Improved lower bound, and proof barrier, for constant depth algebraic circuits

C.S. Bhargav
Indian Institute of Technology, Kanpur, India

Sagnik Dutta
Chennai Mathematical Institute, Chennai, India

Nitin Saxena
Indian Institute of Technology, Kanpur, India

Abstract

We show that any product-depth Δ algebraic circuit for the Iterated Matrix Multiplication Polynomial $\mathrm{IMM}_{n,d}$ (when $d = O(\log n)$) must be of size at least $n^{\Omega\left(d^{1/(\varphi^2)^\Delta}\right)}$ where $\varphi = 1.618\ldots$ is the golden ratio. This improves the recent breakthrough result of Limaye, Srinivasan and Tavenas (FOCS'21) who showed a super polynomial lower bound of the form $n^{\Omega\left(d^{1/4^\Delta}\right)}$ for constant-depth circuits.

One crucial idea of the (LST21) result was to use set-multilinear polynomials where each of the sets in the underlying partition of the variables could be of different sizes. By picking the set sizes more carefully (depending on the depth we are working with), we first show that any product-depth Δ *set-multilinear* circuit for $\mathrm{IMM}_{n,d}$ (when $d = O(\log n)$) needs size at least $n^{\Omega\left(d^{1/\varphi^\Delta}\right)}$. This improves the $n^{\Omega\left(d^{1/2^\Delta}\right)}$ lower bound of (LST21). We then use their Hardness Escalation technique to lift this to general circuits.

We also show that our lower bound cannot be improved significantly using these same techniques. For the *specific* two set sizes used in (LST21), they showed that their lower bound cannot be improved. We show that for any $d^{o(1)}$ set sizes (out of maximum possible d), the scope for improving our lower bound is minuscule: there exists a set-multilinear formula that has product-depth Δ and size almost matching our lower bound such that the value of the measure used to prove the lower bound is maximum for this formula. This results in a barrier to further improvement using the same measure.

2012 ACM Subject Classification Theory of computation Algebraic complexity theory

Keywords and phrases polynomials, lower bounds, algebraic circuits, proof barrier, fibonacci numbers

Digital Object Identifier 10.4230/LIPIcs..2022.23

Funding *Nitin Saxena*: N.S. thanks the funding support from DST-SERB (CRG/2020/000045) and N.Rama Rao Chair

1 Introduction

An *Arithmetic Circuit* is a natural model to compute multivariate polynomials over a field $\mathbb{F}$. It is a layered *directed acyclic graph* with leaves labelled by variables $x_1, \ldots, x_n$ or elements from $\mathbb{F}$. The internal nodes are alternating layers of either addition $(+)$ or multiplication $(\times)$ gates. The circuit computes a polynomial in $\mathbb{F}[x_1, \ldots, x_n]$ in the natural way: the $+$ gates compute arbitrary $\mathbb{F}$-linear combination of their inputs and the $\times$ gates compute the product. The *depth* of the circuit is the number of layers in the circuit and by *product-depth*, we mean the number of layers of multiplication gates (depth is twice the product-depth). *Arithmetic Formulas* are a subclass of circuits whose underlying graph is a *tree*. For general survey of the field of Algebraic Complexity Theory, see [3, 30, 20].

Valiant [33], in a very influential work defined the classes VP and VNP which can be considered the arithmetic analogues of P and NP. Much like in the Boolean world, the

Submitted draft.
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

question of whether VP and VNP are the same is one of the central open problems of algebraic complexity theory. Though the best known lower bounds for general arithmetic circuits [2] ($\Omega(n \log n)$) and formulas [10] ($\Omega(n^2)$) fall far short of the super polynomial lower bounds that we hope to prove, there have been many super polynomial lower bounds known for various restricted classes [22, 23, 24]. See [4, 26] for excellent survey of lower bounds.

