}}$$

We would like to lower bound the dimension of the vector space generated by the shifted k-partial derivatives of F. Clearly, we have

$$\dim(\langle \partial_k F \rangle_{\leqslant \ell}) \geqslant \dim\left(\mathrm{span}(\mathcal{M})\right)$$

where $\mathcal{M} = \left\{ m \cdot \partial_{\mathcal{I}} F \;\middle|\; m \text{ a monomial of degree at most } \ell \text{ and } \mathcal{I} \in [n]^{2k} \right\}$. Since $\mathcal{M}$ is a set of *monomials*, the dimension of the span of $\mathcal{M}$ is exactly $|\mathcal{M}|$.

Another way of looking at $\mathcal{M}$ is as follows. For $\mathcal{I} \in [n]^{2k}$, define

$$\mathcal{M}_{\mathcal{I}} := \left\{ m' \;\middle|\; m' \text{ a monomial of degree at most } \ell + d - k \text{ and } p(\mathcal{I}) \text{ divides } m' \right\}.$$

Since by definition, $p(\mathcal{I})$ is exactly $\partial_{\mathcal{I}} F$, we have

$$\begin{aligned}\mathcal{M} &= \left\{ m \cdot p(\mathcal{I}) \,\middle|\, m \text{ a monomial of degree at most } \ell \text{ and } \mathcal{I} \in [n]^{2k} \right\} \\ &= \left\{ m' \,\middle|\, m' \text{ of degree at most } \ell + d - k \text{ and } \exists\, \mathcal{I} \in [n]^{2k} \text{ such that } p(\mathcal{I}) | m' \right\} = \bigcup_{\mathcal{I} \in [n]^{2k}} \mathcal{M}_{\mathcal{I}}\end{aligned}$$

We have shown the following.

Claim 6. *For F and $\mathcal{M}_{\mathcal{I}}$ ($\mathcal{I} \in [n]^{2k}$) as defined above, we have $\dim(\langle \partial_k F \rangle_{\leqslant \ell}) \geqslant |\mathcal{M}|$, where $\mathcal{M} = \bigcup_{\mathcal{I} \in [n]^{2k}} \mathcal{M}_{\mathcal{I}}$.*

We will also need the following simple technical claim, the intuitive content of which is that any two distinct monomials $p(\mathcal{I})$ and $p(\mathcal{I}')$ are quite different. Recall that we do not distinguish between multilinear monomials over the variable set X and subsets of X.

Claim 7. *For any $\mathcal{I}, \mathcal{I}' \in [n]^{2k}$, we have*

$$|p(\mathcal{I}') \backslash p(\mathcal{I})| \geqslant \Delta(\mathcal{I}, \mathcal{I}') \cdot \left\lfloor \frac{r-1}{2} \right\rfloor$$

where $\Delta(\mathcal{I}, \mathcal{I}')$ denotes the Hamming distance between $\mathcal{I}$ and $\mathcal{I}'$.

Proof. Consider any $\mathcal{I}, \mathcal{I}' \in [n]^{2k}$. Say $\mathcal{I} = (i_1, j_1, \ldots, i_k, j_k)$ and $\mathcal{I}' = (i'_1, j'_1, \ldots, i'_k, j'_k)$. Then, using the notation from the definition of $p(\mathcal{I})$, we have

$$\begin{aligned}p(\mathcal{I}') \backslash p(\mathcal{I}) &\supseteq \dot{\bigcup_{q \in [k]}} (g^{\mathcal{I}'}_{q+1} \backslash g^{\mathcal{I}}_{q+1}) \,\dot\cup\, \dot{\bigcup_{q \in [k]}} (h^{\mathcal{I}'}_{q} \backslash h^{\mathcal{I}}_{q}) \\ &\supseteq \dot{\bigcup_{q \in [k] : j_q \neq j'_q}} (g^{\mathcal{I}'}_{q+1} \backslash g^{\mathcal{I}}_{q+1}) \,\dot\cup\, \dot{\bigcup_{q \in [k] : i_q \neq i'_q}} (h^{\mathcal{I}'}_{q} \backslash h^{\mathcal{I}}_{q}).\end{aligned}$$

(Recall that $A \dot\cup B$ denotes the union of *disjoint* sets A and B.)

Note that when $j_q \neq j'_q$, then the monomials $g^{\mathcal{I}}_{q+1}$ and $g^{\mathcal{I}'}_{q+1}$ do not share *any* variables and hence $|g^{\mathcal{I}'}_{q+1} \backslash g^{\mathcal{I}}_{q+1}| = |g^{\mathcal{I}'}_{q+1}| \geqslant \lfloor \frac{r-1}{2} \rfloor$. Similarly, when $i_q \neq i'_q$, we have $|h^{\mathcal{I}'}_{q} \backslash h^{\mathcal{I}}_{q}| \geqslant \lfloor \frac{r-1}{2} \rfloor$.

$$\begin{aligned}|p(\mathcal{I}') \backslash p(\mathcal{I})| &\geqslant \sum_{q \in [k] : j_q \neq j'_q} |g^{\mathcal{I}'}_{q+1} \backslash g^{\mathcal{I}}_{q+1}| + \sum_{q \in [k] : i_q \neq i'_q} |h^{\mathcal{I}'}_{q} \backslash h^{\mathcal{I}}_{q}| \\ &\geqslant \Delta(\mathcal{I}, \mathcal{I}') \cdot \left\lfloor \frac{r-1}{2} \right\rfloor,\end{aligned}$$

which completes the proof of the claim. □

Claim 8. *For any $\mathcal{I} \in [n]^{2k}$, we have $\mathcal{M}_{\mathcal{I}} = \binom{N+\ell}{\ell}$.*

Proof. A monomial $m \in \mathcal{M}_{\mathcal{I}}$ iff there is a monomial m' of degree at most ℓ such that $m = m' \cdot p(\mathcal{I})$. Thus, $|\mathcal{M}_{\mathcal{I}}|$ is equal to the number of monomials of degree at most ℓ, which is $\binom{N+\ell}{\ell}$. □

Claim 9. *For any $\mathcal{I}, \mathcal{I}' \in [n]^{2k}$, we have $|\mathcal{M}_{\mathcal{I}} \cap \mathcal{M}_{\mathcal{I}'}| = \binom{N+\ell-|p(\mathcal{I}')\backslash p(\mathcal{I})|}{\ell-|p(\mathcal{I}')\backslash p(\mathcal{I})|}$.*

Proof. Fix any $\mathcal{I}, \mathcal{I}'$ as above. Any monomial $m \in \mathcal{M}_{\mathcal{I}} \cap \mathcal{M}_{\mathcal{I}'}$ may be factored as $m = m' \cdot p(\mathcal{I}) \cdot (p(\mathcal{I}')\backslash p(\mathcal{I}))$. Note that the degree of m' can be bounded by $\ell + d - k - (d-k) - |p(\mathcal{I}')\backslash p(\mathcal{I})| = \ell - |p(\mathcal{I}')\backslash p(\mathcal{I})|$.

Thus, $|\mathcal{M}_{\mathcal{I}} \cap \mathcal{M}_{\mathcal{I}'}|$ is equal to the number of monomials of degree at most $\ell - |p(\mathcal{I}')\backslash p(\mathcal{I})|$, from which the claim follows. ☐

Claim 10. *Fix any $k, n \in \mathbb{N}$. Then there exists an $\mathcal{S} \subseteq [n]^{2k}$ such that*

- $|\mathcal{S}| = \left\lfloor \left(\frac{n}{4}\right)^k \right\rfloor$,
- *For all distinct $\mathcal{I}, \mathcal{I}' \in \mathcal{S}$, we have $\Delta(\mathcal{I}, \mathcal{I}') \geqslant k$.*

Proof. Greedily pick vectors which have pairwise Hamming distance at least k. A standard volume argument (see, e.g., [Gur10]) shows that the set picked has size at least $\frac{n^{2k}}{\mathrm{Vol}_n(2k,k)}$, where $\mathrm{Vol}_n(2k, k)$ stands for the volume of the Hamming ball of radius k for strings of length $2k$ over an alphabet of size n. We can upper bound $\mathrm{Vol}_n(2k, k)$ by $n^k\binom{2k}{k}$. This shows that there exists a set $\mathcal{S}$ of size at least $n^k/\binom{2k}{k} \geqslant (n/4)^k$. We choose $\mathcal{S}$ such that it has size exactly $\left\lfloor \left(\frac{n}{4}\right)^k \right\rfloor$. Hence the lemma follows.

☐

Now we are ready to prove lower bound on the dimension of the space of shifted partial derivatives of F.

Lemma 11. *Let $k, \ell \in \mathbb{N}$ be arbitrary parameters such that $20k < d < \ell$ and $k \geqslant 2$. Then,*

$$\dim(\langle \partial_k F \rangle_{\leqslant \ell}) \geqslant M \cdot \binom{N+\ell}{\ell} - M^2 \cdot \binom{N+\ell-d/10}{\ell-d/10},$$

where $M = \left\lfloor \left(\frac{n}{4}\right)^k \right\rfloor$.

