Lower Bounds and Separations for Constant Depth Multilinear Circuits

Ran Raz * Amir Yehudayoff †

Abstract

We prove an exponential lower bound for the size of constant depth multilinear arithmetic circuits computing either the determinant or the permanent (a circuit is called multilinear, if the polynomial computed by each of its gates is multilinear). We also prove a super-polynomial separation between the size of product-depth[1] d and product-depth $d+1$ multilinear circuits (where d is constant). That is, there exists a polynomial f such that

- There exists a multilinear circuit of product-depth $d+1$ and of polynomial size computing f.
- Every multilinear circuit of product-depth d computing f has super-polynomial size.

1 Introduction

Arithmetic circuits are the standard model for computing polynomials. Proving exponential lower bounds for the size of arithmetic circuits is an outstanding open problem. So, restricted classes of arithmetic

*Faculty of Mathematics and Computer Science, Weizmann Institute, Rehovot, Israel. Email: ran.raz@weizmann.ac.il. Research supported by grants from the Binational Science Foundation (BSF), the Israel Science Foundation (ISF), and the Minerva Foundation.

†Faculty of Mathematics and Computer Science, Weizmann Institute, Rehovot, Israel. Email: amir.yehudayoff@weizmann.ac.il. Research supported by grants from the Binational Science Foundation (BSF), the Israel Science Foundation (ISF), the Minerva Foundation, and the Israel Ministry of Science (IMOS) - Eshkol Fellowship.

[1]The product-depth of a circuit is the largest number of product gates in a directed path in it. According to the standard definition of depth, we show a super-polynomial separation between the size of depth d and depth $d+2$ multilinear circuits.

circuits, such as constant depth circuits and multilinear circuits, have been studied. The study of constant depth Boolean circuits gave beautiful insights into Boolean computations; in particular, exponential lower bounds are known for constant depth Boolean circuits [A, FSS, H, R, S, Y]. However, surprisingly, our understanding of constant depth arithmetic circuits is much poorer, and no general lower bounds (better than, say, n^2) are known, even for depth 4 arithmetic circuits.

In this paper we study constant depth multilinear circuits. We prove an exponential lower bound for the size of constant depth multilinear circuits for the determinant and the permanent. We also prove a super-polynomial separation between the size of product-depth d and product-depth $d+1$ multilinear circuits. We consider the notion of product-depth rather than the standard notion of depth for the simplicity of the presentation, and since the notions of depth and product-depth are equivalent up to a factor of two.

1.1 Multilinear Circuits

An *arithmetic circuit* Φ over the field $\mathbb{F}$ and over the set of variables X is a directed acyclic graph as follows. Every vertex in Φ of in-degree 0 is labelled by either a variable in X or a field element in $\mathbb{F}$. Every other vertex in Φ is labelled by either $\times$ or $+$. An arithmetic circuit is called a *formula* if it is a directed tree (whose edges are directed from the leaves to the root).

The vertices of Φ are also called *gates*. Every gate of in-degree 0 is called an *input gate*. Every gate of out-degree 0 is called an *output gate*. Every gate labelled by $\times$ is called a *product gate*. Every gate labelled by $+$ is called a *sum gate*. For two gates u and v in Φ, if (u,v) is an edge in Φ, then u is called a *child* of v, and v is called a *parent* of u. We denote by $\textsc{child}(v)$ the set of children of v. The *size* of Φ, denoted $|\Phi|$, is the number of edges in Φ. The product-depth of a gate v in Φ, denoted $\mathrm{p-depth}(v)$, is the largest number of product gates in a directed path reaching v. The product-depth of Φ is the maximal product-depth of a gate in Φ.

For a gate v in Φ, define Φ_v to be the sub-circuit of Φ rooted at v. Denote by X_v the set of variables that occur in Φ_v. An arithmetic circuit computes a polynomial in a natural way. An input gate labelled by $\alpha \in \mathbb{F} \cup X$ computes the polynomial α. A product gate computes the product of the polynomials computed by its children. A sum gate computes the sum of the polynomials computed by its children. We denote by $\widehat{\Phi}_v$ the polynomial in $\mathbb{F}[X_v]$ computed by the gate v in Φ.

A polynomial $f \in \mathbb{F}[X]$ is called *multilinear* if the degree of each variable in f is at most one. An arithmetic circuit Φ is called *multilinear* if every gate in Φ computes a multilinear polynomial. An

arithmetic circuit Φ is called *syntactically multilinear* if for every product gate in Φ and for every two gates $v_1, v_2 \in \text{CHILD}(v)$, the two sets X_{v_1} and X_{v_2} are disjoint.

1.2 Background

The study of multilinear circuits was initiated by Nisan and Wigderson in [NW96]. A super-polynomial lower bound for the size of multilinear formulas for the determinant and the permanent was proved in [R04A]. Then, [R04B] proved a super-polynomial separation between the size of multilinear formulas and the size of multilinear circuits ([RY] simplified the proof of this separation). Later, [RSY] proved a roughly $n^{4/3}$ lower bound for the size of syntactically multilinear arithmetic circuits.

Constant depth arithmetic circuits have been studied extensively. Over finite fields [GK, GR] proved an exponential lower bound for the size of depth 3 circuits. Over fields of characteristic zero [SW] proved a roughly n^2 lower bound for the size of depth 3 circuits. For arithmetic circuits of arbitrary constant depth [SS, R07] proved a roughly $n^{1+1/d}$ lower bound for the size of depth d arithmetic circuits (over arbitrary fields).

In this paper we study circuits that are both multilinear and of constant depth – in fact, our results hold for non-constant (but bounded) depth as well. We improve over the lower bounds known for each model separately. We also give a super-polynomial separation between product-depth d and product-depth $d+1$ multilinear circuits.

1.3 Methods

We use ideas that where used before to prove lower bounds for multilinear circuits and formulas [R04A, R04B, RSY]. The proof of the separation between multilinear formula and circuit size in [R04B] implies that these ideas fail to prove a super-polynomial lower bound for the size of multilinear circuits. In this paper, we use the properties of constant depth circuits, together with ideas from [R04A, R04B, RSY], to prove an exponential lower bound for the size of constant depth multilinear circuits. In addition, we give a construction of a multilinear polynomial that is computable by a polynomial size multilinear circuit of product-depth $d+1$, and is not computable by a polynomial size multilinear circuit of product-depth d.

1.4 Results

The following theorem gives an exponential lower bound for the size of constant depth multilinear circuits computing either the determinant or the permanent (it also gives a super-polynomial lower bound for non-constant but bounded depth).

Theorem 1.1. *Let $n \in \mathbb{N}$, and let Φ be a multilinear arithmetic circuit of product-depth d for[2] $d = o(\log n / \log\log n)$ over the field $\mathbb{F}$ and over the set of variables $X = \{x_{i,j}\}_{i,j\in[n]}$ computing either the determinant of X or the permanent of X. Then,*

$$|\Phi| \geq 2^{n^{\Omega(1/d)}}.$$

The following theorem gives a super-polynomial separation between the size of product-depth d and product-depth $d+1$ multilinear circuits (the separation remains super-polynomial even for non-constant but bounded d). We note that as a part of the proof of the separation between product-depth d and product-depth $d+1$ multilinear circuits, we prove a tight lower bound for the size of product-depth d multilinear circuits.

Theorem 1.2. *Let $d \in \mathbb{N}$ be a constant, and let $f = f_{d+1}$ be the polynomial in $\mathbb{F}[X, W]$ defined in Section 6.1, and denote $|X| + |W| = N$. Then,*

- *There exists a syntactically multilinear arithmetic formula of product-depth $d+1$ and size poly(N) over the field $\mathbb{F}$ and over the set of variables $X \cup W$ computing f.*
- *Every multilinear arithmetic circuit of product-depth d over the field $\mathbb{F}$ and over the set of variables $X \cup W$ computing f is of super-polynomial size, $N^{\Omega(\log^{1/(2d)}(N))}$.*

2 Preliminaries

For an integer $n \in \mathbb{N}$, denote $[n] = \{1, \ldots, n\}$.

[2]Unless stated otherwise logarithms are of base 2.

2.1 Constant Depth Multilinear Circuits

We will now define *d-normal-form* formulas (where $d \in \mathbb{N}$). A d-normal-form formula has the form

$$\underbrace{\sum\prod\sum\cdots\sum\prod\sum}_{2d+1 \text{ times}}.$$

Formally, the definition is inductive: a 0-normal-form formula is a sum of input gates, and for $d > 0$, a d-normal-form formula rooted in the gate v has the form

$$\sum_{u \in \text{CHILD}(v)} \prod_{u' \in \text{CHILD}(u)} \Phi_{u'},$$

where $\Phi_{u'}$ is a $(d-1)$-normal-form formula.

It will be convenient for us to think of a constant depth circuit as a normal-form formula. The following lemma shows that this view is not misleading.

Lemma 2.1. *Let Φ be a multilinear arithmetic circuit of product-depth d over the field $\mathbb{F}$ and over the set of variables X computing the polynomial f. Then, there exists a d-normal-form syntactically multilinear arithmetic formula of size at most $(d+1)^2 |\Phi|^{2d+1}$ over the field $\mathbb{F}$ and over the set of variables X computing f as well.*

We prove the lemma in Section 2.1.4 below, using the following three claims.

2.1.1 Constant Depth Circuits Are Formulas

The following claim shows that constant depth circuits and formulas are equivalent.

Claim 2.2. *Let Φ be a multilinear arithmetic circuit of product-depth d over the field $\mathbb{F}$ and over the set of variables X computing the polynomial f. Then, there exists a multilinear arithmetic formula of product-depth d and of size at most $|\Phi|^{2d+1}$ over the field $\mathbb{F}$ and over the set of variables X computing f as well.*

Proof. The proof follows by induction on d. If $d = 0$, then without loss of generality Φ is a formula, and the claim follows. Assume $d > 0$. Let v be the gate in Φ computing f. Assume without loss of generality that

$$\Phi = \sum_{u \in \text{CHILD}(v)} \prod_{u' \in \text{CHILD}(u)} \Phi_{u'}$$

(if v is not a sum gate, or if some of the gates $u \in \text{CHILD}(v)$ are input gates, we add some 'dummy' gates). For every gate u' (as above), by induction, there exists a multilinear formula $\Psi_{u'}$ of product-depth at most $d-1$ and size at most $|\Phi_{u'}|^{2(d-1)+1}$ computing $\widehat{\Phi}_{u'}$. Let Ψ be the formula $\sum_{u \in \text{CHILD}(v)} \prod_{u' \in \text{CHILD}(u)} \Psi_{u'}$. Thus,

$$|\Psi| \leq (|\Phi|-1)^2(|\Phi|-1)^{2d-1} + |\Phi| \leq |\Phi|^{2d+1}.$$

In addition, Ψ is multilinear of product-depth d and computes f. □

2.1.2 Multilinear Formulas Are Syntactically Mulitlinear

Every syntactically multilinear arithmetic formula is also multilinear. It is shown in [R04A] that the other direction holds as well. Formally,

Claim 2.3. *For every multilinear arithmetic formula, there exists a syntactically multilinear arithmetic formula of the same size and product-depth computing the same polynomial.*

For completeness we give the main idea of the proof. Let Φ be an arithmetic formula over the set of variables X. Let v be a product gate in Φ with two children v_1 and v_2 such that the polynomial $\widehat{\Phi}_v$ is multilinear. Assume that $x \in X_{v_1} \cap X_{v_2}$. Then, without loss of generality the degree of x in the polynomial $\widehat{\Phi}_{v_1}$ is 0. Since Φ is a formula, after substituting 0 instead of x in Φ_{v_1}, both Φ_{v_1} and Φ_{v_2} still compute the same polynomials.

2.1.3 Formulas Have Normal-Form

Roughly, the following claim shows that a syntactically multilinear formula has normal-form.