One of the most interesting restrictions is that of bounding the *depth* of circuits and formulas. When the depth is a constant, circuits and formulas are equivalent upto polynomial blow up in their size and hence we use them interchangeably in this paper. Unlike the Boolean world though, a very curious phenomenon of *depth reduction* occurs in arithmetic circuits [34, 1, 16, 31, 8] which essentially says that depth 3 and depth 4 circuits are almost as powerful as general ones. Formally, any degree d polynomial f that has a size s circuit can also be computed by a depth 4 *homogeneous* circuit or a depth 3 (possibly non homogeneous) circuit of size $s^{O(\sqrt{d})}$. Hence proving an $n^{\omega(\sqrt{d})}$ lower bound on these special circuits is enough to separate VP from VNP. The extreme importance of bounded depth circuits has led to a large body of work proving lower bounds for these models and their variants [28, 29, 25, 11, 7, 13, 17, 6, 18, 14, 12, 15, 9].

The LST breakthrough. Recently in a remarkable work, Limaye, Srinivasan and Tavenas [19] proved the first superpolynomial lower bound for general constant depth circuits. More precisely, they showed that the Iterated Matrix Multiplication polynomial $\mathrm{IMM}_{n,d}$ (where $d = O(\log n)$) has no product-depth Δ circuits of size $n^{d^{\exp(-O(\Delta))}}$. The polynomial $\mathrm{IMM}_{n,d}$ is defined on $N = dn^2$ variables. The variables are partitioned into d sets $X_1, \ldots, X_d$ of n^2 variables each (viewed as $n \times n$ matrices). The polynomial is defined as the $(1,1)$-th entry of the matrix product $X_1 X_2 \cdots X_d$. All monomials of the polynomial are of the same degree and so $\mathrm{IMM}_{n,d}$ is *homogeneous*. As the the individual degree of any variable is at most 1, it is also *multilinear*. Moreover, every monomial has exactly one variable from each of the sets $X_1, \ldots, X_d$. Hence the polynomial is also *set-multilinear*. For any $\Delta \le \log d$, $\mathrm{IMM}_{n,d}$ has a set-multilinear circuit of product-depth Δ and size $n^{O(d^{1/\Delta})}$. There are no significantly better upper bounds known even if we allow general circuits. It makes sense to conjecture that this upper bound is tight (see [5] for improvement limitations in special cases).

The lower bound of [19] proceeds by first transforming size s, product-depth Δ, general algebraic circuits computing a set-multilinear polynomial of degree d to set-multilinear algebraic circuits of product-depth 2Δ and size $\mathrm{poly}(s)d^{O(d)}$ (which is not huge if d is small). Hence lower bounds on bounded depth set-multilinear circuits translate to bounded depth general circuit lower bounds albeit with some loss. Then they consider set-multilinear circuits with variables partitioned into sets of different sizes and crucially use this discrepancy of set sizes to obtain strong set-multilinear lower bounds.

In a more recent work [32], the same authors prove a product-depth Δ set-multilinear formula lower bound of $(\log n)^{\Omega(\Delta d^{1/\Delta})}$ for $\mathrm{IMM}_{n,d}$. There is no restriction of degree here, but in the small degree regime, the bound is much weaker than [19] and cannot be used for escalation. We will only be interested in the low degree regime where we can translate set-multilinear circuit lower bounds to general circuits.

Our Results. In this work, we improve the lower bound for IMM against constant depth circuits. We also exhibit barriers to improving the bound further using these techniques, which is of importance as this is the only known approach to achieve super polynomial lower bounds for general low depth circuits.

For the rest of this paper, let $\mu(\Delta) = 1/(F(\Delta) - 1)$ where $F(n) = \Theta(\varphi^n)$ is the n-th Fibonacci number (starting with $F(0) = 1$, $F(1) = 2$) and $\varphi = (1 + \sqrt{5})/2 = 1.618\ldots$ is the golden ratio.

▶ **Theorem 1.** *(General circuit lower bound) Fix a field* $\mathbb{F}$ *of characteristic* 0 *or characteristic* $> d$*. Let* N, d, Δ *be such that* $d = o(\log N)$*. Then, any product-depth* Δ *circuit computing* $\mathrm{IMM}_{n,d}$ *on* $N = dn^2$ *variables must have size at least* $N^{\Omega\left(d^{\mu(2\Delta)}/\Delta\right)}$*.*

▶ Remark. Theorem 1 improves on the lower bound of $N^{\Omega\left(d^{1/(2^{2\Delta}-1)}/\Delta\right)}$ of [19] since $F(2\Delta) = \Theta(\varphi^{2\Delta}) \ll 2^{2\Delta}$.