Proof. Fix $\mathcal{S}$ as guaranteed by Claim 10. By Claim 6, it suffices to lower bound $|\mathcal{M}|$. For this, we use inclusion-exclusion. Since $\mathcal{M} = \bigcup_{\mathcal{I}} \mathcal{M}_{\mathcal{I}}$, we have

$$\begin{aligned} |\mathcal{M}| &\geqslant |\bigcup_{\mathcal{I} \in \mathcal{S}} \mathcal{M}_{\mathcal{I}}| \\ &\geqslant \sum_{\mathcal{I} \in \mathcal{S}} |\mathcal{M}_{\mathcal{I}}| - \sum_{\mathcal{I} \neq \mathcal{I}' \in \mathcal{S}} |\mathcal{M}_{\mathcal{I}} \cap \mathcal{M}_{\mathcal{I}'}|. \end{aligned} \tag{2}$$

By Claim 8, we know that $|\mathcal{M}_{\mathcal{I}}| = \binom{N+\ell}{\ell}$. By Claims 9 and 7 and our choice of $\mathcal{S}$, we see that for any distinct $\mathcal{I}, \mathcal{I}' \in \mathcal{S}$, we have

$$|\mathcal{M}_{\mathcal{I}} \cap \mathcal{M}_{\mathcal{I}'}| \leqslant \binom{N+\ell-k\cdot\lfloor(r-1)/2\rfloor}{\ell-k\cdot\lfloor(r-1)/2\rfloor} \leqslant \binom{N+\ell-d/10}{\ell-d/10}$$

where the last inequality follows since $\lfloor (r-1)/2 \rfloor \geqslant d/10k$ for $k \leqslant d/20$.

Plugging the above into (2), we obtain

$$|\mathcal{M}| \geqslant |\mathcal{S}| \cdot \binom{N+\ell}{\ell} - |\mathcal{S}|^2 \cdot \binom{N+\ell-d/10}{\ell-d/10}.$$

Since $|\mathcal{S}| = \left\lfloor \left(\frac{n}{4}\right)^k \right\rfloor$, the lemma follows. ☐

5.2.2 A lower bound for $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas computing F

We now prove the main lemma for $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas.

Lemma 12. *Let $n, d, D, t, k \in \mathbb{N}$ be such that $1 \leqslant t \leqslant d/320$, $n \geqslant 10$, and $k \leqslant d/320t$. Then, any $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formula for F has top fan-in at least $\Omega\left(\left(\frac{n^{3/4}}{4D}\right)^k\right)$.*

Proof. Recall that $N = (d-2)n^2 + 2dn = |X|$. By Claim 5, we can choose ℓ to be a positive integer such that $n^{1/16} \leqslant \frac{N+\ell}{\ell} \leqslant n^{1/4}$. We now analyze $\dim(\langle \partial_k F \rangle_{\leqslant \ell})$. By Lemma 11, we have

$$\dim(\langle \partial_k F \rangle_{\leqslant \ell}) \geqslant \underbrace{M \cdot \binom{N+\ell}{\ell}}_{T_1} - \underbrace{M^2 \cdot \binom{N+\ell-d/10}{\ell-d/10}}_{T_2}$$

where $M = \lfloor (\frac{n}{4})^k \rfloor$.

However, for our choice of parameters, we have

$$\begin{aligned}
\frac{T_1}{T_2} &= \frac{\binom{N+\ell}{\ell}}{M \cdot \binom{N+\ell-d/10}{\ell-d/10}} \\
&\geqslant \frac{1}{M} \cdot \left(\frac{N+\ell}{\ell}\right)^{d/10} && \text{(by Fact 4)} \\
&\geqslant \frac{n^{d/320t}}{M} && \text{(by our choice of } \ell\text{)} \\
&\geqslant \frac{n^k}{M} \geqslant 4^k \geqslant 2.
\end{aligned}$$

Hence, we have

$$\dim(\langle \partial_k F \rangle_{\leqslant \ell}) \geqslant T_1 - T_2 \geqslant T_1/2 = \frac{M}{2} \cdot \binom{N+\ell}{\ell}. \tag{3}$$

Now, let C be a $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formula for F of top fan-in s. Then, by Lemma 3, we have

$$\dim(\langle \partial_k C \rangle_{\leqslant \ell}) \leqslant s \cdot \binom{D}{k} \cdot \binom{N+\ell+(t-1)k}{\ell+(t-1)k} \leqslant s \cdot D^k \cdot \binom{N+\ell+(t-1)k}{\ell+(t-1)k}.$$

An application of inequality (3) implies that we must have

$$s \cdot D^k \cdot \binom{N+\ell+(t-1)k}{\ell+(t-1)k} \geqslant \frac{M}{2} \cdot \binom{N+\ell}{\ell}.$$

Therefore,

$$
\begin{aligned}
s &\geqslant \frac{M}{2D^k} \cdot \frac{\binom{N+\ell}{\ell}}{\binom{N+\ell+(t-1)k}{\ell+(t-1)k}} \\
&\geqslant \frac{M}{2D^k} \cdot \left(\frac{\ell}{N+\ell}\right)^{(t-1)k} && \text{(by Fact 4)} \\
&\geqslant \frac{1}{4D^k} \cdot \left(\frac{n}{4}\right)^k \cdot \left(\frac{\ell}{N+\ell}\right)^{(t-1)k} && \text{(by our choice of } M\text{)} \\
&= \frac{1}{4} \cdot \left(\frac{n}{4D} \cdot \left(\frac{\ell}{N+\ell}\right)^{(t-1)}\right)^k \\
&\geqslant \frac{1}{4} \cdot \left(\frac{n}{4D \cdot \left(\frac{N+\ell}{\ell}\right)^t}\right)^k \geqslant \frac{1}{4} \cdot \left(\frac{n^{3/4}}{4D}\right)^k && \text{(by our choice of } \ell\text{).}
\end{aligned}
$$

This proves the theorem. □

5.2.3 A lower bound for $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas computing nice restrictions of $\mathrm{IMM}_{n,d}$

We will show that any restriction in $\mathcal{R}$, when applied to $\mathrm{IMM}_{n,d}$, yields a polynomial whose complexity is equivalent to that of F.

Definition 13. *For a polynomial $g \in \mathbb{F}[X]$ and a permutation ϕ of X, define $\phi(g)$ as the polynomial obtained by replacing in g each variable $x \in X$ by $\phi(x)$.*
Two polynomials $f, g \in \mathbb{F}[X]$ are said to be equivalent if there exists a permutation ϕ of X such that $f = \phi(g)$.

Note that if two polynomials $f, g \in \mathbb{F}[X]$ are equivalent, then their complexity with regards to $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas is the same. That is, there exists a $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formula of size s for f if and only if there exists a $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formula of size s for g.

Claim 14. *For any restriction $\sigma \in \mathcal{R}$ as defined above, $\mathrm{IMM}_{n,d}|_\sigma$ is equivalent to F.*

Proof. Recall that the polynomial F was obtained from $\mathrm{IMM}_{n,d}$ by the restriction σ:

$$
\sigma(x) = \begin{cases} 0 & \text{if } x = x_j^{(1)} \text{ for some } j \neq 1, \text{ or} \\ & \text{if } x = x_j^{(d)} \text{ for some } j \neq 1, \text{ or} \\ & \text{if } x = x_{i,j}^{(p)} \text{ for } i \neq j \text{ and } p \notin P' \cup \{1, d\}. \\ * & \text{otherwise.} \end{cases}
$$

Consider a restriction τ obtained by picking j_1, j_d and permutations π_p for $p \notin P' \cup \{1, d\}$. We wish to define a permutation ϕ such that $\mathrm{IMM}_{n,d}|_\tau = \phi(F)$. We start necessarily by defining $\phi(x_1^{(1)}) = x_{j_1}^{(1)}$.

We then define ϕ for the matrices X_p for $p \in \{2, \ldots, p'_1 - 1\}$. Once again viewing our polynomials as ABPs: let $\mathcal{A}_\sigma$ be the graph corresponding to F and $\mathcal{A}_\tau$ the graph corresponding to $\mathrm{IMM}_{n,d}|_\tau$. To lighten notations, we only write the index of the vertex at each layer; the path $1, 1, \ldots, 1$ from layer p to layer q is thus the path $v_1^{(p)}, v_1^{(p+1)} \ldots, v_1^{(q)}$.

Note that there are n pairwise edge-disjoint paths $i, i, \ldots, i$ going from layer 1 to layer $p_1' - 1$ in $\mathcal{A}_\sigma$, one path for each $i \in [n]$. There are also n pairwise edge-disjoint paths going from layer 1 to layer $p_1' - 1$ in $\mathcal{A}_\tau$, these paths being defined by the composition of the permutations π_p for $p \in \{2, \ldots, p_1' - 1\}$. We define ϕ successively in each matrix $X_2, X_3, \ldots, X_{p'-1}$ so that it sends the path $1, 1, \ldots, 1$ to the path starting from vertex j_1. Let i' be the end-vertex of this path at layer $p_1' - 1$. We then define ϕ over the variables in $X_{p_1'}$ by only requiring that $\phi(x_{1,j}^{p_1'})$ be equal to $x_{i',j}^{p_1'}$ for all $j \in [n]$.