Claim 2.4. *Let Φ be a syntactically multilinear arithmetic formula of product-depth d over the field $\mathbb{F}$ and over the set of variables X computing the polynomial f. Then, there exists a d-normal-form syntactically multilinear arithmetic formula Ψ of size at most $(d+1)^2|\Phi|$ over the field $\mathbb{F}$ and over the set of variables X computing the polynomial f as well.*

Proof. The proof follows by induction on the product-depth of Φ. If the product-depth of Φ is 0, then without loss of generality Φ is of 0-normal-form. Otherwise, assume without loss of generality that

$$\Phi = \sum_{u \in \text{CHILD}(v)} \prod_{u' \in \text{CHILD}(u)} \Phi_{u'},$$

where v is the gate in Φ computing f (we note that v might not be a sum gate, and that some of the gates $u \in \text{CHILD}(v)$ might be input gates; in these cases we add 'dummy' gates to the circuit). By induction, for every u' there exists a $(d-1)$-normal form syntactically multilinear arithmetic formula $\Psi_{u'}$ of size at most $d^2 |\Phi_{u'}| + 2d - 1$ over the set of variables $X_{u'}$ computing the polynomial $\widehat{\Phi}_{u'}$ (we may need to add at most $2d-1$ 'dummy' gates, if the product-depth of u' is less than $d-1$). Set

$$\Psi = \sum_{u \in \text{CHILD}(v)} \prod_{u' \in \text{CHILD}(u)} \Psi_{u'}.$$

Since Φ is syntactically multilinear, so is Ψ. In addition,

$$|\Psi| \leq \sum_{u \in \text{CHILD}(v)} \sum_{u' \in \text{CHILD}(u)} |\Psi_{u'}| + |\Phi| + 1 \leq (d+1)^2 |\Phi| .$$

□

2.1.4 Proof of Lemma 2.1

Let Φ be a multilinear arithmetic circuit of product-depth d over the field $\mathbb{F}$ and over the set of variables X computing the polynomial f. By Claim 2.2, Claim 2.3 and Claim 2.4, there exists a d-normal-form syntactically multilinear arithmetic formula of size at most $(d+1)^2 |\Phi|^{2d+1}$ over the field $\mathbb{F}$ and over the set of variables X computing f as well. □

2.2 Partitions

Let X, Y, Z be three sets of variables. A *partition* A of X is a map from X to $Y \cup Z \cup \{0,1\}$ such that every $t \in Y \cup Z$ admits $|A^{-1}(t)| = 1$. For a distribution on partitions μ, we will denote by $A \sim \mu$ a partition distributed according to μ.

We will focus on two types of partitions. The first type (used for the lower bounds for the permanent and the determinant) will be of maps from an $n \times n$ matrix of variables X to $Y \cup Z \cup \{0,1\}$, where the size of both Y and Z is roughly $n^{1/3}$. The second type (used for the separation) will be of one-to-one maps from X to $Y \cup Z$, where Y and Z are two sets of equal size.

For a set $X' \subseteq X$ and a partition A, we denote

$$\mathcal{D}_A(X') = \left|A^{-1}(Y) \cap X'\right| - \left|A^{-1}(Z) \cap X'\right| .$$

Thus, $\left|\mathcal{D}_A(\cdot)\right|$ measures the amount of 'unbalanceness' of a set according to A.

Let Φ be an arithmetic circuit over the set of variables X. Given a partition A, we can define a new arithmetic circuit Φ^A, which is the same as Φ except that each variables $x \in X$ is substituted by $A(x)$. Thus, Φ^A is an arithmetic circuit over the set of variables $Y \cup Z$, and if Φ is syntactically multilinear, then so is Φ^A. In addition, there is a one-to-one correspondence between gates in Φ and gates in Φ^A. Namely, for every gate v in Φ there is a corresponding gate in Φ^A and vice versa. Thus, for a gate v in Φ, we will denote by X_v the set of X variables that occur in Φ_v, by Y_v the set of Y variables that occur in Φ_v^A and by Z_v the set of Z variables that occur in Φ_v^A.

Similarly, given a polynomial f in $\mathbb{F}[X]$, we denote by $f^A \in \mathbb{F}[Y,Z]$ the polynomial f after substituting every $x \in X$ by $A(x) \in Y \cup Z \cup \{0,1\}$. Thus, if f is multilinear, then so is f^A.

In the rest of the paper we will consider circuits and polynomials both over the set of variables X and over the two sets of variables Y and Z. One can think of the variables Y and Z as a renaming of some of the variables in X.

2.3 The Partial Derivative Matrix

Let $\mathbb{F}$ be a field and let Y and Z be two sets of variables of size $m \in \mathbb{N}$ each. Let f be a multilinear polynomial in $\mathbb{F}[Y,Z]$. Denote by M_f the following $2^m \times 2^m$ matrix with entries in $\mathbb{F}$: for a multilinear monomial p in the set of variables Y and a multilinear monomial q in the set of variables Z, the (p,q) entry in M_f is the coefficient of the monomial $p \cdot q$ in the polynomial f. The matrix M_f is called the *partial derivative matrix* of f. Of special interest will be polynomials whose partial derivative matrix has full rank. Such polynomials are said to have *full rank*.

For a gate v in a multilinear circuit Ψ over the field $\mathbb{F}$ and over the two sets of variables Y and Z, denote by M_v the partial derivative matrix of the polynomial $\widehat{\Psi}_v$.

The following proposition gives some useful properties of the partial derivative matrix.

Proposition 2.5. *Let $\mathbb{F}$ be a field and let Y and Z be two sets of variables of size $m \in \mathbb{N}$ each. Let Y_1 and Y_2 be two subsets of Y. Let Z_1 and Z_2 be two subsets of Z. Let f_1 be a multilinear polynomial in $\mathbb{F}[Y_1,Z_1]$ and let f_2 be a multilinear polynomial in $\mathbb{F}[Y_2,Z_2]$. Then,*

1. $$Rank(M_{f_1}) \leq 2^{\min(|Y_1|,|Z_1|)}.$$

2.

$$Rank(M_{f_1+f_2}) \leq Rank(M_{f_1}) + Rank(M_{f_2}).$$

3. If $Y_1 \cap Y_2 = \emptyset$ and $Z_1 \cap Z_2 = \emptyset$, then

$$Rank(M_{f_1 \cdot f_2}) = Rank(M_{f_1}) \cdot Rank(M_{f_2}).$$

Proof.

1. M_{f_1} has at most $2^{|Y_1|}$ nonzero rows and at most $2^{|Z_1|}$ nonzero columns.
2. Note that $M_{f_1+f_2} = M_{f_1} + M_{f_2}$ and that $\text{Rank}(M_{f_1} + M_{f_2}) \leq \text{Rank}(M_{f_1}) + \text{Rank}(M_{f_2})$.
3. Let M be the submatrix of $M_{f_1 \cdot f_2}$ of nonzero rows and columns, let M_1 be the submatrix of M_{f_1} of nonzero rows and columns and let M_2 be the submatrix of M_{f_2} of nonzero rows and columns. Since $Y_1 \cap Y_2 = \emptyset$ and $Z_1 \cap Z_2 = \emptyset$, M is the tensor product of M_1 and M_2. Finally, $\text{Rank}(M_1 \otimes M_2) = \text{Rank}(M_1) \cdot \text{Rank}(M_2)$, where $\otimes$ denotes tensor product.

□

3 Central Paths and Weak Formulas

Let Ψ be a d-normal-form syntactically multilinear arithmetic formula over the two sets of variables Y and Z (that are of size $m \in \mathbb{N}$ each). For a gate v in Ψ, denote by Y_v the set of Y variables that occur in Ψ_v, denote by Z_v the set of Z variables that occur in Ψ_v, and denote

$$\text{avg}(v) = \frac{|Y_v| + |Z_v|}{2}.$$

For $k \in \mathbb{N}$, a simple directed path γ in Ψ is called *k-central*, if γ starts at an input gate, and every edge (u, v) in γ admits the following:

· if v is a product gate, every gate $u' \in \text{CHILD}(v)$ admits $\text{avg}(u') \leq \text{avg}(u)$,

· if v is a sum gate, $\text{avg}(u) \geq \text{avg}(v) - k/2$.

Roughly, a central path goes to a child with many variables. We note that not every gate has a central path reaching it.

For $k \in \mathbb{N}$, we say that a gate v in Ψ has *k-low-rank* if

$$\mathrm{Rank}(M_v) \leq 2^{\mathrm{avg}(v)-k}.$$

We say that a gate v in Ψ is k-weak, if every k-central path that reaches v contains a k-low-rank gate. We say that a formula is k-weak if its output gate is k-weak.

The following theorem shows that weak formulas do not compute full rank polynomials.

Theorem 3.1. *Let Φ be a d-normal-form syntactically multilinear arithmetic formula over the field $\mathbb{F}$ and over the two sets of variables Y and Z, and let v be a gate in Φ. For every $k \in \mathbb{N}$, if v is k-weak, then*

$$\mathit{Rank}(M_v) \leq |\Phi_v| \, 2^{avg(v)-k/2}.$$

Proof. The proof follows by induction on the size of Φ_v. Consider the following three cases:

Case one: v has k-low-rank. Since $|\Phi_v| \geq 1$,

$$\mathrm{Rank}(M_v) \leq |\Phi_v| \, 2^{\mathrm{avg}(v)-k/2}.$$

For the rest of the proof we assume that v does not have k-low-rank.

Case two: v is a sum gate. Partition the children of v to two sets

$$C_1 = \{u \in \textsc{child}(v) \ : \ \mathrm{avg}(u) \geq \mathrm{avg}(v) - k/2\} \quad \text{and} \quad C_2 = \textsc{child}(v) \setminus C_1.$$

Since v does not have k-low-rank, and since every k-central path reaching a gate in C_1 can be extended to a k-central path reaching v, every gate $u \in C_1$ is k-weak. Thus, by induction, every $u \in C_1$ admits

$$\mathrm{Rank}(M_u) \leq |\Phi_u| \, 2^{\mathrm{avg}(v)-k/2}$$

(since $\mathrm{avg}(u) \leq \mathrm{avg}(v)$). In addition, for every gate $u \in C_2$,

$$\min(|Y_u|, |Z_u|) \leq \mathrm{avg}(u) < \mathrm{avg}(v) - k/2.$$

Thus, by property 1. of Proposition 2.5,

$$\mathrm{Rank}(M_u) \leq 2^{\mathrm{avg}(v)-k/2}.$$

Thus, since $|\Phi_v| \geq \sum_{u \in C_1} |\Phi_u| + |C_2|$, using property 2. of Proposition 2.5,

$$\text{Rank}(M_v) \leq \sum_{u \in \text{CHILD}(v)} \text{Rank}(M_u) \leq |\Phi_v| \, 2^{\text{avg}(v) - k/2}.$$

Case three: v is a product gate. Since v does not have k-low-rank, there is a gate $v' \in \text{CHILD}(v)$ that is k-weak. Thus, by induction,

$$\text{Rank}(M_{v'}) \leq |\Phi_{v'}| \, 2^{\text{avg}(v') - k/2}.$$

Since Φ is syntactically multilinear,

$$\text{avg}(v) = \sum_{u \in \text{CHILD}(v)} \text{avg}(u),$$

and, by property 3. of Proposition 2.5,

$$\text{Rank}(M_v) = \prod_{u \in \text{CHILD}(v)} \text{Rank}(M_u),$$

Thus, since $|\Phi_v| \geq |\Phi_{v'}|$, using property 1. of Proposition 2.5,

$$\text{Rank}(M_v) \leq |\Phi_{v'}| \, 2^{\text{avg}(v') - k/2} \cdot 2^{\sum_{u \in \text{CHILD}(v) : u \neq v'} \text{avg}(u)} \leq |\Phi_{v'}| \, 2^{\text{avg}(v) - k/2} \leq |\Phi_v| \, 2^{\text{avg}(v) - k/2}$$

(since $\min(|Y_u|, |Z_u|) \leq \text{avg}(u)$). □

4 Lower Bounds for Permanent and Determinant

In this section we will prove Theorem 1.1 – every constant depth multilinear circuit computing either the permanent or the determinant has exponential size.