To prove Theorem 1, we use the hardness escalation given by (Lemma 6) which allows for a way to convert general circuits to set-multilinear ones without too much size blow up, provided the degree is small. The actual lower bound is proved on set-multilinear circuits.

▶ **Theorem 2.** *(Set-multilinear circuit lower bound) Let* $d \leq (\log n)/4$*. Any product-depth* Δ *set-multilinear circuit computing* $\mathrm{IMM}_{n,d}$ *must have size at least* $n^{\Omega\left(d^{\mu(\Delta)}/\Delta\right)}$*.*

▶ Remark. This is an improvement over the $n^{\Omega\left(d^{1/(2^{\Delta}-1)}/\Delta\right)}$ bound of [19, Lemma 15]. Also, the result holds over any field $\mathbb{F}$. The restriction on the characteristic in Theorem 1 comes from the conversion to set-multilinear circuits. The difference between $\mu(2\Delta)$ in Theorem 1 and $\mu(\Delta)$ in Theorem 2 is also due to the doubling of product-depth during this conversion.

Proof of Theorem 1. From Lemma 6 and Theorem 2, for a circuit of product-depth Δ and size s computing $\mathrm{IMM}_{n,d}$ we get that $d^{O(d)}\mathrm{poly}(s) \geq N^{\Omega\left(d^{\mu(2\Delta)}/2\Delta\right)}$. Following the proof of [19, Corollary 4], if $2\Delta \geq \frac{1}{2}\log_\varphi \log_2 d$, then $d^{\mu(2\Delta)} = d^{(1/\log d)^{\Theta(1)}} < 1$ and hence the $N^{d^{\mu(2\Delta)}}$ bound is trivial. Otherwise $d^{\mu(2\Delta)} \geq d^{1/\varphi^{2\Delta}} \geq \omega(\log d)$ and by the assumption $\log N \geq d$, we get $N^{d^{\mu(2\Delta)}} = 2^{d^{\mu(2\Delta)}\log N} \geq 2^{\omega(d \log d)} \geq d^{\omega(d)}$. Hence $\mathrm{poly}(s) \geq N^{\Omega\left(d^{\mu(2\Delta)}/2\Delta\right)}/d^{O(d)} = N^{\Omega\left(d^{\mu(2\Delta)}/4\Delta\right)}$ implying the required lower bound on s and thus, also Theorem 1. ◀

The hard polynomial for which we prove set-multilinear lower bound is actually a word polynomial (Definition 4) which is a set-multilinear restriction of IMM (Lemma 5). Hence the lower bound gets translated to $\mathrm{IMM}_{n,d}$. These word polynomials are set-multilinear with respect to $(X_1, \ldots, X_d)$ where each of the X_is could potentially have different sizes.

For the two specific set sizes that they consider in [19], they also exhibit polynomials that match their lower bound. It still leaves open the question whether we can improve the lower bound if we choose some other set sizes. We show that this is indeed possible in Theorem 2. It is plausible that using many more set sizes could improve the lower bound further. We answer this question in the negative for most cases. Suppose there are γ different set sizes among the X_is. We show that there are set-multilinear polynomials which can be computed by product-depth Δ circuits having roughly the same size as the size lower bound of Theorem 2, provided γ is not too large. Formally,

▶ **Theorem 3.** *(Barrier) Let* $s_1, \ldots, s_\gamma$ *be positive integers. Fix sets* $X_1, \ldots, X_d$ *where for all* i*,* $|X_i| \in \{s_1, \ldots, s_\gamma\}$*. For any fixed positive integer* Δ*, there exist polynomials* P_Δ *and* Q_Δ *that are set-multilinear with respect to* $X_1, \ldots, X_d$ *such that* P_Δ *can be computed by product-depth* Δ *circuits of size* $n^{O\left(\Delta\gamma d^{\mu(\Delta)}\right)}$ *and* Q_Δ *can be computed by product-depth* Δ *circuits of size* $n^{O\left(\Delta d^{\mu(\Delta-1)}+\gamma\right)}$*. Moreover, both* P_Δ *and* Q_Δ *maximise the measure used to prove lower bounds.*