We will then define the permutation separately over the following intervals of matrix indices: $\{p_1' + 1, \ldots, p_1\}, \{p_1 + 1, \ldots, p_2'\}, \ldots, \{p_k + 1, \ldots, p_{k+1}' - 1\}$. We will describe the case of the interval $\{p_1' + 1, \ldots, p_1\}$ in some detail, other intervals can be treated in a similar fashion, with a slight exception for the last one.

Once again in $\mathcal{A}_\sigma$ there are n pairwise edge-disjoint paths from layer p_1' to layer $p_1 - 1$. Similarly, there is a set of n pairwise edge-disjoint paths from from layer p_1' to layer $p_1 - 1$ in the graph $\mathcal{A}_\tau$. We define ϕ such that it sends the path $i, i, \ldots, i$ to the path starting at vertex i of layer p_1' in $\mathcal{A}_\tau$. For example, we send the path $1, 1, \ldots, 1$ in $\mathcal{A}_\sigma$ to the path $1, \pi_{p_1'+1}(1), \pi_{p_1'+2} \circ \pi_{p_1'+1}(1), \ldots, \pi_{p_1-1} \circ \pi_{p_1-2} \circ \cdots \circ \pi_{p_1'+2} \circ \pi_{p_1'+1}(1)$. Next we define the effect of ϕ on the matrix X_{p_1}. Define the permutation α by setting $\alpha(i)$ to be the index of the end-vertex of the path starting at vertex i of layer p_1' in $\mathcal{A}_\tau$, i.e., $\alpha(i) = \pi_{p_1-1} \circ \pi_{p_1-2} \circ \cdots \circ \pi_{p_1'+2} \circ \pi_{p_1'+1}(i)$. We then let $\phi(x_{i,j}^{(p_1)}) = x_{\alpha(i),j}^{(p_1)}$ for all $i \in [n]$. A path $i, i, \ldots, i$ in $\mathcal{A}_\sigma$ is sent by ϕ to the path starting at i and ending at $\alpha(i)$ in $\mathcal{A}_\tau$. This path $i, i, \ldots, i$ in $\mathcal{A}_\sigma$ could then be extended by any edge (i, j) in X_{p_1}. Since we have sent the path $i, i, \ldots, i$ to the path ending in $\alpha(i)$, by setting $\phi(x_{i,j}^{(p_1)}) = x_{\alpha(i),j}^{(p_1)}$ we ensure that any path from vertex i of layer p_1' to vertex j of layer p_1 in $\mathcal{A}_\sigma$ is sent to a path from vertex i of layer p_1' to vertex j of layer p_1 in $\mathcal{A}_\tau$. We can then start the process again with the next interval, with the exception of the last one, where we do not set ϕ for the variables in $X_{p_{k+1}'}$ yet. Let α be the permutation obtained as above for the interval $\{p_k + 1, \ldots, p_{k+1}' - 1\}$

To define ϕ on the end of the graph, we will start from the end. Clearly, we must set $\phi(x_1^{(d)}) = x_{j_d}^{(d)}$. We then need to do the interval $\{p_{k+1}' + 1, \ldots, d - 1\}$. We define ϕ on the interval $\{p_{k+1}' + 1, \ldots, d - 1\}$ by sending the path $1, 1, \ldots, 1$ to the unique path ending at vertex j_d of layer $d - 1$ in $\mathcal{A}_\tau$. Let j' be the index at layer p_{k+1}' of the starting vertex of this path. To define ϕ on $X_{p_{k+1}'}$, we only require of ϕ that it send $x_{i,1}^{(p_{k+1}')}$ to $x_{\alpha(i),j'}^{(p_{k+1}')}$ for all $i \in [n]$.

Then $\mathrm{IMM}_{n,d}|_\tau = \phi(F)$. $\square$

The following lemma is then obvious.

Lemma 15. *Let $n, d, D, t, k \in \mathbb{N}$ be such that $1 \leqslant t \leqslant d/320$, $n \geqslant 10$, and $k \leqslant d/320t$. Let τ be a restriction in $\mathcal{R}$. Then any $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ for $\mathrm{IMM}_{n,d}|_\tau$ has top fan-in at least $\Omega\left(\left(\frac{n^{3/4}}{4D}\right)^k\right)$.*

5.3 From set-multilinear formulas to $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas

In this section we reduce the case of a depth 4 set-multilinear formula to the case of bounded bottom fan-in by finding a suitable nice restriction.

Lemma 16. *Let n, d be large enough integers, and $k, t \in \mathbb{N}$ be such that $d \leqslant n^{1/10}$ and $t \geqslant 4k$. Let C be a set-multilinear $\Sigma\Pi\Sigma\Pi$ formula of size $s < n^{t/10}$. Then there exists a restriction $\tau \in \mathcal{R}$ such that $C|_\tau$ is a $\Sigma\Pi\Sigma\Pi^{(t)}$ formula.*

Proof. We consider the uniform distribution over restrictions in $\mathcal{R}$ and prove that with high probability the property in Lemma 16 holds.

Let us fix any bottom level Π gate G in C that has fan-in $t' > t$. Let m be the (set-multilinear) monomial computed by this Π gate. We can write m as a product of t' variables, each coming from a different variable set. That is, there exists a set $S \subseteq [d], |S| = t'$ such that $m = \Pi_{i \in S} y^{(i)}$, where $y^{(i)}$ is a variable from X_i. We claim the following:

$$\Pr_\tau[G \text{ not set to } 0 \text{ in } C|_\tau] \leqslant \frac{1}{n^{t/3}} \tag{4}$$

To see this, first note for all $p \in P'$, each variable in X_p survives with probability 1, i.e., the restriction does not set any variable in X_p to 0. But from the definition of our restriction and choice of t, $|P'| = 2k+1$ and $t' \geqslant t \geqslant 4k$. Therefore, the monomial m has at least $t/3$ variables coming from matrices X_p ($p \notin P'$). And for all $p \notin P'$, the probability over τ that a variable survives in X_p is exactly $1/n$. Therefore, the probability that the monomial m survives is at most $(1/n)^{t/3}$. Therefore we have (4).

Since there are at most $s < n^{t/10}$ bottom level Π gates of fan-in greater than t, by a union bound, the probability that any such Π gate survives is at most $n^{t/10} \cdot \frac{1}{n^{t/3}} = o(1)$.

□

We can now bring everything together.

Theorem 17. *Let $n, d \in \mathbb{N}$ be such that $d \leqslant n^{1/10}$. Let C be any set-multilinear $\Sigma\Pi\Sigma\Pi$ formula computing $\mathrm{IMM}_{n,d}$. Then C has size $n^{\Omega(\sqrt{d})}$.*

Proof. Let us choose k, t such that $d/640 \leqslant kt \leqslant d/320$ and $t = 4k$. Let C be a $\Sigma\Pi\Sigma\Pi$ set-multilinear formula of size s computing $\mathrm{IMM}_{n,d}$ and say $s < n^{t/10}$. Let σ be the restriction guaranteed by Lemma 16. Therefore, we have that $C|_\sigma$ is a $\Sigma\Pi\Sigma\Pi^{(t)}$ formula computing $\mathrm{IMM}_{n,d}|_\sigma$, which is equivalent to F. Since $C|_\sigma$ is also a set-multilinear formula computing a degree d polynomial, every Π gate at layer 3 has fan-in at most d. From Theorem 12, we have that any $\Sigma\Pi^{(d)}\Sigma\Pi^{(t)}$ circuit computing F has size at least $\frac{1}{4} \cdot \left(\frac{n^{3/4}}{4d}\right)^k$, i.e., $n^{\Omega(k)}$ (as $d \leqslant n^{1/10}$). Therefore, we get that $s > \min\left\{n^{\Omega(t)}, n^{\Omega(k)}\right\}$. Since $kt = \Theta(d)$ and $t = 4k$, we have proved $s = n^{\Omega(\sqrt{d})}$. □

Remark 18. *Raz and Yehudayoff [RY09] proved that any $\Sigma\Pi\Sigma\Pi$ multilinear formula computing the determinant polynomial has size $\exp(\Omega(n^{1/27}))$ (note that their bound also hold in the stronger model of $\Sigma\Pi\Sigma\Pi\Sigma$ formulas). Using a carefully defined restriction (which is a function from $\{x_{i,j}\}_{i,j \in [n]}$ to $\{0, 1, *\}$ instead of to $\{0, *\}$) along the above lines together with the result of [GKKS13], this lower bound can be improved to $\exp(\Omega(n^{1/2}))$ in the set-multilinear case.*

We consider the determinant polynomial of a generic $n \times n$ matrix which is set-multilinear with respect to its columns. Let $S, T \subseteq [n]$ and $|S| = |T| = n/2$ be chosen uniformly at random.