4.1 A Distribution on Partitions

Let $n \in \mathbb{N}$, and let $X = \{x_{i,j}\}_{i,j \in [n]}$ be a set of variables. Let $Y = \{y_1, \ldots, y_m\}$ and $Z = \{z_1, \ldots, z_m\}$ be two sets of variables of size

$$m = \lfloor n^{1/3} \rfloor$$

each. Recall that a partition A is a map from X to $Y \cup Z \cup \{0, 1\}$ such that $|A^{-1}(t)| = 1$, for every $t \in Y \cup Z$. We will now define a distribution μ on partitions A.

A partition A is distributed according to μ if it is chosen by the following random process. Let

$$R_1 = \{r_1(1), \ldots, r_1(m)\} \text{ and } R_2 = \{r_2(1), \ldots, r_2(m)\}$$

be two disjoint subsets of $[n]$ of size m chosen uniformly at random, and let

$$C_1 = \{c_1(1), \ldots, c_1(m)\} \text{ and } C_2 = \{c_2(1), \ldots, c_2(m)\}$$

be two disjoint subsets of $[n]$ of size m chosen uniformly at random and independently of R_1 and R_2. We note that R_1 and R_2 correspond to random rows of the matrix X, and that C_1 and C_2 correspond to random columns of the matrix X. For every $i \in [m]$, with probability one half set

$$\begin{pmatrix} x_{r_1(i),c_1(i)} & x_{r_1(i),c_2(i)} \\ x_{r_2(i),c_1(i)} & x_{r_2(i),c_2(i)} \end{pmatrix} \to \begin{pmatrix} y_i & z_i \\ 1 & 1 \end{pmatrix}$$

(that is, $A(x_{r_1(i),c_1(i)}) = y_i$, $A(x_{r_1(i),c_2(i)}) = z_i$ and $A(x_{r_2(i),c_1(i)}) = A(x_{r_2(i),c_2(i)}) = 1$), and with probability one half set

$$\begin{pmatrix} x_{r_1(i),c_1(i)} & x_{r_1(i),c_2(i)} \\ x_{r_2(i),c_1(i)} & x_{r_2(i),c_2(i)} \end{pmatrix} \to \begin{pmatrix} y_i & 1 \\ z_i & 1 \end{pmatrix}$$

(that is, $A(x_{r_1(i),c_1(i)}) = y_i$, $A(x_{r_2(i),c_1(i)}) = z_i$ and $A(x_{r_1(i),c_2(i)}) = A(x_{r_2(i),c_2(i)}) = 1$). Denote

$$\{j_1 < \cdots < j_{n-2m}\} = [n] \setminus (R_1 \cup R_2)$$

and

$$\{\ell_1 < \cdots < \ell_{n-2m}\} = [n] \setminus (C_1 \cup C_2).$$

For every $i \in [n-2m]$, set $A(x_{j_i,\ell_i}) = 1$. For every variable $x_{i,j} \in X$ that A is not already defined on, set $A(x_{i,j}) = 0$.

Remark 4.1. *Up to a permutation, the matrix X after the substitution A looks like*

$$\begin{pmatrix} B_1 & 0 & \ldots & & 0 & 0 & & \ldots & & 0 \\ & & \ldots & & & & & \ldots & & \\ 0 & & \ldots & 0 & B_m & 0 & & \ldots & & 0 \\ 0 & & \ldots & & 0 & 1 & 0 & \ldots & & 0 \\ & & \ldots & & & & & \ldots & & \\ 0 & & \ldots & & 0 & 0 & & \ldots & 0 & 1 \end{pmatrix},$$

where B_i is a 2×2 matrix that is either

$$\begin{pmatrix} y_i & z_i \\ 1 & 1 \end{pmatrix} \quad or \quad \begin{pmatrix} y_i & 1 \\ z_i & 1 \end{pmatrix}.$$

Therefore,

$$per(X)^A = \prod_{i \in [m]} (y_i + z_i) \quad and \quad det(X)^A = \pm \prod_{i \in [m]} (y_i - z_i),$$

for every A in the support of μ (where per *is the permanent and* det *is the determinant).*

4.2 Small Formulas Can Be Made Weak

The following theorem shows that constant depth formulas are weak for a random partition $A \sim \mu$.

Theorem 4.2. *Let Φ be a d-normal-form syntactically multilinear arithmetic formula over the field $\mathbb{F}$ and over the set of variables $X = \{x_{i,j}\}_{i,j \in [n]}$ such that all the variables occur in Φ. Then,*

$$\mathbb{P}[\Phi^A \text{ is not } k\text{-weak}] \leq |\Phi| \, 2^{-n^{\Omega(1/d)}} + o(1),$$

for $k = \lfloor n^{\frac{2}{27d}}/2 \rfloor$, where $A \sim \mu$.

Using the theorem above we now prove the lower bound for the determinant and the permanent.

4.3 Proof of Theorem 1.1

Let Φ be a multilinear arithmetic circuit of product-depth d over the field $\mathbb{F}$ and over the set of variables $X = \{x_{i,j}\}_{i,j \in [n]}$ computing the determinant of X. By Lemma 2.1, there exists a d-normal-form syntactically multilinear arithmetic formula Ψ of size at most $(d+1)^2 \, |\Phi|^{2d+1}$ computing the determinant of X. Assume towards a contradiction that the size of Ψ is at most

$$2^{n^{\varepsilon/d}},$$

for some small enough constant $\varepsilon > 0$. Since Ψ computes the determinant, all the variables occur in Ψ. By Theorem 4.2, there exists a partition A in the support of μ such that Ψ^A is k-weak for $k = \lfloor n^{\frac{2}{27d}}/2 \rfloor$. By Theorem 3.1, since Ψ computes the determinant,

$$\text{Rank}(M_{\det(X)^A}) \leq |\Psi| \, 2^{m-k/2} < 2^m.$$

However, using property 3. of Proposition 2.5, and using Remark 4.1 above,

$$\mathrm{Rank}(M_{\det(X)^A}) = \prod_{i\in[m]} \mathrm{Rank}(M_{y_i - z_i}) = 2^m,$$

which is a contradiction. Thus,

$$|\Phi| \geq (d+1)^{-2/(2d+1)} \, |\Psi|^{1/(2d+1)} \geq 2^{n^{\Omega(1/d)}}$$

(since $d = o(\log n / \log\log n)$). A similar argument shows that if Φ computes the permanent, then $|\Phi| \geq 2^{n^{\Omega(1/d)}}$. □

For the rest of this section we prove Theorem 4.2. The proof of the theorem will be roughly as follows. Choose a random partition $A \sim \mu$. Using Claim 4.5, every central path in Φ^A contains an ineffectual gate. We will roughly show that every ineffectual gate has low-rank with probability $1 - 2^{-n^{1/d}}$. Since the number of central paths is at most the size of the formula, the theorem will follow using a union bound.

The actual proof is more complicated, and so we need some more definitions, and a few lemmas.

4.4 Definition of W

For a partition A and for $i \in [m]$, denote

$$W_i = A^{-1}(\{y_i, z_i\}),$$

and denote

$$W = \bigcup_{i\in[m]} W_i = A^{-1}(Y \cup Z)$$

(for simplicity, we omit the dependency on A from the notation).

The set $W \subseteq X$ has a special structure: Recall that we think of X as a matrix of variables. Every variable $x_{i,j} \in W$ has a unique variable $x_{i',j'} \neq x_{i,j}$ in W that is either in the same row or in the same column as $x_{i,j}$. The set $\{x_{i,j}, x_{i',j'}\}$ is one of the sets $W_1, \ldots, W_m$. The sets $W_1, \ldots, W_m$ form a partition of W to m disjoint pairs.

For a set $\tilde{X} \subseteq X$, denote

$$W_i(\tilde{X}) = W_i \cap \tilde{X} \quad \text{and} \quad W(\tilde{X}) = W \cap \tilde{X},$$

and denote

$$W'(\tilde{X}) = \bigcup_{i\in[m]:\left|W_i(\tilde{X})\right|=1} W_i(\tilde{X}).$$

The set $W'(\tilde{X})$ is the set of all $x_{i,j} \in W(\tilde{X})$ such that no other variable in $W(\tilde{X})$ is in the same row or column as $x_{i,j}$.

For a gate v in Φ, denote

$$W(v) = W(X_v).$$

4.5 Ineffectual Gates and Good W's

For $\ell, \tau \in \mathbb{N}$, a gate v in Φ^A is called (ℓ,τ)-*ineffectual*, if v is a product gate such that $|W(v)| \geq \ell\tau$, and every gate $u \in \textsc{child}(v)$ admits $|W(u)| \leq |W(v)|/\tau$. Roughly, an ineffectual gate is a gate with many children each with few variables.

We define good W's. We say that W as above is *good* for Φ, if for every (ℓ,τ)-ineffectual gate v in Φ^A for $\ell = \lfloor n^{\frac{2}{9}} \rfloor$ and $\tau = \lfloor n^{\frac{2}{27d}}/2 \rfloor$, there exist an integer $\ell' \geq \ell$, and a family of pairwise disjoint sets $C_1, \ldots, C_{\lfloor \tau/2 \rfloor} \subseteq \textsc{child}(v)$ such that for every $i \in \{1, \ldots, \lfloor \tau/2 \rfloor\}$,

$$\ell' \leq \left|W(i)\right| \leq 2\ell' \text{ and } \left|W'(i)\right| \geq \frac{\ell'}{48},$$

where $W(i) = W\big(\bigcup_{u\in C_i} X_u\big)$ and $W'(i) = W'\big(\bigcup_{u\in C_i} X_u\big)$ for every $i \in \{1, \ldots, \lfloor \tau/2 \rfloor\}$. The condition that W is good will help us to show that an ineffectual gate has low-rank with high probability.

4.6 Proof of Theorem 4.2

The following proposition states that small formulas have good W.

Proposition 4.3. *Let Φ be a d-normal-form syntactically multilinear arithmetic formula over the set of variables $X = \{x_{i,j}\}_{i,j\in[n]}$ such that all the variables occur in Φ. Then,*

$$\mathbb{P}[W \textit{ is not good for } \Phi] \leq |\Phi| \, 2^{-n^{\Omega(1)}} + o(1),$$

where $W = A^{-1}(Y \cup Z)$ and $A \sim \mu$.

The following proposition states that small formulas with good W can be made weak. We will use the following notation. For a set $W \subseteq X$ that is good for Φ, denote by $\mu(W)$ the distribution on partitions $A \sim \mu$ conditioned on $A^{-1}(Y \cup Z) = W$.

Proposition 4.4. *Let Φ be a d-normal-form syntactically multilinear arithmetic formula over the set of variables $X = \{x_{i,j}\}_{i,j\in[n]}$ such that all the variables occur in Φ. Assume that $W \subseteq X$ is good for Φ. Then,*

$$\mathbb{P}[\Phi^A \text{ is not } k\text{-weak}] \leq |\Phi| \, 2^{-n^{\Omega(1/d)}}$$

for $k = \lfloor n^{\frac{2}{27d}}/2 \rfloor$, where $A \sim \mu(W)$.