▶ Remark. The two different polynomials with slightly different sizes will imply barriers to improving the lower bound in different regimes of γ. When $\gamma = O(1)$, the size of P_Δ matches our lower bound essentially implying the tightness of the bound. When γ is $d^{o(1)}$, the size of Q_Δ is only slightly larger than the lower bound (note $\mu(\Delta - 1)$ vs $\mu(\Delta)$). Hence even using multiple set sizes, the scope for improvement is tiny.

2 Preliminaries

For any positive integer n, we denote by $F(n)$ the n-th Fibonacci number with $F(0) = 1$, $F(1) = 2$ and $F(n) = F(n-1) + F(n-2)$. The nearest integer to any real number r is denoted by $\lfloor r \rceil$. We follow the notation of [19] as much as possible for better readability.

Consider set-multilinear polynomials in the sets $(X_1, \ldots, X_d)$. Words are defined as tuples $(w_1, \ldots, w_d)$ of length d where the w_i's are real numbers. These words define the actual set sizes of the set-multilinear polynomials we will be working with. Given a word w, let $\overline{X}(w)$ denote the tuple of sets of variables $(X_1(w), \ldots, X_d(w))$ where the size of each $X_i(w)$ is $2^{|w_i|}$, obtained by arbitrarily removing the rest of the variables. We denote the space of these polynomials by $\mathbb{F}_{sm}[\overline{X}(w)]$.

For any subset $S \subseteq [d]$, the sum of weights at its indices is denoted by $w_S = \sum_{i \in S} w_i$. If for all $t \leq d$, $|w_{[t]}| \leq b$, then we call w b-unbiased. Denote by $w_{|S}$ the sub-word indexed by w. The positive and negative indices of w are denoted $\mathcal{P}_w = \{i \mid w_i \geq 0\}$ and $\mathcal{N}_w = \{i \mid w_i < 0\}$ respectively. Furthermore, $\mathcal{M}_w^{\mathcal{P}}$ (resp. $\mathcal{M}_w^{\mathcal{N}}$) is the set of all multilinear monomials over the positive (resp. negative) variables.

The *partial derivative matrix* $\mathcal{M}_w(f)$ has rows indexed by $\mathcal{M}_w^{\mathcal{P}}$ and columns by $\mathcal{M}_w^{\mathcal{N}}$. The entry corresponding to row $m_+ \in \mathcal{M}_w^{\mathcal{P}}$ and $m_- \in \mathcal{M}_w^{\mathcal{N}}$ is the coefficient of the monomial $m_+ m_-$ in f. The complexity measure we use is the *relative rank*, same as [19]:

$$\mathrm{relrk}_w(f) := \frac{\mathrm{rank}(\mathcal{M}_w(f))}{\sqrt{|\mathcal{M}_w^{\mathcal{P}}| \cdot |\mathcal{M}_w^{\mathcal{N}}|}} = \frac{\mathrm{rank}(\mathcal{M}_w(f))}{2^{\frac{1}{2}\sum_{i \in [d]} |w_i|}} \leq 1$$

The following properties of relrk_w will be useful (see [19] for the proofs).

1. (Imbalance) For any $f \in \mathbb{F}_{sm}[\overline{X}(w)]$, $\mathrm{relrk}_w(f) \leq 2^{-|w_{[d]}|/2}$.
2. (Additivity) For any $f, g \in \mathbb{F}_{sm}[\overline{X}(w)]$, $\mathrm{relrk}_w(f + g) \leq \mathrm{relrk}_w(f) + \mathrm{relrk}_w(g)$.
3. (Multiplicativity) Suppose $f = f_1 f_2 \cdots f_t$ where $f_i \in \mathbb{F}_{sm}[\overline{X}(w_{|S_i})]$ and $(S_1, \ldots, S_t)$ is a partition of $[d]$. Then, $\mathrm{relrk}_w(f) = \mathrm{relrk}_w(f_1 f_2 \cdots f_t) = \prod_{i \in [t]} \mathrm{relrk}_{w_{|S_i}}(f_i)$