Let ϕ be a random bijection from $[n]\backslash S$ to $[n]\backslash T$. Now consider a restriction σ as given below: $\sigma(x_{i,j}) = *$ if $i \in S, j \in T$, $\sigma(x_{i,\phi(i)}) = 1$ if $i \in [n]\backslash S$, and $\sigma(x_{i,j}) = 0$ otherwise. Under this restriction, an $n \times n$ determinant reduces to an $n/2 \times n/2$ determinant of the matrix defined by S, T. And under this restriction a $\Sigma\Pi\Sigma\Pi$ set-multilinear formula of size at most $2^{o(\sqrt{n})}$ computing the determinant reduces to a $\Sigma\Pi\Sigma\Pi^{(t)}$ set-multilinear formula where $t = O(\sqrt{n})$ whp. Therefore, there exists a restriction which along with the result of [GKKS13] gives a lower bound of $\exp(\Omega(n^{1/2}))$ *for* $\Sigma\Pi\Sigma\Pi$ *set-multilinear formula computing the determinant polynomial.*

6 Homogeneous multilinear depth-4 formulas

We will follow the same strategy as in Section 5, first defining a set of nice restrictions, then proving a lower bound for $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas computing them, and finally showing that there exists a restriction in the set which changes a homogeneous multilinear formula to a $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formula.

6.1 Nice restrictions of $\mathrm{IMM}_{n,d}$

Our restrictions are once again related to the evenly-spaced matrices chosen before. We will now choose new indices p''_q in a slightly different way. For $q \in [k+1]$, let $p''_q \in [d]$ be defined so that $p_{q-1} < p''_q < p_q$ and $\min\{p''_q - (p_{q-1}+1), p_q - (p''_q + 2)\} \geqslant \left\lfloor \frac{r}{2} \right\rfloor - 1$. Let P''_1 denote the set $\{p_q \mid q \in [k]\}$ and P''_2 denote $\left\{p''_q \,\middle|\, q \in [k+1]\right\}$. We define P'' to be $P''_1 \cup \{p, p+1 \mid p \in P''_2\}$.

Definition 19. *Let $\mathcal{R}$ be the set of restrictions τ such that:*

1. *for $p \in \{1, d\}$, there is a* unique *$j_p \in [n]$ such that $\tau(x^{(p)}_{j_p}) = *$,*
2. *for $p \notin P'' \cup \{1, d\}$, there is a permutation π_p of $[n]$ such that for any $i, j \in [n]$, $\tau(x^{(p)}_{i,j}) = *$ iff $j = \pi_p(i)$,*
3. *for $p \in P''_2$ and for all $i, j \in [n]$, there is* at least one *h in $[n]$ such that $\tau(x^{(p)}_{i,h}) = \tau(x^{(p+1)}_{h,j}) = *$,*
4. *for $p \in P''_1$, $|X_p \cap \tau^{-1}(*)| \geqslant n^{1.7}$.*

6.2 A lower bound for nice restrictions of $\mathrm{IMM}_{n,d}$

We will not proceed exactly as we did in Section 5, where we chose a specific nice restriction and showed the lower bound for the resulting polynomial. Instead, we study the polynomial obtained from a nice restriction in general.

6.2.1 The dimension of the space of shifted partial derivatives of nice restrictions of $\mathrm{IMM}_{n,d}$

Let σ be a restriction in $\mathcal{R}$ and F the polynomial $\mathrm{IMM}_{n,d}|_\sigma$. Let $\mathcal{A}_\sigma$ be the ABP corresponding to F.

As in Section 5.1, we first analyze $\partial_{\mathcal{I}} F$ for $\mathcal{I} \in [n]^{2k}$. Let $\mathcal{I} = (i_1, j_1, \dots, i_k, j_k)$. Clearly, if there is a $q \in [k]$ such that $\sigma(x_{i_q,j_q}^{(p_q)}) = 0$, then we have $\partial_{\mathcal{I}} F = 0$. Consequently we say $\mathcal{I}$ is *surviving* if for all $q \in [k]$, we have $\sigma(x_{i_q,j_q}^{(p_q)}) = *$. Let $\mathcal{T} = \{\mathcal{I} \in [n]^{2k} \mid \mathcal{I} \text{ surviving}\}$. By property 3 in Definition 19, we have $|\mathcal{T}| = |\bigtimes_{q=1}^{k}(X_p \cap \sigma^{-1}(*))| \geqslant n^{(1.7)\cdot k}$.

For any $\mathcal{I} = (i_1, j_1, \dots, i_k, j_k) \in \mathcal{T}$, the polynomial $\partial_{\mathcal{I}} F$ is the sum of all monomials m such that $m = \rho_1 \rho_2 \dots \rho_{k+1}$, where ρ_q is a path from $v_{j_{q-1}}^{(p_q-1)}$ to $v_{i_q}^{(p_q-1)}$ in $\mathcal{A}_\sigma$ for all $2 \leqslant q \leqslant k$, ρ_1 is a path from vertex $v^{(0)}$ to $v_{i_1}^{(p_1)}$ in $\mathcal{A}_\sigma$, and ρ_{k+1} is a path from $v_{j_k}^{(p_k)}$ to vertex $v^{(d)}$ in $\mathcal{A}_\sigma$. Clearly, $\partial_{\mathcal{I}} F$ is a homogeneous polynomial of degree $d-k$.

Moreover, given any $\mathcal{I} \in \mathcal{T}$, the polynomial $\partial_{\mathcal{I}} F$ is non-zero; that is, there is a path from $v^{(0)}$ to $v^{(d)}$ in $\mathcal{A}_\sigma$ which contains each edge $(v_{i_q}^{(p_q-1)}, v_{j_q}^{(p_q)})$ for $q \in [k]$ by properties 1, 2, and 3 in Definition 19.

We would like to lower bound $\dim(\langle \partial_k F \rangle_{\leqslant \ell})$, which is at least $\dim(\mathcal{V})$, where

$$\mathcal{V} = \operatorname{span}\{m \cdot \partial_{\mathcal{I}} F \mid \mathcal{I} \in \mathcal{T} \text{ and } m \text{ a monomial of degree at most } \ell\}.$$

To lower bound $\dim(\mathcal{V})$, we will use the monomial-ordering technique as in [GKKS13] (see also [CLO97]). Let $\succeq$ be an arbitrary linear ordering of the variables in X and extend it to the lexicographic ordering on the set of all monomials in $\mathbb{F}[X]$ — the linear order on monomials is also denoted $\succeq$. Given this monomial ordering $\succeq$, for any polynomial $f \in \mathbb{F}[X]$, we denote by $\mathrm{LM}(f)$ the leading monomial of f under this ordering (the ordering will be clear from context). The following fact will be useful.

Fact 20. *Let $\succeq$ be any ordering as described above. Let $m_1, m_2 \in \mathbb{F}[X]$ be arbitrary monomials such that $m_1 \succeq m_2$. Then, for any monomial m, we have*

$$m_1 \cdot m \succeq m_2 \cdot m$$

This immediately implies that for $f, g \in \mathbb{F}[X]$, we have $\mathrm{LM}(f \cdot g) = \mathrm{LM}(f) \cdot \mathrm{LM}(g)$.

Now, to lower bound $\dim(\mathcal{V})$, note that by Gaussian elimination, we know that

$$\begin{aligned}\dim(\mathcal{V}) &= |\{\mathrm{LM}(f) \mid f \in \mathcal{V}\}| \\ &\geqslant |\{\mathrm{LM}(m \cdot \partial_{\mathcal{I}} F) \mid \mathcal{I} \in \mathcal{T} \text{ and } m \text{ a monomial of degree at most } \ell\}| \\ &= |\{m \cdot \mathrm{LM}(\partial_{\mathcal{I}} F) \mid \mathcal{I} \in \mathcal{T} \text{ and } m \text{ a monomial of degree at most } \ell\}|\end{aligned}$$

where the last equality follows from Fact 20. We denote by $p(\mathcal{I})$ the monomial $\mathrm{LM}(\partial_{\mathcal{I}} F)$. From the above, we see that $\dim(\mathcal{V}) \geqslant |\mathcal{M}|$, where

$$\mathcal{M} = \{m' \mid m' \text{ a monomial of degree at most } \ell + d - k \text{ and } \exists\, \mathcal{I} \in \mathcal{T} \text{ such that } p(\mathcal{I}) | m'\}.$$

Also, for $\mathcal{I} \in \mathcal{T}$, if we let

$$\mathcal{M}_{\mathcal{I}} = \{m' \mid m' \text{ a monomial of degree at most } \ell + d - k \text{ such that } p(\mathcal{I}) | m'\},$$

then we have $\mathcal{M} = \bigcup_{\mathcal{I} \in \mathcal{T}} \mathcal{M}_{\mathcal{I}}$.

The above arguments prove the following claim.

Claim 21. $\dim(\langle \partial_k F \rangle_{\leqslant \ell}) \geqslant |\mathcal{M}|$.

We need some technical claims, which are analogous to the claims proved in Section 5.