We will now prove Theorem 4.2. Let $A \sim \mu$, let $W = A^{-1}(Y \cup Z)$, and let $k = \lfloor n^{\frac{2}{27d}}/2 \rfloor$. By the two proposition above,

$$\begin{aligned}
\mathbb{P}[\Phi^A \text{ is not } k\text{-weak}] &= \mathbb{P}[\Phi^A \text{ is not } k\text{-weak} | W \text{ is good for } \Phi] \cdot \mathbb{P}[W \text{ is good for } \Phi] \\
&\quad + \mathbb{P}[\Phi^A \text{ is not } k\text{-weak} | W \text{ is not good for } \Phi] \cdot \mathbb{P}[W \text{ is not good for } \Phi] \\
&\leq |\Phi| \, 2^{-n^{\Omega(1/d)}} + |\Phi| \, 2^{-n^{\Omega(1)}} + o(1) \\
&\leq |\Phi| \, 2^{-n^{\Omega(1/d)}} + o(1).
\end{aligned}$$

□

For the rest of this section we will prove Proposition 4.3 and Proposition 4.4. We will first prove Proposition 4.4.

4.7 Proof of Proposition 4.4

Before proving the proposition we need the following property of central paths.

4.7.1 Central Paths and Ineffectual Gates

The following claim shows that a central path contains an ineffectual gate.

Claim 4.5. *Let $d, \ell, \tau \in \mathbb{N}$ be such that $\tau \geq 2$. Let $A \in \text{supp}(\mu)$, and let $\Psi = \Phi^A$. Let v be a gate in Ψ such that Ψ_v is of d-normal-form and $|W(v)| > \ell \cdot (2\tau)^d$. Then, every k-central path for $0 \leq k \leq \ell$ from an input gate to v contains an (ℓ, τ)-ineffectual gate.*

Proof. The claim follows by induction on d.

$d = 0$ Since $|W(v)| > \ell$, and since for every input gate $|W(v)| \leq 1$, there are no k-central paths reaching v. The claim follows (in an empty sense).

$d > 0$ If there are no k-central paths reaching v, the claim follows. Otherwise, let γ be a k-central path reaching v, and let u be the child of v in γ. Since Φ_v has d-normal-from, u is a product gate. Since γ is k-central, $|W(u)| \geq |W(v)| - k$. Let w be the child of u in γ. If $|W(w)| \leq |W(u)|/\tau$, since γ is k-central, the same holds for every child of u. Thus, since $|W(u)| \geq \ell\tau$, u is (ℓ, τ)-ineffectual. Otherwise, $|W(w)| \geq |W(u)|/\tau > (\ell \cdot (2\tau)^d - k)/\tau \geq \ell \cdot (2\tau)^{d-1}$. Now, by induction, every k-central path from an input gate reaching w contains an (ℓ, τ)-ineffectual gate. Thus, γ contains an (ℓ, τ)-ineffectual gate, and the claim follows.

□

4.7.2 Proof of Proposition 4.4

Let $A \sim \mu(W)$, where W is good for Φ. Denote $\Psi = \Phi^A$, and recall that we think of the gates of Φ and of the gates of Ψ as the 'same' set of gates (see discussion in Section 2.2). We first note that for every gate v in Φ, regardless of A,

$$|W(v)| = 2 \cdot \mathrm{avg}(v).$$

Let γ be a ℓ-central path for $\ell = \lfloor n^{\frac{2}{9}} \rfloor$ from an input gate to the output gate of Ψ (if there are no such paths, the proposition follows). Let $\tau = \lfloor n^{\frac{2}{27d}}/2 \rfloor$. Since all the variables occur in Φ (and since $\ell \cdot (2\tau)^d < m$ for large enough n), by Claim 4.5, γ contains an (ℓ, τ)-ineffectual gate v. We will show that the probability that v does not have k-low-rank is $2^{-n^{\Omega(1/d)}}$. Then, using the union bound, the proposition will follow.

Since W is good, there exist an integer $\ell' \geq \ell$, and a family of pairwise disjoint sets $C_1, \ldots, C_{\lfloor \tau/2 \rfloor} \subseteq \textsc{child}(v)$ such that for every $i \in \{1, \ldots, \lfloor \tau/2 \rfloor\}$,

$$\ell' \leq \left|W(i)\right| \leq 2\ell' \text{ and } \left|W'(i)\right| \geq \frac{\ell'}{48},$$

where $W(i) = W\big(\bigcup_{u \in C_i} X_u\big)$ and $W'(i) = W'\big(\bigcup_{u \in C_i} X_u\big)$ for every $i \in \{1, \ldots, \lfloor \tau/2 \rfloor\}$. Denote $C_0 = \textsc{child}(v) \backslash (\bigcup_{i \in \{1, \ldots, \lfloor \tau/2 \rfloor\}} C_i)$. Denote $D_i = \left|\mathcal{D}_A(W(i))\right|$, and denote by E the event that $\sum_{i \in \{1, \ldots, \lfloor \tau/2 \rfloor\}} D_i < 2k$.

We will prove that the probability of E is at most $2^{-n^{\Omega(1/d)}}$. This will finish the proof of the proposition: For every $i \in \{0, \ldots, \lfloor \tau/2 \rfloor\}$, denote by M_i the partial derivative matrix of $\prod_{u \in C_i} \widehat{\Phi}_u$, and denote $\mathrm{avg}(i) = \sum_{u \in C_i} \mathrm{avg}(u)$. By property 1. of Proposition 2.5, $\mathrm{Rank}(M_0) \leq 2^{\mathrm{avg}(0)}$, and for every $i \in \{1, \ldots, \lfloor \tau/2 \rfloor\}$,

$$\mathrm{Rank}(M_i) \leq 2^{\mathrm{avg}(i) - D_i/2}.$$

Since Φ is syntactically multilinear, $\mathrm{avg}(v) = \sum_{i \in \{0,\ldots,\lfloor \tau/2 \rfloor\}} \mathrm{avg}(i)$. Furthermore, using property 3. of Proposition 2.5,

$$\mathrm{Rank}(M_v) = \prod_{i \in \{0,\ldots,\lfloor \tau/2 \rfloor\}} \mathrm{Rank}(M_i) \leq 2^{\mathrm{avg}(v) - \sum_{i \in \{1,\ldots,\lfloor \tau/2 \rfloor\}} D_i/2}.$$

Hence,

$$\mathbb{P}[v \text{ doesn't have } k\text{-low-rank}] = \mathbb{P}[\mathrm{Rank}(M_v) > 2^{\mathrm{avg}(v)-k}] \leq \mathbb{P}[E] \leq 2^{-n^{\Omega(1/d)}}.$$

Thus, it remains to bound the probability of E. We will now define $W''(i)$ to be (roughly) the set of variables on which D_i depends. For every variable $x \in W'(i)$, there exists a unique variables $\tilde{x} \in W$ such that $W_j = \{x, \tilde{x}\}$ for some $j \in [m]$ ($\tilde{x}$ is the variable in W in the same row or column as x). Denote by $W''(i)$ the set of all such $\tilde{x}$'s; that is, $W''(i)$ is the set of $\tilde{x} \in W$ such that there exists $x \in W'(i)$ such that $W_j = \{x, \tilde{x}\}$ for some $j \in [m]$.

We now identify a set $T \subseteq \{1, \ldots, \lfloor \tau/2 \rfloor\}$ of size $|T| \geq \tau/400$ such that the random variables $\{D_i\}_{i \in T}$ are 'almost' independent. We will find T by the following iterative process: T_0 is the empty set. Given T_h, we will define T_{h+1} as follows. Since for every $i \in T_h$, we have $|W(i)| \leq 2\ell'$, it follows that

$$\Big| \bigcup_{i \in T_h} W(i) \Big| \leq 2\ell' h.$$

In addition, for every $i \in \{1, \ldots, \lfloor \tau/2 \rfloor\} \setminus T_h$,

$$|W''(i)| = |W'(i)| \geq \frac{\ell'}{48}.$$

Thus, as long as $h < \tau/400$, there exists $j \in \{1, \ldots, \lfloor \tau/2 \rfloor\} \setminus T_h$ such that

$$\Big| W''(j) \setminus \bigcup_{i \in T_h} W(i) \Big| \geq \frac{\ell'}{200}.$$

Set $T_{h+1} = T_h \cup \{j\}$. Finally, set $T = T_h$ for $h = \lceil \tau/400 \rceil$. For simplicity, assume without loss of generality that $T_h = \{1, \ldots, h\}$ for every h.

For every $i \in T$, denote by E_i the event that $D_i < 2k$. Thus, $E \subseteq \bigcap_{i \in T} E_i$, which implies

$$\mathbb{P}[E] \leq \mathbb{P}[E_1] \cdot \mathbb{P}[E_2|E_1] \cdots \mathbb{P}[E_{|T|}|E_1, \ldots, E_{|T|-1}].$$

We will now upper bound $\mathbb{P}[E_i|E_1, \ldots, E_{i-1}]$ for $i \in \{2, \ldots, |T|\}$. Fix i, and partition $W'(i)$ to two sets: the set I' of variables $x \in W'(i)$ such that there exists

$$\tilde{x} \in W''(i) \cap \bigcup_{h \in [i-1]} W(h)$$

such that $W_j = \{x, \tilde{x}\}$ for some $j \in [m]$, and the set $I = W'(i) \setminus I'$. For every $x \in W$, define $\chi(x)$ to be the random variable that takes the value 1 if $A(x) \in Y$ and the value -1 if $A(x) \in Z$. The random variables in $\{\chi(x)\}_{x \in I}$ are independent of $E_1, \ldots, E_{i-1}$ (since each $\chi(x)$ depends only on $\chi(\tilde{x})$). Thus,

$$D_i = \Big| \sum_{x \in W(i)} \chi(x) \Big| = \Big| \sum_{x \in W'(i)} \chi(x) \Big| = \Big| \sum_{x \in I} \chi(x) + \sum_{x \in I'} \chi(x) \Big|.$$

By the construction of T, we have that, conditioned on $E_1, \ldots, E_{i-1}$, the sum $\sum_{x \in I} \chi(x)$ is a sum of at least $\ell'/200$ independent random variables each takes the value 1 with probability $1/2$ and the value -1 with probability $1/2$. Thus, conditioned on $E_1, \ldots, E_{i-1}$, the random variable D_i takes any specific value with probability at most $O(\ell'^{-1/2}) \leq O(\ell^{-1/2})$. By the union bound, conditioned on $E_1, \ldots, E_{i-1}$, the probability of E_i is at most $O(k \cdot \ell^{-1/2}) \leq 1/2$ for large enough n (since $k \leq \ell^{1/3}$). Since $\tau = n^{\Omega(1/d)}$, for large enough n we have

$$\mathbb{P}[E] \leq 2^{-\lfloor \tau/400 \rfloor} \leq 2^{-n^{\Omega(1/d)}}.$$

□

4.8 Proof of Proposition 4.3

4.8.1 A Different Distribution on W

We are interested in showing that if a formula is small enough, then a random W is good for it. It will be more convenient to work with a different distribution on W than the one defined by μ. It will also be more convenient to think of W as a multi-set. We will now define such a distribution, which we shall denote by μ^*. The two distributions on multi-sets (defined by μ and by μ^*) will be similar in that the statistical difference between the two is $o(1)$. However, for the proof of Proposition 4.3, it will be more convenient to think of W as distributed according to μ^*.

A multi-set $T \subseteq X$ is distributed according to μ^* if it is obtained by the following random process. Let $t_1(1), \ldots, t_1(m)$ be m independent uniformly distributed elements of X. For every $i \in [m]$, with probability one half, let $t_2(i)$ be a random element in the same row of $t_1(i)$, and with probability one half, let $t_2(i)$ be a random element in the same column of $t_1(i)$ (chosen uniformly and independently of all other variables). Set T to be the multi-set

$$T = \{t_1(1), t_2(1), t_1(2), t_2(2), \ldots, t_1(m), t_2(m)\}.$$

We will denote by $T \sim \mu^*$ a multi-set T distributed according to μ^*. We will abuse notation and denote by $W \sim \mu$ a multi-set distributed as $A^{-1}(Y \cup Z)$, where $A \sim \mu$.