We define the hard polynomials we prove lower bounds for. For any monomial $m \in \mathbb{F}_{sm}[\overline{X}(w)]$, let $m_+ \in \mathcal{M}_w^{\mathcal{P}}$ and $m_- \in \mathcal{M}_w^{\mathcal{N}}$ be as above. As $|X_i| = 2^{|w_i|}$, the variables of X_i can be indexed using boolean strings of length $|w_i|$. This gives a way to associate a boolean string with any monomial. Let $\sigma(m_+)$ and $\sigma(m_-)$ be the strings associated with m_+ and m_- respectively. We write $\sigma(m_+) \sim \sigma(m_-)$ if one is a prefix of the other.

▶ **Definition 4.** *[19, Word polynomials] Let w be any word. The polynomial P_w is defined as the sum of all monomials m such that $\sigma(m_+) \sim \sigma(m_-)$.*

The matrices $M_w(P_w)$ have full rank (equal to either the number of rows or columns, whichever is smaller) and hence $\mathrm{relrk}_w(P_w) = 2^{-w_{[d]}/2}$. We also note (without proof) that these polynomials can be obtained as *set-multilinear* restrictions of $\mathrm{IMM}_{n,d}$.

▶ **Lemma 5.** *[19, Lemma 8] Let w be any b-unbiased word. If there is a set-multilinear circuit computing $\mathrm{IMM}_{2^b,d}$ of size s and product-depth Δ, then there is also a set-multilinear circuit of size s and product-depth Δ computing a polynomial $P_w \in \mathbb{F}_{sm}[\overline{X}(w)]$ such that $\mathrm{relrk}_w(P_w) \geq 2^{-b/2}$.*

We also state the set-multilinearization lemma alluded to before:

▶ **Lemma 6.** *[19, Proposition 9] Let s, N, d, Δ be growing parameters with $s \geq Nd$. If C is a circuit of size at most s and product-depth at most Δ computing a set-multilinear polynomial P over the sets of variables $(X_1, \ldots, X_d)$ (with $|X_i| \leq N$), then there is a* set-multilinear circuit *$\tilde{C}$ of size $d^{O(d)}\mathrm{poly}(s)$ and product-depth at most 2Δ computing P.*

3 Proof outline

From the discussion in Section 1 and Lemmas 5 and 6, in order to prove general circuit lower bounds, it suffices to prove that there is a high rank word polynomial that needs large set-multilinear formula. For a word (and hence set sizes) of our choice, we show that relrk_w is small for set-multilinear formulas of a certain size.

Let k be an integer close to $\log_2 n$. In [19], the authors choose the positive entries of the word w to be an integer close to $k/\sqrt{2}$ and the negative entries to be $-k$. Evidently, these entries are independent of the product-depth Δ. In this paper, we take the positive entries to be $(1-p/q)k$ and the negative entries to be $-k$ where p and q are suitable integers dependent on Δ. This *depth-dependent* construction of the word enables us to improve the lower bound. We demonstrate the high level proof strategy of the lower bound for the case of product-depth 3.

Proof overview of Theorem 2 for $\Delta = 3$. Define $G(i) = 1/\mu(i) = F(i) - 1$ for all i and let $\lambda = \lfloor d^{1/G(3)} \rfloor$. Consider a set-multilinear forumula C of product-depth 3 and let v be a gate in it. Suppose that the subformula $C^{(v)}$ rooted at v has product-depth $\delta \le 3$, size s and degree $\ge \lambda^{G(\delta)}/2$. We will prove that $\mathrm{relrk}_w(C^{(v)}) \le s2^{-k\lambda/48}$ by induction on δ. This will give us the desired upper bound of the form $s2^{-k\lambda/48} = sn^{-\Omega(d^{\mu(3)})}$ on the relative rank of the whole formula when v is taken to be the output gate. Write $C^{(v)} = C_1 + \cdots + C_t$ where each C_i is a subformula of size s_i rooted at a product gate. Because of the subadditivity of relrk_w, it suffices to show that $\mathrm{relrk}_w(C_i) \le s_i 2^{-k\lambda/48}$ for all i.