Claim 22. *For any $\mathcal{I}, \mathcal{I}' \in \mathcal{T}$, we have*

$$|p(\mathcal{I}') \backslash p(\mathcal{I})| \geqslant \Delta(\mathcal{I}, \mathcal{I}') \cdot \left(\left\lfloor \frac{r}{2} \right\rfloor - 1 \right),$$

where $\Delta(\mathcal{I}, \mathcal{I}')$ denotes the Hamming distance between $\mathcal{I}$ and $\mathcal{I}'$.

Proof. Let $\mathcal{I}, \mathcal{I}' \in \mathcal{T}$. Say $\mathcal{I} = (i_1, j_1, \ldots, i_k, j_k)$ and $\mathcal{I}' = (i'_1, j'_1, \ldots, i'_k, j'_k)$. In the ABP $\mathcal{A}_\sigma$, let $g_1^{\mathcal{I}}$ be the unique path from $v^{(0)}$ to $V_{p''_1 - 1}$ and for all $q \in [k]$, let $g_{q+1}^{\mathcal{I}}$ be the unique path from $v_{j_q}^{(p_q)}$ to $V_{p''_{q+1}-1}$. For all $q \in [k]$, let $h_q^{\mathcal{I}}$ be the unique path from $V_{p''_q + 1}$ to $v_{i_q}^{(p_q - 1)}$, and let $h_{k+1}^{\mathcal{I}}$ be the unique path from $V_{p''_{k+1}+1}$ to $v^{(d)}$. (These paths are unique by property 2 in Definition 19). For $q \in [k+1]$, define $g_q^{\mathcal{I}'}$ and $h_q^{\mathcal{I}'}$ in the same way for $\mathcal{I}'$. We have $p(\mathcal{I}) = m \cdot \prod_{q \in [k+1]} g_q^{\mathcal{I}} h_q^{\mathcal{I}}$ and $p(\mathcal{I}') = m' \cdot \prod_{q \in [k+1]} g_q^{\mathcal{I}'} h_q^{\mathcal{I}'}$ where m and m' are monomials on the variables $\bigcup_{q \in P''_2} (X_q \cup X_{q+1})$. Hence $|p(\mathcal{I}') \backslash p(\mathcal{I})| \geqslant \sum_{q \in [k+1]} (|g_q^{\mathcal{I}'} \backslash g_q^{\mathcal{I}}| + |h_q^{\mathcal{I}'} \backslash h_q^{\mathcal{I}}|)$.

If $i_q \neq i'_q$, the paths $h_q^{\mathcal{I}}$ and $h_q^{\mathcal{I}'}$ are edge disjoint (again by property 2 in Definition 19). In the same way, $g_{q+1}^{\mathcal{I}}$ and $g_{q+1}^{\mathcal{I}'}$ are edge disjoint if $j_q \neq j'_q$. Now all paths $g_q^{\mathcal{I}}$, $h_q^{\mathcal{I}}$ (and $g_q^{\mathcal{I}'}, h_q^{\mathcal{I}'}$) are of length at least $\lfloor r/2 \rfloor - 1$ by the choice of p_q and p''_q. ☐

Claim 23. *For any $\mathcal{I} \in \mathcal{T}$, we have $|\mathcal{M}_{\mathcal{I}}| = \binom{N+\ell}{\ell}$.*

Claim 24. *For any $\mathcal{I}, \mathcal{I}' \in \mathcal{T}$, we have $|\mathcal{M}_{\mathcal{I}} \cap \mathcal{M}_{\mathcal{I}'}| = \binom{N+\ell-|p(\mathcal{I}') \backslash p(\mathcal{I})|}{\ell - |p(\mathcal{I}') \backslash p(\mathcal{I})|}$.*

Claim 25. *Fix any $k, n \in \mathbb{N}$. Then there exists an $\mathcal{S} \subseteq \mathcal{T}$ such that*

- $|\mathcal{S}| = \left\lfloor \left(\frac{\sqrt{n}}{4} \right)^k \right\rfloor$,
- *For all distinct $\mathcal{I}, \mathcal{I}' \in \mathcal{S}$, we have $\Delta(\mathcal{I}, \mathcal{I}') \geqslant k$.*

Proof. As in the proof of Claim 10, a volume argument shows that we can pick $\left\lfloor \frac{n^{1.7k}}{n^k \binom{2k}{k}} \right\rfloor \geqslant \frac{n^{k/2}}{2^{2k}}$ elements in $\mathcal{T}$ with pairwise Hamming distance at least k. ☐

Now we are ready to prove lower bound on the dimension of the space of shifted partial derivatives of F.

Lemma 26. *Let $k, \ell \in \mathbb{N}$ be arbitrary parameters such that $20k < d < \ell$ and $k \geqslant 2$. Then,*

$$\dim(\langle \partial_k F \rangle_{\leqslant \ell}) \geqslant M \cdot \binom{N+\ell}{\ell} - M^2 \cdot \binom{N+\ell-d/10}{\ell - d/10},$$

where $M = \left\lfloor \left(\frac{\sqrt{n}}{4} \right)^k \right\rfloor$.

Proof. By Claim 21, it suffices to lower bound $|\mathcal{M}|$. Since $\mathcal{M} = \bigcup_{\mathcal{I}} \mathcal{M}_{\mathcal{I}}$, we have

$$\begin{aligned} |\mathcal{M}| &\geqslant |\bigcup_{\mathcal{I}\in\mathcal{S}} \mathcal{M}_{\mathcal{I}}| \\ &\geqslant \sum_{\mathcal{I}\in\mathcal{S}} |\mathcal{M}_{\mathcal{I}}| - \sum_{\mathcal{I}\neq\mathcal{I}'\in\mathcal{S}} |\mathcal{M}_{\mathcal{I}} \cap \mathcal{M}_{\mathcal{I}'}|. \end{aligned} \tag{5}$$

By Claim 23, we know that $|\mathcal{M}_{\mathcal{I}}| = \binom{N+\ell}{\ell}$. By Claims 24 and 22 and our choice of $\mathcal{S}$ (Claim 25), we see that for any distinct $\mathcal{I}, \mathcal{I}' \in \mathcal{S}$, we have

$$|\mathcal{M}_{\mathcal{I}} \cap \mathcal{M}_{\mathcal{I}'}| \leqslant \binom{N+\ell-k(\lfloor r/2\rfloor - 1)}{\ell - k(\lfloor r/2\rfloor - 1)} \leqslant \binom{N+\ell-d/10}{\ell-d/10}$$

where the last inequality follows since $\lfloor r/2\rfloor - 1 \geqslant d/10k$ for $k \leqslant d/20$.

Plugging the above into (5), we obtain

$$|\mathcal{M}| \geqslant |\mathcal{S}| \cdot \binom{N+\ell}{\ell} - |\mathcal{S}|^2 \cdot \binom{N+\ell-d/10}{\ell-d/10}.$$

Since $|\mathcal{S}| = \left\lfloor \left(\frac{\sqrt{n}}{4}\right)^k \right\rfloor$, the lemma follows. $\square$

6.2.2 A lower bound for $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas computing nice restrictions of $\mathrm{IMM}_{n,d}$

Lemma 27. *Let $n, D, k, t, d \in \mathbb{N}$ be such that $1 \leqslant k, t \leqslant d \leqslant n^{1/10}$, $D \leqslant n^{1/4}/100$, and $kt \leqslant d/320$. Let σ be a restriction in $\mathcal{R}$. Then any $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formula C for $\mathrm{IMM}_{n,d}|_\sigma$ has top fan-in $\Omega\left(\left(\frac{n^{1/4}}{4D}\right)^k\right)$.*

Proof. Let $F = \mathrm{IMM}_{n,d}|_\sigma$. We proceed as in Theorem 12. Fix $\ell \in \mathbb{N}$ such that $n^{1/16} \leqslant \left(\frac{N+\ell}{\ell}\right)^t \leqslant n^{1/4}$, which exists by Claim 5. By Lemma 26, we have

$$\dim(\langle \partial_k F\rangle_{\leqslant \ell}) \geqslant \underbrace{M \cdot \binom{N+\ell}{\ell}}_{T_1} - \underbrace{M^2 \cdot \binom{N+\ell-d/10}{\ell-d/10}}_{T_2}$$

where $M = \left\lfloor (\frac{\sqrt{n}}{4})^k \right\rfloor$.