The following claim shows that the statistical distance between μ and μ^* is small.

Claim 4.6. *For every event E, $\left|\mathbb{P}[W \in E] - \mathbb{P}[T \in E]\right| = o(1)$, where $W \sim \mu$ and $T \sim \mu^*$.*

Proof. The distribution of W is the same as the distribution of T conditioned on two events: the event E_1 in which $|T| = 2m$, and the event E_2 in which for every $i \in [m]$, the two variables $t_1(i), t_2(i)$ do not share a row or a column with $t_1(j), t_2(j)$ for all $j \neq i$.

Since $m \leq n^{1/3}$, both $\mathbb{P}[E_1] = 1 - o(1)$ and $\mathbb{P}[E_2] = 1 - o(1)$. Thus, since $\mathbb{P}[W \in E] = \mathbb{P}\big[T \in E \big| E_1, E_2\big]$, the claim follows. □

4.8.2 Some Probability Bounds

In this section we bound the probabilities of several events. These bounds will help us to prove that W is good with high probability.

We will use the following lemma that is known as Chernoff's bound.

Lemma 4.7. *Let $N \in \mathbb{N}$ and let $p > 0$. Let $\chi_1, \ldots, \chi_N$ be N independent random variables such that for every $i \in [N]$, $\mathbb{P}[\chi_i = 1] = p$ and $\mathbb{P}[\chi_i = 0] = 1 - p$. Then for every $c > 0$,*

$$\mathbb{P}\Big[\Big|\sum_{i \in [N]} \chi_i - pN\Big| > cpN\Big] < 2e^{-2(cp)^2 N}.$$

The probability bounds are given by the following two claims.

Claim 4.8. *Let $\tilde{X}$ be a subset of X, let $T \sim \mu^*$, and let $L > 0$. Denote by $\tilde{T}$ the multi-set $T \cap \tilde{X}$, and denote $\tilde{\alpha} = \left|\tilde{X}\right|/n^2$. Then,*

1. *If $\tilde{\alpha} \geq L$, then with probability at least $1 - 4e^{-0.5L^2 m}$, $m\tilde{\alpha} \leq \left|\tilde{T}\right| \leq 3m\tilde{\alpha}$.*

2. *If $\tilde{\alpha} < L$, then with probability at least $1 - 4e^{-0.5L^2 m}$, $\left|\tilde{T}\right| < 3Lm$.*

Proof. Recall that $T = \{t_1(1), \ldots, t_1(m)\} \cup \{t_2(1), \ldots, t_2(m)\}$. Denote by $\tilde{T}_1$ and by $\tilde{T}_2$ the following two multi-sets:

$$\tilde{T}_1 = \{t_1(1), \ldots, t_1(m)\} \cap \tilde{X} \quad \text{and} \quad \tilde{T}_2 = \{t_2(1), \ldots, t_2(m)\} \cap \tilde{X}.$$

Since $t_1(1), \ldots, t_1(m)$ are independent random variables uniformly distributed in X, and since the probability for each $t_1(i)$ to be in $\tilde{X}$ is $\tilde{\alpha}$, using Lemma 4.7,

$$\mathbb{P}\left[\left|\left|\tilde{T}_1\right| - m\tilde{\alpha}\right| > 0.5m\tilde{\alpha}\right] < 2e^{-0.5\tilde{\alpha}^2 m}.$$

Similarly,

$$\mathbb{P}\left[\left|\left|\tilde{T}_2\right| - m\tilde{\alpha}\right| > 0.5m\tilde{\alpha}\right] < 2e^{-0.5\tilde{\alpha}^2 m}.$$

So, by the union bound, since $\left|\tilde{T}\right| = \left|\tilde{T}_1\right| + \left|\tilde{T}_2\right|$, if $\tilde{\alpha} \geq L$, then

$$\mathbb{P}\left[\left|\left|\tilde{T}\right| - 2m\tilde{\alpha}\right| > m\tilde{\alpha}\right] < 4e^{-0.5L^2 m}.$$

In the same way, if $\tilde{\alpha} < L$,

$$\mathbb{P}\left[\left|\tilde{T}_1\right| \geq 1.5Lm\right] < 2e^{-0.5L^2 m} \quad \text{and} \quad \mathbb{P}\left[\left|\tilde{T}_2\right| \geq 1.5Lm\right] < 2e^{-0.5L^2 m},$$

which implies

$$\mathbb{P}\left[\left|\tilde{T}\right| \geq 3Lm\right] < 4e^{-0.5L^2 m}.$$

□

Claim 4.9. *Let $\tilde{X}$ be a subset of X of size $\left|\tilde{X}\right| \leq n^2/8$, and let $T \sim \mu^*$. Denote by $\tilde{T}'$ the multi-set of elements $t_1(i) \in T$ such that $t_1(i) \in \tilde{X}$ and $t_2(i) \notin \tilde{X}$. Denote $\tilde{\alpha} = \left|\tilde{X}\right|/n^2$. Then, with probability at least $1 - 2e^{-\tilde{\alpha}^2 m/128}$,*

$$\left|\tilde{T}'\right| \geq \frac{\tilde{\alpha} m}{16}.$$

Proof. Recall that we think of X as a matrix. We say that a row of X is *dense* if at least half of the variables in it are in $\tilde{X}$, and we say that a column of X is *dense* if at least half of the variables in it are in $\tilde{X}$. Since $\left|\tilde{X}\right| = \tilde{\alpha} n^2$, at most $2\tilde{\alpha} n$ rows are dense, and at most $2\tilde{\alpha} n$ columns are dense. Thus, at most

$4\tilde{\alpha}^2 n^2$ of the variables of $\tilde{X}$ are both in a dense row and in a dense column. Since $\tilde{\alpha} \leq 1/8$, we have $4\tilde{\alpha}^2 n^2 \leq \left|\tilde{X}\right|/2$. In other words, at least half of the variables in $\tilde{X}$ are either not in a dense row or not in a dense column.

We will now show that the probability that $t_1(i) \in \tilde{X}$ and $t_2(i) \notin \tilde{X}$ is at least $\tilde{\alpha}/8$. With probability $\tilde{\alpha}$, $t_1(i)$ is in $\tilde{X}$. Conditioned on $t_1(i) \in \tilde{X}$, with probability at least $1/2$ (without loss of generality, by the discussion above) $t_1(i)$ is not is a dense row. Conditioned on $t_1(i) \in \tilde{X}$ and on the event that $t_1(i)$ is not in a dense row, with probability $1/2$, $t_2(i)$ is chosen in the same row of $t_1(i)$. Finally, conditioned on $t_1(i) \in \tilde{X}$, on the event that $t_1(i)$ is not in a dense row, and on the event that $t_2(i)$ is chosen in the same row of $t_1(i)$, with probability at least $1/2$, we have $t_2(i) \notin \tilde{X}$.

Thus, since the events $\left\{t_1(1) \in \tilde{X} \text{ and } t_2(1) \notin \tilde{X}\right\}, \ldots, \left\{t_1(m) \in \tilde{X} \text{ and } t_2(m) \notin \tilde{X}\right\}$ are independent, using Lemma 4.7,

$$\mathbb{P}\left[\left|\tilde{T}'\right| < \tilde{\alpha}m/16\right] < 2e^{-\tilde{\alpha}^2 m/128}.$$

□

4.8.3 Proof of Proposition 4.3

We will show that if we choose W according to the distribution μ^*, then W is good with probability at least $1 - |\Phi| \, 2^{-n^{\Omega(1)}}$. The proposition will then follow since the statistical distance between μ and μ^* is $o(1)$ (see Claim 4.6). For the rest of the proof we assume that W is distributed according to μ^*. Set $\ell = \lfloor n^{\frac{2}{9}} \rfloor$ and $\tau = \lfloor n^{\frac{2}{27d}}/2 \rfloor$ (we will assume that n is large enough).

Before considering ineffectual gates we need to make the following observation. Let v be a product gate in Φ, and let $u_1, \ldots, u_t$ be the set of children of v ordered arbitrarily (note that without loss of generality $t \leq n+1$). For an interval $I = [i_1, i_2] = \{i_1, i_1+1, \ldots, i_2 - 1, i_2\}$ in $[t]$, denote by $X(I)$ the set of variables $\bigcup_{j \in I} X_{u_j}$. By Claim 4.8, for every interval $I \subseteq [t]$ such that $|X(I)| < \frac{\ell}{3m} n^2$, with probability at least $1 - 4e^{-\frac{\ell^2}{18m}} \geq 1 - e^{-\Omega(n^{1/9})}$, we have $|W(X(I))| < \ell$. Thus, by the union bound, we can condition (without loss of generality) on the event that for every product gate v, and every interval I of its children, if $|W(X(I))| \geq \ell$, then $|X(I)| \geq \frac{\ell}{3m} n^2$. We will call this event G, for the rest of the proof.

We now go back to ineffectual gates. For an (ℓ, τ)-ineffectual gate v in Φ^A, define E_v to be the event that there exists an integer $\ell' \geq \ell$, and a family of pairwise disjoint sets $C_1, \ldots, C_{\lfloor \tau/2 \rfloor} \subseteq \textsc{child}(v)$ such that for every $i \in \{1, \ldots, \lfloor \tau/2 \rfloor\}$,

$$\ell' \leq \left|W(i)\right| \leq 2\ell' \quad \text{and} \quad \left|W'(i)\right| \geq \frac{\ell'}{48},$$

where $W(i) = W\big(\bigcup_{u\in C_i} X_u\big)$ and $W'(i) = W'\big(\bigcup_{u\in C_i} X_u\big)$ for every $i \in \{1,\ldots,\lfloor\tau/2\rfloor\}$.

Fix an (ℓ,τ)-ineffectual gate v in Φ^A. We will show that the probability of E_v happening is at least $1-2^{-n^{\Omega(1)}}$. The proof will then follow by the union bound. Set $\ell' = |W(v)|/\tau \geq \ell$. Denote the children of v by $\{u_1,\ldots,u_t\}$ (by the order determined above). Since v is (ℓ,τ)-ineffectual, there exists a family of pairwise disjoint intervals $I_1, I_2, \ldots, I_{\lfloor\tau/2\rfloor} \subseteq [t]$ such that for every $i \in \{1,\ldots,\lfloor\tau/2\rfloor\}$, we have $\ell' \leq \big|W(i)\big| \leq 2\ell'$, where we abbreviated $W(i) = W(X(I_i))$. Set $C_i = \{u_j \in \textsc{child}(v) \,:\, j \in I_i\}$ (this coincides with the above definition of $W(i)$). Since we conditioned on G, $|X(i)| \geq \frac{\ell}{3m}n^2$, where we abbreviated $X(i) = X(I_i)$. Thus, by Claim 4.8 with $L = \frac{\ell}{3m}$, with probability at least $1-4e^{-\frac{\ell^2}{18m}} \geq 1-e^{-\Omega(n^{1/9})}$,

$$\frac{m}{n^2}|X(i)| \leq |W(i)| \leq \frac{3m}{n^2}|X(i)|$$

which implies

$$\frac{\ell' n^2}{3m} \leq |X(i)| \leq \frac{2\ell' n^2}{m}.$$

Since $\sum_{i\in\{1,\ldots,\lfloor\tau/2\rfloor\}} |X(i)| \leq n^2$, by the union bound, with probability at least $1-2^{-n^{\Omega(1)}}$, we have $\ell' \leq 6m/\tau$. Hence, with probability at least $1-2^{-n^{\Omega(1)}}$, for every $i \in \{1,\ldots,\lfloor\tau/2\rfloor\}$

$$\frac{\ell' n^2}{3m} \leq |X(i)| \leq \frac{12n^2}{\tau} \leq \frac{n^2}{8}$$

(if $\tau < 96$, the proposition holds trivially). Thus, by Claim 4.9 and by the union bound, with probability at least $1-2^{-n^{\Omega(1)}}$, for every $i \in \{1,\ldots,\lfloor\tau/2\rfloor\}$,

$$|W'(i)| \geq \frac{\ell'}{48}.$$

Thus, the probability of E_v is at least $1-2^{-n^{\Omega(1)}}$, and the proposition follows. □

5 A Tight Lower Bound

In this section we prove a 'tight' lower bound for the size of constant depth multilinear circuits (see Section 6 below for the appropriate upper bound). The general scheme is similar to that of the lower bound for the determinant and the permanent. However, the previous ideas are not sufficient to prove neither the lower bound nor the separation we are aiming at. So, in this part we need to alter the previous definitions, and we need to investigate more cases.

For the rest of this section we will have the following conventions. The set X will be a set of variables of size $n = 2m$. The sets Y and Z will be two sets of variables of size m each. Partitions of X will be one-to-one maps from X to $Y \cup Z$.

Theorem 5.1. *Let Φ be a multilinear arithmetic circuit of product-depth d for $d = o(\log n / \log\log n)$ over the field $\mathbb{G}$ and over the set of variables X computing the polynomial f. If for every one-to-one map A from X to $Y \cup Z$ the polynomial f^A has full rank, then*

$$|\Phi| \geq n^{\Omega((n/\log(n))^{1/d}/d^2)}.$$

For the rest of this section we will prove the above theorem.

5.1 Partitions

Let $m \in \mathbb{N}$ and let $n = 2m$. Let $X = \{x_1, \ldots, x_n\}$ be a set of variables. Let $Y = \{y_1, \ldots, y_m\}$ and let $Z = \{z_1, \ldots, z_m\}$ be two sets of variables. In this section, partitions will be one-to-one maps from X to $Y \cup Z$. In other words, partitions will just be renamings of the variables. We will define ν to be the uniform probability measure on such partitions; that is, ν is the uniform probability measure on one-to-one maps from X to $Y \cup Z$. As usual, we will denote by $A \sim \nu$ a random partition distributed according to ν.

5.2 Small Formulas are Weak

The following theorem shows that small formulas are weak for a random partition.

Theorem 5.2. *Let Φ be a d-normal-form (for $d = o(\log n / \log\log n)$) syntactically multilinear arithmetic formula over the set of variables X such that all the variables occur in Φ, and let $A \sim \nu$ be a random partition of X. Then,*

$$\mathbb{P}[\Phi^A \text{ is not } k\text{-weak}] \leq |\Phi| \, 2^{-k/d} + o(1),$$

for $k = \Omega\big(\big(\frac{n}{\log(n)}\big)^{1/d} \log(n)\big)$.

The above theorem yields the lower bound.

5.3 Proof of Theorem 5.1

Let Φ be a multilinear arithmetic circuit of product-depth d over the set of variables X computing the polynomial f. By Lemma 2.1, there exists a d-normal-form syntactically multilinear arithmetic formula Ψ of size at most $(d+1)^2 |\Phi|^{2d+1}$ computing the polynomial f as well. Assume towards a contradiction that for every one-to-one map A from X to $Y \cup Z$, the polynomial f^A has full rank, and that the size of Ψ is at most $n^{\varepsilon(n/\log(n))^{1/d}/d}$ for some small enough constant $\varepsilon > 0$. Since f^A has full rank for all A, all the variables occur in Ψ. Thus, by Theorem 5.2, there exists a one-to-one map from X to $Y \cup Z$ such that Ψ^A is k-weak for $k = \Omega\big(\big(\frac{n}{\log(n)}\big)^{1/d} \log(n)\big)$. Thus, by Theorem 3.1,

$$\mathrm{Rank}(M_{f^A}) \leq |\Psi| \, 2^{m-k/2} < 2^m,$$

which is a contradiction. Thus, if for every one-to-one map from X to $Y \cup Z$, the polynomial f^A has full rank, then

$$|\Phi| \geq n^{\Omega((n/\log(n))^{1/d}/d^2)}.$$

□

For the rest of this section we prove Theorem 5.2. To prove the theorem we need a probability bound for the hypergeometric distribution, and a 'new' definition of ineffectual gates.

5.4 The Hypergeometric Distribution

Let $N, M_1, M_2 \in \mathbb{N}$ be three integers such that $M_1 \leq N$ and $M_2 \leq N$. Denote by $\mathcal{H}(N, M_1, M_2)$ the hypergeometric distribution with parameters N, M_1, M_2; that is, $\mathcal{H}(N, M_1, M_2)$ is the distribution of $|S_1 \cap S_2|$, where S_1 is a random subset of $[N]$ of size M_1 (chosen uniformly at random from all subsets of $[N]$ of size M_1), and S_2 is a fixed subset of $[N]$ of size M_2.

The following proposition shows that a hypergeometric random variable does not take any specific value with high probability.

Proposition 5.3. *Let $N \in \mathbb{N}$ be an integer. Let χ be a random variable that has the hypergeometric distribution $\mathcal{H}(N, M_1, M_2)$, where*

$$N/4 \leq M_1 \leq N/2 \quad \textit{and} \quad 0 \leq M_2 \leq N/4. \tag{5.1}$$

Then, every $j \in \mathbb{N}$ admits $\mathbb{P}[\chi = j] = O\big((M_2)^{-1/2}\big)$.

Proof. Denote by $j_{\max}$ the maximal value that χ takes with nonzero probability. For every $j \in \mathbb{N}$, denote $P(j) = \mathbb{P}[\chi = j]$. Every $j \in \{0, \ldots, j_{\max}\}$ admits

$$P(j) = \frac{\binom{M_2}{j}\binom{N-M_2}{M_1-j}}{\binom{N}{M_1}}, \tag{5.2}$$

and every $j \notin \{0, \ldots, j_{\max}\}$ admits $P(j) = 0$. Set $j^* \in \{0, \ldots, j_{\max}\}$ to be the integer that maximizes $P(j)$; that is, every $j \in \mathbb{N}$ admits $P(j) \leq P(j^*)$. To find j^* consider $P(j+1)/P(j)$. By (5.2), for all $j \in \{0, \ldots, j_{\max} - 1\}$,

$$\frac{P(j+1)}{P(j)} = \frac{(M_2 - j)(M_1 - j)}{(j+1)(N - M_1 - M_2 + j + 1)},$$

which implies that

$$P(j) \leq P(j+1) \Leftrightarrow j \leq \frac{M_1 M_2 + M_1 + M_2 - N - 1}{N+2} = \frac{M_1 M_2}{N+2} - \frac{N + 1 - M_1 - M_2}{N+2}.$$

Hence, by (5.1),

$$\frac{M_1 M_2}{N+2} - 1 \leq j^* \leq \frac{M_1 M_2}{N+2}.$$

Using Stirling's formula, for every $M \in \mathbb{N}$ and every $k \in [M]$,

$$\binom{M}{k} = \Theta\left(\sqrt{\frac{M}{k(M-k)}}\frac{(M/k-1)^k}{(1-k/M)^M}\right).$$

Thus, using (5.1) and (5.2),

$$P(j^*) \;=\; O\left(\sqrt{\frac{M_2}{j^*(M_2-j^*)}}\sqrt{\frac{N-M_2}{(M_1-j^*)(N-M_2-M_1+j^*)}}\sqrt{\frac{M_1(N-M_1)}{N}}\right) = O\big((M_2)^{-1/2}\big).$$

Hence, every $j \in \mathbb{N}$ admits $P(j) \leq P(j^*) = O\big((M_2)^{-1/2}\big)$. □

5.5 Central Paths and Ineffectual Gates

We will now define a different notion of ineffectual gates than the one defined above in Section 4.5. We will use this notion only for the rest of the proof.

For $\ell, \tau \in \mathbb{N}$, a gate v in Φ is called (ℓ, τ)-*ineffectual*, if v is a product gate such that one of the following holds

· either $|X_v| \geq \ell\tau$, and every gate $u \in \text{CHILD}(v)$ admits $|X_u| \leq |X_v|/\tau$,

· or the product-depth of v is 1 and $|X_v| \geq \ell$.

Remark 5.4. *In Section 3 we defined the notion of a central path for circuits over the two sets of variables Y and Z. Since the definition of a central path depends only on $|Y_v \cup Z_v|$ for the gates v, we will use the notion of central paths also for circuits over the set of variables X.*

The following claim shows that a central path contains an ineffectual gate.

Claim 5.5. *Let $k, \ell, \tau, d \in \mathbb{N}$ be such that $\tau, d \geq 1$. Let Φ be an arithmetic formula over the set of variables X. Let v be a gate in Φ such that Φ_v is of d-normal-form and $|X_v| \geq 2 \cdot \ell \cdot (2\tau)^{d-1}$. Then, every k-central path for $0 \leq k \leq \ell$ from an input gate to v contains an (ℓ, τ)-ineffectual gate.*

Proof. The claim follow by induction on d.

$d = 1$ Let γ be a k-central path reaching v. Let u be the child of v in γ. Since Φ_v has 1-normal-from, u is a product gate of product-depth 1. Since γ is k-central, $|X_u| \geq |X_v| - k \geq 2 \cdot \ell - k \geq \ell$. Thus, u is (ℓ, τ)-ineffectual.

$d > 1$ If there are no k-central paths reaching v, the claim follows. Otherwise, let γ be a k-central path reaching v, and let u be the child of v in γ. Since Φ_v has d-normal-from, u is a product gate. Since γ is k-central, $|X_u| \geq |X_v| - k$. Let w be the child of u in γ. If $|X_w| \leq |X_u|/\tau$, since γ is k-central, the same holds for every child of u. Thus, since $|X_u| \geq \ell\tau$, u is (ℓ, τ)-ineffectual. Otherwise, $|X_w| \geq |X_u|/\tau \geq (2 \cdot \ell \cdot (2\tau)^{d-1} - k)/\tau \geq 2 \cdot \ell \cdot (2\tau)^{d-2}$. Now, by induction, every k-central path from an input gate reaching w contains an (ℓ, τ)-ineffectual gate. Thus, γ contains an (ℓ, τ)-ineffectual gate, and the claim follows.

□

5.6 Proof of Theorem 5.2

Let Φ be a d-normal-form syntactically multilinear arithmetic formula over the set of variables X such that all the variables occur in it, and let $A \sim \nu$ be a random partition of X. Denote $\Psi = \Phi^A$, and

recall that we think of the gates of Φ and of the gates of Ψ as the 'same' set of gates (see discussion in Section 2.2). Fix

$$\tau = \left\lfloor \left(\frac{n}{\log(n)}\right)^{1/d}/2 \right\rfloor,\ \ell = \lfloor \tau \log(n)/2 \rfloor \text{ and } k = c \cdot \ell$$

for some small enough universal constant $c > 0$. We assume that n is large enough.

Let γ be a k-central path reaching the output gate of Φ (note that γ is also a k-central path in Ψ). Since all the variables occur in Φ, by Claim 5.5, γ contains an (ℓ, τ)-ineffectual gate.

By the definition of ineffectual gate, there are two cases to consider. We will show below that in both cases, with probability at most $2^{-k/d}$, v does not have k-low-rank. The proof of the theorem will then follow by the union bound.