However, for our choice of parameters, we have

$$\begin{aligned} \frac{T_1}{T_2} &= \frac{\binom{N+\ell}{\ell}}{M \cdot \binom{N+\ell-d/10}{\ell-d/10}} \\ &\geqslant \frac{1}{M} \cdot \left(\frac{N+\ell}{\ell}\right)^{d/10} && \text{(by Fact 4)} \\ &\geqslant \frac{n^{d/320t}}{M} && \text{(by our choice of } \ell\text{)} \\ &\geqslant \frac{n^k}{M} \geqslant 4^k \geqslant 2. \end{aligned}$$

Hence, we have

$$\dim(\langle \partial_k F \rangle_{\leqslant \ell}) \geqslant T_1 - T_2 \geqslant T_1/2 = \frac{M}{2} \cdot \binom{N+\ell}{\ell}. \tag{6}$$

Now, let C be a $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formula for F of top fan-in s. Then, by Lemma 3, we have

$$\dim(\langle \partial_k C \rangle_{\leqslant \ell}) \leqslant s \cdot \binom{D}{k} \cdot \binom{N+\ell+(t-1)k}{\ell+(t-1)k} \leqslant s \cdot D^k \cdot \binom{N+\ell+(t-1)k}{\ell+(t-1)k}.$$

An application of inequality (6) implies that we must have

$$s \cdot D^k \cdot \binom{N+\ell+(t-1)k}{\ell+(t-1)k} \geqslant \frac{M}{2} \cdot \binom{N+\ell}{\ell}.$$

Therefore,

$$\begin{aligned}
s &\geqslant \frac{M}{2D^k} \cdot \frac{\binom{N+\ell}{\ell}}{\binom{N+\ell+(t-1)k}{\ell+(t-1)k}} \\
&\geqslant \frac{M}{2D^k} \cdot \left(\frac{\ell}{N+\ell}\right)^{(t-1)k} && \text{(by Fact 4)} \\
&\geqslant \frac{1}{4D^k} \cdot \left(\frac{\sqrt{n}}{4}\right)^k \cdot \left(\frac{\ell}{N+\ell}\right)^{(t-1)k} && \text{(by our choice of } M) \\
&= \frac{1}{4} \cdot \left(\frac{\sqrt{n}}{4D} \cdot \left(\frac{\ell}{N+\ell}\right)^{(t-1)}\right)^k \\
&\geqslant \frac{1}{4} \cdot \left(\frac{\sqrt{n}}{4D \cdot \left(\frac{N+\ell}{\ell}\right)^t}\right)^k \geqslant \frac{1}{4} \cdot \left(\frac{n^{1/4}}{4D}\right)^k && \text{(by our choice of } \ell).
\end{aligned}$$

□

6.3 From homogeneous multilinear formulas to $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas

Lemma 28. *The following holds for any large enough $n \in \mathbb{N}$, and any k, t, d such that $1 \leqslant k, t \leqslant d \leqslant n^{1/10}$. Let C be a homogeneous multilinear $\Sigma\Pi\Sigma\Pi$ formula of size $s < n^{t/10}$, there is a restriction $\tau \in \mathcal{R}$ such that $C|_\tau$ is $\Sigma\Pi\Sigma\Pi^{(t)}$.*

Proof. As in Section 5, we will use a probabilistic argument. We define a suitable distribution $\mathcal{D}$ over restrictions $\sigma : X \to \{0, *\}$ in general and show that with high probability over the choice of $\sigma \sim \mathcal{D}$, the restriction $F := \mathrm{IMM}_{n,d}|_\sigma$ belongs to $\mathcal{R}$ and satisfies the required property. We specify the distribution $\mathcal{D}$ by describing how to sample a single restriction σ.

- For $p \in \{1, d\}$, pick $j_p \in [n]$ uniformly at random. Set $\sigma(x_j^{(p)}) = *$ if $j = j_p$ and 0 otherwise.
- For $p \notin P'' \cup \{1, d\}$, pick an independent and uniformly random permutation π_p of $[n]$. Set $\sigma(x_{i,j}^{(p)}) = *$ if $j = \pi_p(i)$ and 0 otherwise.
- For each $x \in \bigcup_{p \in P''} X_p$, set $\sigma(x) = *$ independently with probability $\frac{1}{n^{0.2}}$.

We denote by $\mathcal{A}_\sigma$ the ABP corresponding to this restriction.

σ belongs to $\mathcal{R}$ with high probability. From the description of $\mathcal{D}$ above, it follows that for any $\sigma \sim \mathcal{D}$, the restriction σ always satisfies properties 1 and 2 of Definition 19. So we need only consider properties 3 and 4.

Let $\mathcal{E}_3$ denote the event that property 3 is not satisfied. Fix any $p \in P_2''$ and any $i, j \in [n]$. For each $v \in V_p$, define the $\{0,1\}$-valued random variable $Y_{v,i,j}$ so that $Y_{v,i,j} = 1$ iff the edges $(v_i^{(p-1)}, v)$ and $(v, v_j^{(p+1)})$ both survive in the ABP $\mathcal{A}_\sigma$ and let $Y'_{p,i,j} = \sum_{v \in V_p} Y_{v,i,j}$. Since each of the edges $(v_i^{(p-1)}, v)$ and $(v, v_j^{(p+1)})$ survives independently with probability $1/n^{0.2}$, it follows that $\Pr_\sigma[Y_{v,i,j} = 1] = 1/n^{0.4}$.

Note that the random variables $Y_{v,i,j}$ are mutually independent. Hence, we have

$$\begin{aligned}
\Pr_\sigma[\text{There is no path from } v_i^{(p-1)} \text{ to } v_j^{(p+1)}] &= \Pr_\sigma[Y'_{p,i,j} = 0] = \Pr_\sigma[\bigwedge_v Y_{v,i,j} = 0] \\
&= \prod_v \Pr_\sigma[Y_{v,i,j} = 0] \\
&= \left(1 - \frac{1}{n^{0.4}}\right)^n = \exp(-n^{\Omega(1)}).
\end{aligned}$$

Union bounding over the choices of p, i, j, we see that $\Pr[\mathcal{E}_3] = \exp(-n^{\Omega(1)}) = o(1)$.

Let $\mathcal{E}_4$ denote the event that property 4 is not satisfied. Fix any $p \in P_1''$ and for each $x \in X_p$, define the $\{0,1\}$-valued random variable $Z(x)$ that is 1 iff $\sigma(x) = *$; let $Z_p = \sum_{x \in X_p} Z(x)$.

We have $\mathbf{E}_\sigma[Z_p] = \sum_{x \in X_p} \Pr_\sigma[Z(x) = 1] = |X_p| \cdot n^{-0.2} = n^{1.8}$. Moreover, the random variables $Z(x)$ are mutually independent and hence, by a Chernoff bound (see, e.g., [DP09, Chapter 1]), we have for large enough n,

$$\Pr_\sigma[Z_p < n^{1.7}] \leqslant \Pr_\sigma[Z_p < \mathbf{E}[Z_p]/2] = \exp(-n^{\Omega(1)}).$$

Thus, union bounding over the at most $d \leqslant n^{1/10}$ choices of $p \in P_1''$, we have $\Pr[\mathcal{E}_4] \leqslant n^{1/10} \cdot \exp(-n^{\Omega(1)}) = \exp(-n^{\Omega(1)}) = o(1)$.

Thus, we have

$$\Pr_\sigma[\sigma \text{ not valid}] = \Pr_\sigma[\mathcal{E}_3 \vee \mathcal{E}_4] = o(1).$$

$C|_\sigma$ is $\Sigma\Pi\Sigma\Pi^{(t)}$ with high probability. We need to show that, with high probability, the fan-in of bottom Π gates of $C|_\sigma$ is at most t. In other words, we need to show that with high probability, all bottom Π gates in C that have fan-in greater than t are set to 0 by σ.

Let us fix any bottom level Π gate G in C that has fan-in greater than t. Let m be the (multilinear) monomial computed by this Π gate.[1] Write $m = m_1 \cdot m_2 \cdots m_d$, where $m_p = \prod_{x \in X_p : x | m} x$. Let t_p denote $\deg(m_p)$. We have $\sum_{p \in [d]} t_p = \deg(m) > t$. We claim that

$$\Pr_\sigma[G \text{ not set to 0 in } C|_\sigma] \leqslant \frac{1}{n^{t/5}}. \tag{7}$$

[1]This is the only place where multilinearity is necessary.

To see this, note that by the independence of the random restriction σ across the different X_p ($p \in [d]$), we have

$$\Pr_{\sigma}[G \text{ not set to } 0 \text{ in } C|_{\sigma}] = \prod_{p \in [d]} \underbrace{\Pr_{\sigma}[\text{No variable in } m_p \text{ set to } 0]}_{\alpha_p}. \tag{8}$$

Now, fix any $p \in [d]$. We upper bound α_p based on a case analysis.

- If $p \in \{1, d\}$, $\alpha_p = 0$ if $t_p > 1$ and $\alpha_p = 1/n^{t_p}$ otherwise.
- If $p \in P''$, $\alpha_p = 1/n^{t_p/5}$.
- If $p \notin P'' \cup \{1, d\}$, then $\alpha_p = 0$ if the monomial m_p contains at least two variables from any row or column of X_p. Otherwise, $\alpha_p = \prod_{z=0}^{t_p-1} \frac{1}{n-z} \leqslant \left(\frac{1}{n-t_p}\right)^{t_p} \leqslant \frac{1}{n^{t_p/2}}$, where the last inequality follows since $t_p \leqslant t \leqslant d \leqslant n^{1/10}$.

Thus, we see that in all cases, we have $\alpha_p \leqslant \frac{1}{n^{t_p/5}}$. Substituting in (8), we have $\Pr_{\sigma}[G \text{ not set to } 0] \leqslant 1/n^{\sum_{p \in [d]}(t_p/5)} \leqslant 1/n^{t/5}$, which proves (7).