5.6.1 Case one

The gate v admits $|X_v| \geq \ell\tau$, and every gate $u \in \text{CHILD}(v)$ admits $|X_u| \leq |X_v|/\tau$ (the analysis in this case is similar to the one in the lower bound for the determinant and the permanent). Thus, there exist an integer $\ell' = |X_v|/\tau \geq \ell$, and a family of pairwise disjoint sets $C_1, \ldots, C_{\lfloor \tau/2 \rfloor} \subseteq \text{CHILD}(v)$ such that for every $i \in \{1, \ldots, \lfloor \tau/2 \rfloor\}$,

$$\ell' \leq |X(i)| \leq 2\ell',$$

where $X(i) = \bigcup_{u \in C_i} X_u$ for all $i \in \{1, \ldots, \lfloor \tau/2 \rfloor\}$. Let $\tau' \in \mathbb{N}$ be the largest integer in $\{0, \ldots, \lfloor \tau/2 \rfloor\}$ such that $\tau'\ell' < n/8$, and consider only $C_1, \ldots, C_{\tau'}$ (note that $\tau' = \Omega(\tau)$). Denote $D(i) = \left|\mathcal{D}_A(X(i))\right|$, and denote by E the event that $\sum_{i \in \{1,\ldots,\tau'\}} D(i) < 2k$.

Below we will upper bound the probability of E by $2^{-k/d}$. We now show that this bound on the probability of E yields a bound on the probability that v does not have k-low-rank. Since Φ is syntactically multilinear and since v is a product gate, by properties 1. and 3. of Proposition 2.5,

$$\text{Rank}(M_v) \leq 2^{\text{avg}(v) - \sum_{i \in [\tau']} D(i)/2},$$

which implies that

$$\mathbb{P}[v \text{ doesn't have } k\text{-low-rank}] = \mathbb{P}[\text{Rank}(M_v) > 2^{\text{avg}(v)-k}] \leq \mathbb{P}[E] \leq 2^{-k/d}.$$

It remains to upper bound the probability of E. Write

$$E = \bigcup_{k_1, \ldots, k_{\tau'}} \{D(1) = k_1\} \cap \{D(2) = k_2\} \cap \cdots \cap \{D(\tau') = k_{\tau'}\},$$

where the union is over all $k_1, \ldots, k_{\tau'} \in \{0, 1, \ldots\}$ such that $k_1 + k_2 + \cdots + k_{\tau'} < 2k$. For every $i \in [\tau']$ and for every $k_1, \ldots, k_i$ as above, using Proposition 5.3, since $\ell' \leq |X(i)| \leq 2\ell'$, and since $\left|\bigcup_{j \in [i-1]} X(j)\right| \leq n/4$ (by the choice of τ'),

$$\mathbb{P}[D(i) = k_i | D(1) = k_1, \ldots, D(i-1) = k_{i-1}] = O(\ell'^{-1/2}).$$

Hence, since the number of tuples $k_1, \ldots, k_{\tau'}$ such that $k_1 + \cdots + k_{\tau'} < 2k$ is at most $\binom{2k+\tau'}{\tau'} \leq \binom{3k}{\tau'} \leq \left(\frac{3ek}{\tau'}\right)^{\tau'}$, using the union bound,

$$\mathbb{P}[E] \leq \left(\frac{3ek}{\tau' \ell^{1/2}}\right)^{\Omega(\tau')} \leq 2^{-k/d}$$

(since $d = o(\log n / \log\log n)$).

5.6.2 Case two

The product-depth of v is 1 and $|X_v| \geq \ell$. Every gate $u \in \textsc{child}(v)$ such that $|X_u| = 1$ admits

$$\mathrm{Rank}(M_u) \leq 1 \leq 2^{\mathrm{avg}(u) - \mathrm{avg}(u)/3}.$$

Every gate $u \in \textsc{child}(v)$ such that $|X_u| = 2$ admits

$$\mathrm{Rank}(M_u) \leq 2 \leq 2^{\mathrm{avg}(u)}.$$

Since the product-depth of v is 1, every gate $u \in \textsc{child}(v)$ such that $|X_u| > 2$ admits

$$\mathrm{Rank}(M_u) \leq 2 \leq 2^{\mathrm{avg}(u) - \mathrm{avg}(u)/3}$$

(since the product gate of v is 1, u is a sum on input gates, and hence computes a linear function; the partial derivative matrix of a linear function has at most one nonzero row and at most one nonzero column). Let $R = R(v)$ be the set of $u \in \textsc{child}(v)$ such that $|X_u| = 2$. Since Φ is syntactically multilinear, $\mathrm{avg}(v) = \sum_{u \in \textsc{child}(v)} \mathrm{avg}(u)$, and by properties 1. and 3. of Proposition 2.5,

$$\mathrm{Rank}(M_v) = \prod_{u \in \textsc{child}(v)} \mathrm{Rank}(M_u) \leq 2^{\mathrm{avg}(v) - \sum_{u \in \textsc{child}(v) \setminus R} \mathrm{avg}(u)/3}.$$

Hence, if $\sum_{u \in \textsc{child}(v) \setminus R} \mathrm{avg}(u)/3 \geq k$, then

$$\mathrm{Rank}(M_v) \leq 2^{\mathrm{avg}(v) - k}$$

(which implies that v has k-low-rank, with probability 1). Otherwise, assume

$$\sum_{u \in \text{CHILD}(v) \setminus R} \text{avg}(u) < 3k,$$

which implies

$$|R| > \frac{|X_v| - 6k}{2}.$$

Since $k \leq \ell/20$,

$$|R| > \frac{|X_v| - 6k}{2} \geq 7\ell/20.$$

We will now partition the gates in R to $\Omega(\ell)$ sets. Let $C \in \mathbb{N}$ be a constant large enough so that in the conclusion of Proposition 5.3, $O(C^{-1/2}) \leq 1/8$ (C takes the role of M_2). There are $r = \lfloor \frac{7\ell}{20C} \rfloor$ disjoint sets $R_1, \ldots, R_r \subseteq R$ such that every $i \in [r]$ admits $|X(i)| = 2C$, where $X(i) = \bigcup_{u \in R_i} X_u$. For $i \in [r]$, denote $D(i) = \left|\mathcal{D}_A(X(i))\right|$. Denote by E the event that $\sum_{i \in [r]} D(i) < 2k$.

As before, our next step is to prove an upper bound on the probability of E. By the choice of C and by Proposition 5.3, since $|X(1)| + |X(2)| + \cdots + |X(r)| \leq n/4$, every $i \in \{2, \ldots, r\}$ admits

$$\mathbb{P}[D(i) = 0 | D(1) = k_1, \ldots, D(i-1) = k_{i-1}] \leq 1/8,$$

for any values of $k_1, \ldots, k_{i-1} \in \{0, 1, \ldots\}$. Thus, by the union bound,

$$\mathbb{P}[E] \leq \sum_{S \subseteq [r] : |S| > r - 2k} \mathbb{P}[\forall\ i \in S \ \ D(i) = 0] \leq r \binom{r}{2k} 2^{-3(r-2k)} \leq 2^{-k}.$$

We now use the bound on the probability of E to conclude a bound on the probability that v does not have k-low-rank. Since Φ is syntactically multilinear and since v is a product gate, by properties 1. and 3. of Proposition 2.5,

$$\text{Rank}(M_v) \leq 2^{\text{avg}(v) - \sum_{i \in [r]} D(i)/2}.$$

Hence,

$$\mathbb{P}[v \text{ doesn't have } k\text{-low-rank}] \leq \mathbb{P}\left[\text{Rank}(M_v) > 2^{\text{avg}(v) - k}\right] \leq \mathbb{P}[E] \leq 2^{-k}.$$

□

6 The Construction

In this section we define a polynomial $f = f_d$ (for a constant d) over the set of variables X such that

- For every one-to-one map A from X to $Y \cup Z$, the polynomial f^A has full rank.
- The polynomial f can be computed by a syntactically multilinear arithmetic formula of product-depth d and size
$$n^{O((n/\log(n))^{1/d})}.$$

The polynomial f implies the tightness of the lower bound proved in the previous section (for constant d). Indeed, since f^A has full rank for all partitions, using Theorem 5.1, every multilinear circuit of product-depth d computing f is of size at least $n^{\Omega((n/\log(n))^{1/d})}$. On the other hand, f has a multilinear circuit of product-depth d and roughly the same size computing it.

6.1 Definition of f

Preliminaries In this section we think of d as a constant independent of n. For the simplicity of the presentation we assume that $n \in \mathbb{N}$ is such that

$$\tau = \left(\frac{n}{\log(n)}\right)^{1/d} \quad \text{and} \quad \ell = \tau \log(n)$$

are two integers and ℓ is even. Note that

$$\tau^{d-1}\ell = n \quad \text{and} \quad n^{\tau} = 2^{\ell}.$$

Let $X = \{x_1, \ldots, x_n\}$ be a set of variables. Let W be a new set of variables that will be defined implicitly below. Let $\mathbb{F}$ be a field and denote by $\mathbb{G} = \mathbb{F}(W)$ the field of rational functions in the variables W over the field $\mathbb{F}$.

Generalized intervals For $i, j \in [n]$, denote by $[i, j] \subseteq [n]$ the interval of integers from i to j modulo n. More precisely,
if $i \leq j$, then

$$[i, j] = \{a \in [n] \ : \ i \leq a \leq j\},$$

and if $i > j$, then

$$[i,j] = \{a \in [n] \ : \ i \le a \le n \ \text{ or } \ 1 \le a \le j\}\,.$$

For a set $T \subseteq [n]$, define a set of generalized intervals that corresponds to T as

$$\mathcal{I}(T) = \{I = [i,j] \cap T \ : \ i,j \in [n]\}\,.$$

Inductive definition of f We will define a family of polynomials $\{f_I\}$ for sets $I \subseteq [n]$. The definition will be by induction.

Induction Base: For a set $I \subseteq [n]$ of size $|I| = \ell$, define

$$f_I = \sum_{S \subseteq I : |S| = \ell/2} w_S \cdot (x_{i_1} + x_{j_1}) \cdot (x_{i_2} + x_{j_2}) \cdots (x_{i_{\ell/2}} + x_{j_{\ell/2}}),$$

where w_S is a variable in W and in the sum above

$$S = \left\{i_1 < \cdots < i_{\ell/2}\right\} \ \text{ and } \ I \setminus S = \left\{j_1 < \cdots < j_{\ell/2}\right\}.$$

Induction Step: For a set $I \subseteq [n]$ of size $|I| = \ell\tau^{d'}$ for some integer $d' \ge 1$, define

$$f_I = \sum_{I_1,\ldots,I_\tau} w_{I,I_1,\ldots,I_\tau} \cdot f_{I_1} \cdot f_{I_2} \cdots f_{I_\tau},$$

where $w_{I,I_1,\ldots,I_\tau}$ is a variable in W and the sum is over $I_1,\ldots,I_\tau$ as follows.

- The generalized interval I_1 varies over all generalized intervals in $\mathcal{I}(I)$ of size $\ell\tau^{d'-1}$.
- Given $I_1,\ldots,I_j$ for $j \in \{1,\ldots,\tau-1\}$, the generalized interval I_{j+1} varies over all generalized intervals in
$$\mathcal{I}\Big(I \setminus \big(I_1 \cup I_2 \cup \cdots \cup I_j\big)\Big)$$
of size $\ell\tau^{d'-1}$.

Finally, we define f to be

$$f = f_{[n]}$$

(note that f depends both on n and on d).

The set W We define the set W to be *only* the variables w that occur in the polynomial f (in the definition above the set W is 'much larger', but for the purpose of defining f we do not need 'most' of the variables). We note that the size of W is roughly the size of the circuit computing f.

6.2 Properties of f

We note the following properties of f.

1. f is a multilinear polynomial over the set of variables X and W.
2. The coefficients of the monomials in f are in $\{0, 1\}$.
3. There exists a d-normal-form syntactically multilinear arithmetic formula of size
$$n^{O(d\tau)} = n^{O(d(n/\log(n))^{1/d})}$$
computing f (since the in-degree of each gate is at most $n^{O(\tau)} = 2^{O(\ell)}$). The size of the circuit computing f is polynomial in the number of variables that occur in f (roughly, for every edge there is a variable that correspond to it).

The following theorem shows that f has full rank for every partition.