Since there are at most $s < n^{t/10}$ bottom level Π gates of fan-in greater than t, by a union bound, the probability that any such Π gate survives is at most $n^{t/10} \cdot \frac{1}{n^{t/5}} = o(1)$.

Finally, another union bound proves the lemma. □

Theorem 29. *Let $n, d \in \mathbb{N}$ be such that $d \leqslant n^{1/10}$. Let C be a homogeneous multilinear $\Sigma\Pi\Sigma\Pi$ formula computing $\mathrm{IMM}_{n,d}$. Then C has size $n^{\Omega(\sqrt{d})}$.*

Proof. Let us choose k, t such that $d/640 \leqslant kt \leqslant d/320$ and $t = 4k$. Let C be a $\Sigma\Pi\Sigma\Pi$ homogeneous multilinear formula of size s computing $\mathrm{IMM}_{n,d}$ and say $s < n^{t/10}$. Let σ be the nice restriction guaranteed by Lemma 28. Therefore, we have that $C|_{\sigma}$ is a $\Sigma\Pi\Sigma\Pi^{(t)}$ formula computing $F = \mathrm{IMM}_{n,d}|_{\sigma}$. Since $C|_{\sigma}$ is also a homogeneous multilinear formula computing a degree d polynomial, every Π gate at layer 3 has fan-in at most d. From Lemma 27, we have that any $\Sigma\Pi^{(d)}\Sigma\Pi^{(t)}$ circuit computing F has size at least $\frac{1}{4} \cdot \left(\frac{n^{1/4}}{4d}\right)^k$, i.e., $n^{\Omega(k)}$ (as $d \leqslant n^{1/10}$).

Therefore, we get that $s > \min\left\{n^{\Omega(t)}, n^{\Omega(k)}\right\}$. Since $kt = \Theta(d)$ and $t = 4k$, we have proved $s = n^{\Omega(\sqrt{d})}$. □

Remark 30. *Note that we used multilinearity only in the proof of Lemma 28. Even there, mutilinearity was not strictly necessary. We only needed that the $\Sigma\Pi\Sigma\Pi$ formula C has the property that all the Π gates on layer 1, just above the input variables, are multilinear.*

7 Lower bounds for $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formulas and regular formulas

In this section, we derive our lower bounds for some flavors of $\Sigma\Pi\Sigma\Pi$ formulas. We start with a specific case that has been the focus of a few recent results ([GKKS13, KSS13]), the $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ model, where the Π gates at layers 3 and 1 have fan-ins bounded by D and t respectively.

Corollary 31. *Let $n, d, D, t \in \mathbb{N}$ be such that $1 \leqslant t \leqslant d/320$ and $n \geqslant 10$. Then, any $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ for $\mathrm{IMM}_{n,d}$ has top fan-in at least $\left(\frac{n^{3/4}}{4D}\right)^{\Omega(d/t)}$.*

Proof. Fix $k = \lfloor d/320t \rfloor$. We will show that any $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formula C for $\mathrm{IMM}_{n,d}$ has top fan-in at least $\Omega\left(\left(\frac{n^{3/4}}{4D}\right)^k\right)$, which will prove the corollary.

But this follows from Theorem 12 and the simple fact that if $\mathrm{IMM}_{n,d}$ has a $\Sigma\Pi^{(D)}\Sigma\Pi^{(t)}$ formula with top fan-in at most s, then so does any of its restrictions, and in particular F does. $\square$

For the next result of this section we need a few additional definitions. The *degree* of any node is defined to be the degree of the polynomial computed by it. The degree of the circuit (formula) is the degree of the output node. The *syntactic degree* is defined inductively. The syntactic degree of an input node is 1. The syntactic degree of $+$ gate is the maximum of the syntactic degrees of its children. The syntactic degree of $\times$ gate is the sum of the syntactic degrees of its children. The syntactic degree of the circuit (formula) is the syntactic degree of its output node.

A formula is called *regular* if it is a layered formula, the alternate layers in the formula are labeled by $+$ and $\times$, for every layer the fan-in of all the gates at that layer is the same, and the syntactic degree of the formula is at most twice the degree of the formula.

Regular formulas were defined and studied recently by Kayal, Saha, and Saptharishi [KSS13]. They show the existence of a certain polynomial in VNP of degree d over N variables that has no regular formula of size less than $N^{\Omega(\log d)}$.

They also explicitly ask the following: is it true that any degree d polynomial in N variables that has a polynomial-sized ABP also has a regular formula of size $N^{o(\log d)}$? Here, we answer this question in the negative for $d \leqslant n^{1/10}$ by showing that $\mathrm{IMM}_{n,d}$ has no regular formulas of size less than $n^{\Omega(\log d)}$. We will need the following theorem of [KSS13].

Theorem 32 ([KSS13], Theorem 15)**.** *Let X be any set of N variables and let $F \in \mathbb{F}[X]$ be a polynomial of degree d with the property that there exists a $\delta > 0$ such that for any $t < d/100$, any $\Sigma\Pi^{(O(d/t))}\Sigma\Pi^{(t)}$ formula computing the polynomial F has top fan-in at least $\exp(\delta\left(\frac{d}{t}\right)\log N)$. Then, any regular formula computing F must be of size $N^{\Omega(\log d)}$.*

Though the theorem above is stated for $t < d/100$, it holds for $t < d/C$ for any constant C. We leave this check to the interested reader. Putting the above theorem together with Corollary 31, we have

Corollary 33. *For large enough $n, d \in \mathbb{N}$ such that $d \leqslant n^{1/10}$, any regular formula for $\mathrm{IMM}_{n,d}$ has size at least $n^{\Omega(\log d)}$.*

Note that the above is tight, up to the constant in the exponent, since the standard construction of an $n^{O(\log d)}$ sized formula for $\mathrm{IMM}_{n,d}$ yields a regular formula.

8 Discussion

Our aims in this paper were twofold: to explore the limits of depth reduction, and to understand better the arithmetic circuit complexity of $\mathrm{IMM}_{n,d}$. We have made progress on both fronts, but many interesting questions remain unanswered.

We have shown that Tavenas' result [Tav13] is optimal up to polynomial factors, even for polynomials in the class $\mathrm{VP_S}$, by showing that any $\Sigma\Pi^{(O(\sqrt{d}))}\Sigma\Pi^{(O(\sqrt{d}))}$ formulas for $\mathrm{IMM}_{n,d}$ has size $\exp(\Omega(\sqrt{d}\log N))$. Our results also answer a question of Kayal, Saha, and Saptharishi [KSS13] regarding the simulation of polynomial-sized circuits by regular formulas. Thus, in order to use depth reduction based techniques to prove a separation between VP and VNP, we will need to exhibit a polynomial in VNP that requires $\Sigma\Pi^{(O(\sqrt{d}))}\Sigma\Pi^{(O(\sqrt{d}))}$ formulas of size $\exp(\omega(\sqrt{d}\log N))$ to compute it.

One might also wonder whether lower bounds for weaker models, such as arithmetic formulas, might follow from either the shifted partial derivative technique or by depth reduction. Can one show non-trivial upper bounds on the dimension of the shifted partial derivative space of polynomials computed by small formulas? Do polynomials of degree d over N variables computed by $\mathrm{poly}(N)$-sized formulas have $\Sigma\Pi^{(O(\sqrt{d}))}\Sigma\Pi^{(O(\sqrt{d}))}$ formulas of size $\exp(o(\sqrt{d}\log N))$? As far as we know, even the case of lower bounds for $\Sigma\Pi\Sigma\Pi$ homogeneous formulas is open.

Coming to the question of the complexity of $\mathrm{IMM}_{n,d}$, we have been able to pin down almost exactly the $\Sigma\Pi\Sigma\Pi$ complexity of $\mathrm{IMM}_{n,d}$ in the set-multilinear and more generally, in the homogeneous multilinear setting. Can we extend these results to show that, in general, that set-multilinear formulas of product-depth r (for constant r) computing $\mathrm{IMM}_{n,d}$ must have size $\exp(\Omega(d^{1/r}\log n))$?

This would count as tangible progress towards the goal of showing that set-multilinear formulas for $\mathrm{IMM}_{n,d}$ must have size $n^{\Omega(\log d)}$. Raz [Raz10] has shown that, for $d = o(\log n/\log\log n)$, the set-multilinear formula complexity of $\mathrm{IMM}_{n,d}$ and the formula complexity of $\mathrm{IMM}_{n,d}$ are polynomially related and hence, a superpolynomial lower bound for set-multilinear formulas in this regime would immediately imply a separation between $\mathrm{VP_S}$ and $\mathrm{VP_E}$.

References

[AV08] Manindra Agrawal and V. Vinay. Arithmetic circuits: A chasm at depth four. In *FOCS*, pages 67–75. IEEE Computer Society, 2008.

[CLO97] David A. Cox, John Little, and Donal O'Shea. *Ideals, varieties, and algorithms - an introduction to computational algebraic geometry and commutative algebra (2. ed.).* Undergraduate texts in mathematics. Springer, 1997.

[DMPY12] Zeev Dvir, Guillaume Malod, Sylvain Perifel, and Amir Yehudayoff. Separating multilinear branching programs and formulas. In Howard J. Karloff and Toniann Pitassi, editors, *STOC*, pages 615–624. ACM, 2012.

[DP09] Devdatt P. Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms.* Cambridge University Press, 2009.

[FSS84] Merrick L. Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, 1984.

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

[Gur10] Venkatesan Guruswami. Introduction to coding theory, Lecture 2: Gilbert-Varshamov bound. `http://www.cs.cmu.edu/~venkatg/teaching/codingtheory/`, 2010.

[Hås87] Johan Håstad. *Computational limitations of small-depth circuits.* The MIT Press, Cambridge(MA)-London, 1987.

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

[Koi12] Pascal Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theor. Comput. Sci.*, 448:56–65, 2012.

[KS13] Mrinal Kumar and Shubhangi Saraf. Lower Bounds for Depth 4 Homogenous Circuits with Bounded Top Fanin. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:68, 2013.

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

[MP08] Guillaume Malod and Natacha Portier. Characterizing valiant's algebraic complexity classes. *J. Complex.*, 24(1):16–38, 2008.

[MV97] Meena Mahajan and V. Vinay. Determinant: Combinatorics, algorithms, and complexity. *Chicago J. Theor. Comput. Sci.*, 1997, 1997.

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

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

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

[Raz10] Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. In Leonard J. Schulman, editor, *STOC*, pages 659–666. ACM, 2010.

[RY08] Ran Raz and Amir Yehudayoff. Balancing syntactically multilinear arithmetic circuits. *Computational Complexity*, 17(4):515–535, 2008.

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

[SY10] Amir Shpilka and Amir Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5(3-4):207–388, 2010.

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

[Tod92] S. Toda. Classes of Arithmetic Circuits Capturing the Complexity of Computing the Determinant. *IEICE Transactions on Information and Systems*, E75-D:116–124, 1992.

[Val79] L. G. Valiant. Completeness Classes in Algebra. In *STOC '79: Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 249–261, New York, NY, USA, 1979. ACM Press.

[Val82] L. G. Valiant. Reducibility by Algebraic Projections. In *Logic and Algorithmic: an International Symposium held in honor of Ernst Specker*, volume 30 of *Monographies de l'Enseignement Mathémathique*, pages 365–380, 1982.

Appendix

In this section, we prove a lemma which states a lower bound on the dimension of the shifted partial derivative space of $\mathrm{IMM}_{n,d}$. Though this lemma is not necessary to obtain any of our lower bounds, it may be of general interest. We first need the following intuitive statement.

Lemma 34. *Let X be a set of variables with $|X| = N$ and $f, g \in \mathbb{F}[X]$ be such that g is a restriction of f. Then, for any $k, \ell \geqslant 1$ we have $\dim(\langle \partial_k f\rangle_{\leqslant \ell}) \geqslant \dim(\langle \partial_k g\rangle_{\leqslant \ell})$.*

Proof. Given $x \in X$, let $S_x : \mathbb{F}[X] \to \mathbb{F}[X]$ be the linear map that maps any polynomial f' to the polynomial g' obtained by setting x to 0 in f'. Similarly, we use $S_\sigma : \mathbb{F}[X] \to \mathbb{F}[X]$ to denote the linear map that maps f' to the polynomial g' obtained by setting all $x \in \sigma^{-1}(0)$ to 0 in f'. Note that S_σ is the composition of all the S_x for $x \in \sigma^{-1}(0)$ (the order of composition is irrelevant).

We want to show that for any $f \in \mathbb{F}[X]$, we have $\dim(\langle \partial_k f\rangle_{\leqslant \ell}) \geqslant \dim(\langle \partial_k (S_\sigma f)\rangle_{\leqslant \ell})$. From the reasoning in the previous paragraph, it suffices to show that for any $x \in X$, we have $\dim(\langle \partial_k f\rangle_{\leqslant \ell}) \geqslant \dim(\langle \partial_k (S_x f)\rangle_{\leqslant \ell})$. We can write f as $f = \sum_{j=0}^{a} x^j f_j$ where $f_j \in \mathbb{F}[X\backslash\{x\}]$ for $j \leqslant a$. Using this notation, $S_x f$ is simply the polynomial f_0. Thus, what we want to show is that $\dim(\langle \partial_k f\rangle_{\leqslant \ell}) \geqslant \dim(\langle \partial_k (f_0)\rangle_{\leqslant \ell})$.

We introduce some useful notation here. Let the variable in X be denoted $x_1, \ldots, x_N$. For $\mathbf{i} \in \mathbb{N}^N$ such that $i_1 + \cdots + i_N = k$ and $g \in \mathbb{F}[X]$, we denote by $\mathbf{x}^{\mathbf{i}}$ the monomial $x_1^{i_1} \cdots x_n^{i_n}$ and by $\frac{\partial^k g}{\partial \mathbf{x}^{\mathbf{i}}}$ the polynomial $\partial^k g/(\partial^{i_1} x_1 \cdots \partial^{i_N} x_N)$.

Let L denote the dimension of $\langle \partial_k (f_0)\rangle_{\leqslant \ell}$. Choose an arbitrary basis B for this space. Such a basis may be written as

$$B = \left\{ m_1 \cdot \frac{\partial^k f_0}{\partial \mathbf{x}^{\mathbf{i^{(1)}}}}, m_2 \cdot \frac{\partial^k f_0}{\partial \mathbf{x}^{\mathbf{i^{(2)}}}}, \ldots, m_L \cdot \frac{\partial^k f_0}{\partial \mathbf{x}^{\mathbf{i^{(L)}}}} \right\}$$

for some monomials $m_1, \ldots, m_L$ of degree at most ℓ and $\mathbf{i^{(1)}}, \ldots, \mathbf{i^{(L)}} \in \mathbb{N}^N$ such that for each $r \in [L]$, $i_1^{(r)} + \cdots + i_N^{(r)} = k$. In particular, we must have $\frac{\partial^k f_0}{\partial \mathbf{x}^{\mathbf{i^{(r)}}}} \neq 0$ for each $r \in [L]$. As $f_0 \in \mathbb{F}[X\backslash\{x\}]$, this implies that $x \nmid \mathbf{x}^{\mathbf{i^{(r)}}}$.

We claim that the elements

$$B' = \left\{ m_1 \cdot \frac{\partial^k f}{\partial \mathbf{x}^{\mathbf{i}^{(1)}}}, m_2 \cdot \frac{\partial^k f}{\partial \mathbf{x}^{\mathbf{i}^{(2)}}}, \ldots, m_L \cdot \frac{\partial^k f}{\partial \mathbf{x}^{\mathbf{i}^{(L)}}} \right\}$$

of $\langle \partial_k f \rangle_{\leqslant \ell}$ are linearly independent. This would prove that $\dim(\langle \partial_k f \rangle_{\leqslant \ell}) \geqslant L$ and finish the proof of the lemma.

To see that the elements of B' are linearly independent, we partition the monomials m_r ($r \in [L]$) depending on the highest power of x dividing them. For $j \leqslant \ell$, let $T_j = \left\{ r \in [L] \;\middle|\; x^j | m_r \text{ but } x^{j+1} \nmid m_r \right\}$. Now, fix any non-zero linear combination of the elements of B', say

$$F = \sum_{r \in [L]} \alpha_r \cdot m_r \cdot \frac{\partial^k f}{\partial \mathbf{x}^{\mathbf{i}^{(r)}}}$$

where $\alpha_r \in \mathbb{F}$ for each $r \in [L]$. Let j be the least element of $\{0\} \cup [\ell]$ such that there is an $r \in T_j$ with $\alpha_r \neq 0$. Consider the coefficient F_j of x^j in F as a polynomial in $\mathbb{F}[X \backslash \{x\}]$. Since $x \nmid \mathbf{x}^{\mathbf{i}^{(r)}}$ for any $r \in [L]$, it can be seen that

$$x^j F_j = \sum_{r \in T_j} \alpha_r \cdot m_r \cdot \frac{\partial^k f_0}{\partial \mathbf{x}^{\mathbf{i}^{(r)}}}$$

(Notice that f has been replaced by f_0 in the equation above.) But by the linear independence of the elements in B, we have $F_j \neq 0$. Hence so is F. Thus, the elements of B' are linearly independent. $\square$

Lemma 35. *Let $k, \ell \in \mathbb{N}$ be arbitrary parameters such that $20k < d < \ell$ and $k \geqslant 2$. Then,*

$$\dim(\langle \partial_k \mathrm{IMM}_{n,d} \rangle_{\leqslant \ell}) \geqslant M \cdot \binom{N + \ell}{\ell} - M^2 \cdot \binom{N + \ell - d/10}{\ell - d/10},$$

where $M = \left\lfloor \left(\frac{n}{4}\right)^k \right\rfloor$.

Proof. Straightaway follows from Lemmas 34 and 11, since the polynomial F from Lemma 11 is a restriction of $\mathrm{IMM}_{n,d}$. $\square$

ECCC ISSN 1433-8092

http://eccc.hpi-web.de