Theorem 6.1. *For every one-to-one map $A : X \to Y \cup Z$, the polynomial f^A has full rank over the field $\mathbb{G} = \mathbb{F}(W)$.*

6.3 Proof of Theorem 6.1

For a set $I \subseteq [n]$, define $\mathcal{D}_A(I)$ to be $\mathcal{D}_A(\{x_i \ : \ i \in I\})$.

The proof of the theorem will follow by the following lemma (which we will prove below).

Lemma 6.2. *Let A be a one-to-one map from X to $Y \cup Z$. Let $I \subseteq [n]$ of size $\ell\tau^{d'}$ for some integer $d' \geq 0$. If $\mathcal{D}_A(I) = 0$, then*
$$Rank(M_{f_I^A}) \geq 2^{|I|/2},$$
where the rank is over $\mathbb{G}$.

By the lemma above, for every one-to-one map $A : X \to Y \cup Z$, since $\mathcal{D}_A([n]) = 0$, the polynomial f^A has full rank over the field $\mathbb{G} = \mathbb{F}(W)$. □

For the rest of this section we will prove Lemma 6.2.

6.4 A Technical Claim

Claim 6.3. *Let A be a one-to-one map from X to $Y \cup Z$. Let $I \subseteq [n]$ be of size $2k$ for some integer $k \in \mathbb{N}$, and let $a \in [k]$. If $\mathcal{D}_A(I) = 0$, then there exists a generalized interval $I' \in \mathcal{I}(I)$ of size $2a$ such that $\mathcal{D}_A(I') = 0$.*

Proof. For every $i \in I$, let $I(i)$ be the unique generalized interval of size $2a$ starting at i; that is, $I(i) = [i, j] \cap I$, where j is the unique element of I such that $|[i, j] \cap I| = 2a$. In addition, let $D(i) = \mathcal{D}_A(I(i))$. Since

$$\sum_{i \in I} D(i) = 2a \cdot \mathcal{D}_A(I) = 0,$$

assume without loss of generality that i and j in I are such that $i < j$, $D(i) > 0$ and $D(j) < 0$ (otherwise, every $i \in I$ admits $D(i) = 0$). For every $i' \in \{i, \ldots, j-1\}$, we have that $|D(i') - D(i'+1)| \leq 2$ and that $D(i')$ is an even number. Hence, there exists $i' \in \{i, \ldots, j\}$ such that $D(i') = 0$. □

6.5 Proof of Lemma 6.2

The proof follows by induction on d'.

Induction Base: Assume $d' = 0$. Thus,

$$f_I = \sum_{S \subseteq I : |S| = \ell/2} w_S \cdot (x_{i_1} + x_{j_1}) \cdot (x_{i_2} + x_{j_2}) \cdots (x_{i_{\ell/2}} + x_{j_{\ell/2}}),$$

where w_S is a variable in W and in the sum above

$$S = \left\{ i_1 < \cdots < i_{\ell/2} \right\} \quad \text{and} \quad I \setminus S = \left\{ j_1 < \cdots < j_{\ell/2} \right\}.$$

Denote

$$B = \{ i \in I \ : \ A(x_i) \in Y \},$$

and denote

$$g = (x_{i_1} + x_{j_1}) \cdot (x_{i_2} + x_{j_2}) \cdots (x_{i_{\ell/2}} + x_{j_{\ell/2}}),$$

where

$$B = \left\{ i_1 < \cdots < i_{\ell/2} \right\} \quad \text{and} \quad I \setminus B = \left\{ j_1 < \cdots < j_{\ell/2} \right\}.$$

For every $k \in [\ell/2]$, there exist $y \in Y$ and $z \in Z$ such that

$$(x_{i_k} + x_{j_k})^A = y + z,$$

and hence the partial derivative matrix of $(x_{i_k} + x_{j_k})^A$ is of rank 2. Using property 3. of Proposition 2.5,

$$\mathrm{Rank}(M_{g^A}) = 2^{\ell/2}.$$

Since

$$f_I = w_B g + g',$$

where the degree of w_B in g' is 0,

$$\mathrm{Rank}(M_{f_I^A}) \geq \mathrm{Rank}(M_g) = 2^{\ell/2} = 2^{|I|/2},$$

where the rank is over $\mathbb{G}$.

Induction Step: Assume $d' > 0$. Thus,

$$f_I = \sum_{I_1,\ldots,I_\tau} w_{I,I_1,\ldots,I_\tau} \cdot f_{I_1} \cdot f_{I_2} \cdots f_{I_\tau},$$

where $w_{I,I_1,\ldots,I_\tau}$ is a variable in W and the sum is over $I_1, \ldots, I_\tau$ as follows.

- The generalized interval I_1 varies over all generalized intervals in $\mathcal{I}(I)$ of size $\ell\tau^{d'-1}$.
- Given $I_1, \ldots, I_j$ for $j \in \{1, \ldots, \tau - 1\}$, the generalized interval I_{j+1} varies over all generalized intervals in
$$\mathcal{I}\Big(I \setminus \big(I_1 \cup I_2 \cup \cdots \cup I_j\big)\Big)$$
of size $\ell\tau^{d'-1}$.

We will first prove that there exists $I_1, \ldots, I_\tau$ as above such that every $i \in [\tau]$ admits $\mathcal{D}_A(I_i) = 0$. By Claim 6.3, there exists $I_1 \in \mathcal{I}(I)$ of size $\ell\tau^{d'-1}$ such that $\mathcal{D}_A(I_1) = 0$. Since $\mathcal{D}_A(I) = 0$ and since $\mathcal{D}_A(I_1) = 0$, we have that $\mathcal{D}_A(I \setminus I_1) = 0$. Thus, by Claim 6.3, there exists $I_2 \in \mathcal{I}(I \setminus I_1)$ of size $\ell\tau^{d'-1}$ such that $\mathcal{D}_A(I_2) = 0$. Since $\mathcal{D}_A(I_2) = 0$ and since $\mathcal{D}_A(I \setminus I_1) = 0$, we have that $\mathcal{D}_A(I \setminus (I_1 \cup I_2)) = 0$. Continuing by induction, there exists $I_1, \ldots, I_\tau$ as above such that every $i \in [\tau]$ admits $\mathcal{D}_A(I_i) = 0$.

For every $i \in [\tau]$, denote $f_i = f_{I_i}^A$. By induction,

$$\mathrm{Rank}(M_{f_i}) \geq 2^{|I_i|/2}.$$

Since

$$f_I^A = w_{I,I_1,\ldots,I_\tau} \cdot f_1 \cdot f_2 \cdots f_\tau + f',$$

where the degree of $w_{I,I_1,\ldots,I_\tau}$ in f' is 0, using property 3. of Proposition 2.5

$$\mathrm{Rank}(M_{f_I^A}) \geq \mathrm{Rank}(M_{f_1}) \cdot \mathrm{Rank}(M_{f_2}) \cdots \mathrm{Rank}(M_{f_\tau}).$$

Hence,

$$\mathrm{Rank}(M_{f_I^A}) \geq 2^{(|I_1|+|I_2|+\cdots+|I_\tau|)/2} = 2^{|I|/2}.$$

□

7 The Separation

In this section we will use the previous two sections to prove a super-polynomial separation for the size of product-depth d and product-depth $d+1$ multilinear circuits.

7.1 Proof of Theorem 1.2

- By the construction of $f = f_{d+1}$, there exists a product-depth $d+1$ syntactically multilinear arithmetic formula Φ of size $|\Phi| = \mathrm{poly}(N)$ computing f (roughly, every edge has a variable in W that correspond to it).

- Let Φ be a multilinear arithmetic circuit of product-depth d over the field $\mathbb{F}$ and over the set of variables $X \cup W$ computing f. Think of $\mathbb{F}[X, W]$ as $\mathbb{G}[X]$, where $\mathbb{G} = \mathbb{F}(W)$ is the field of rational functions in the variables W over the field $\mathbb{F}$. Think of f as a polynomial in $\mathbb{G}[X]$. Think of Φ as a circuit over the field $\mathbb{G}$ and over the set of variables X computing f.

  By Theorem 6.1, for every one-to-one map $A : X \to Y \cup Z$, the polynomial f^A has full rank over the field $\mathbb{G}$. Thus, by Theorem 5.1,

  $$|\Phi| \geq n^{\Omega((n/\log(n))^{1/d}/d^2)}.$$

  Since

  $$N \leq n^{O((d+1)(n/\log(n))^{1/(d+1)})}$$

(see Section 6.1),

$$|\Phi| \geq N^{\Omega(\log^{1/(2d)}(N)))}$$

(recall the we think of d as constant).

□

References

[A] M. Ajtai. Σ_1^1-Formulae on Finite Structures. *Ann. Pure Appl. Logic* 24: 1–48, 1983.

[FSS] M. L. Furst, J. B. Saxe and M. Sipser. Parity, Circuits, and the Polynomial-Time Hierarchy. *Mathematical Systems Theory* 17(1): 13–27, 1984.

[GR] D. Grigoriev and A. A. Razborov. Exponential Lower Bounds for Depth 3 Arithmetic Circuits in Algebras of Functions over Finite Fields. *Appl. Algebra Eng. Commun. Comput.* 10(6): 465–487, 2000.

[GK] D. Grigoriev and M. Karpinski. An Exponential Lower Bound for Depth 3 Arithmetic Circuits. *STOC*: 577–582, 1998.

[H] J. Håstad. Computational Limitations for Small Depth Circuits. MIT Press, Cambridge, MA, 1987.

[N91] N. Nisan. Lower Bounds for Non-Commutative Computation. *Proceeding of the 23th STOC*: 410–418, 1991.

[NW96] N. Nisan and A. Wigderson. Lower Bound on Arithmetic Circuits via Partial Derivatives. *Computational Complexity*, 6: 217–234, 1996 (preliminary version in Proceeding of the 36th FOCS 1995).

[R04A] R. Raz. Multi-Linear Formulas for Permanent and Determinant are of Super-Polynomial Size. *Proceeding of the 36th STOC*: 633–641, 2004.

[R04B] R. Raz. Separation of Multilinear Circuit and Formula Size. *Theory Of Computing* Vol. 2, article 6 (2006) (preliminary version in the *Proceeding of the 45th FOCS*: 344–351, 2004 (title: "Multilinear-$NC_1 \neq$ Multilinear-NC_2")).

[R07] R. Raz. Elusive Functions and Lower Bounds for Arithmetic Circuits. Manuscript.

[RSY] R. Raz, A. Shpilka and A. Yehudayoff. A Lower Bound for the Size of Syntactically Multilinear Arithmetic Circuits. *Proceedings of the 48th FOCS*: 438–448, 2007.

[RY] R. Raz and A. Yehudayoff. Balancing Syntactically Multilinear Arithmetic Circuits. To appear in Computational Complexity.

[R] A. Razborov. Lower bounds for the size of circuits of bounded depth with basis $\{\oplus, \wedge\}$. *Math. notes of the Academy of Sciences of the USSR*, 41: 333-338, 1987.

[SW] A. Shpilka and A. Wigderson. Depth-3 Arithmetic Formulae over Fields of Characteristic Zero. *Computational Complexity* 10(1): 1–27, 2001.

[SS] V. Shoup and R. Smolensky. Lower Bounds for Polynomial Evaluation and Interpolation Problems. *Proceedings of the 32nd FOCS*: 378–383, 1991.

[S] R. Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. *Proceedings of the 19th STOC*: 77–82, 1987.

[S73] V. Strassen. Die Berechnungskomplexitat von Elementarsymmetrischen Funktionen und von Interpolationskoeffizienten. *Numerische Mathematik*, 20: 238–251, 1973.

[Y] A. C. C. Yao. Separating the Polynomial-Time Hierarchy by Oracles. *Proceedings of the 26th FOCS*: 1–10, 1985.