Base case: If $\delta = 1$, then C_i is a product of linear forms. Thus, it has rank 1 and hence low relative rank.

Induction step: $\delta \in \{2, 3\}$. Write $C_i = C_{i,1} \ldots C_{i,t}$ where each $C_{i,j}$ is a subformula of product-depth $\delta - 1$. If any $C_{i,j}$ has degree $\ge \lambda^{G(\delta-1)}/2$, then by induction hypothesis, the relative rank of $C_{i,j}$ and hence C_i will have the desired upper bound and we are done.

Otherwise each $C_{i,j}$ has degree $D_{ij} < \lambda^{G(\delta-1)}/2$. As the formula is set-multilinear, there is a collection of variable-sets $(X_l)_{l\in S_j}$ with respect to which $C_{i,j}$ is set-multilinear. For $j \in [t_i]$, let a_{ij} be the number of positive indices in S_j i.e. the number of positive sets in the collection $(X_l)_{l\in S_j}$. Then the number of negative indices is $(D_{ij} - a_{ij})$.

We consider two cases: if $a_{ij} \le D_{ij}/3$, then $w_{S_j} \le (D_{ij}/3)\cdot\alpha k + (2D_{ij}/3)\cdot(-k) \le -D_{ij}k/3$. Otherwise $a_{ij} > D_{ij}/3$ and if we can prove that $|w_{S_j}| \ge a_{ij}k/(4\lambda^{G(\delta)-1})$,then in both of the above cases, we would have $|w_{S_j}| \ge D_{ij}k/(12\lambda^{G(\delta)-1})$. By the multiplicativity and imbalance property of relrk_w, it would follow that $\mathrm{relrk}_w(C_i) \le 2^{\sum_{j=1}^{t_i} -\frac{1}{2}|w_{S_j}|} \le 2^{-k\lambda/48}$ and we would be done. Thus, we now only have to show that $|w_{S_j}| \ge a_{ij}k/(4\lambda^{G(\delta)-1})$. We have

$$|w_{S_j}| = |a_{ij}(1-p/q) - (D_{ij} - a_{ij})|\, k.$$

Notice that $|w_{S_j}|/k$ is the distance of $a_{ij}p/q$ from some integer, so it must be at least the minimum of $\{a_{ij}p/q\}$ and $1 - \{a_{ij}p/q\}$ where $\{.\}$ denotes the fractional part. The number $a_{ij}p/q$ being rational, has a fractional part $\zeta = (a_{ij}p \bmod q)/q$ and hence it comes down to solving the following system of inequalities:

$$\min(\zeta,\ 1-\zeta) \ge a_{ij}/(4\lambda^{G(\delta)-1}) \text{ for } \delta \in \{2,3\} \text{ when } a_{ij} \le D_{ij} < \lambda^{G(\delta-1)}/2$$

Assign $p = \lambda$, $q = \lambda^2 + 1$. The $\delta = 2$ case is clearly satisfied as $(a_{ij}\lambda \mod (\lambda^2+1)) = a_{ij}\lambda$ when $0 \le a_{ij} \le \lambda/2$.

Consider the case of $\delta = 3$ and $a_{ij} < \lambda^2/2$. Write $a_{ij} = y_1\lambda + y_0$ for integers $y_1 = \lfloor a_{ij}/\lambda \rfloor < \lambda/2$ and $y_0 \le \lambda - 1$. Thus, $a_{ij}\lambda \equiv -y_1 + y_0\lambda \mod (\lambda^2+1)$. Through some case analysis, one can show that $\min\Big(|y_0\lambda - y_1|,\ \lambda^2 + 1 - |y_0\lambda - y_1|\Big) \ge y_1$ which immediately implies the inequality for the $\delta = 3$ case as $y_1 = \lfloor a_{ij}/\lambda \rfloor \geq a_{ij}/(2\lambda)$.

We can attempt to extend this proof technique to product-depth 4 as follows:
We would similarly want to express a_{ij} as $a_{ij} = y_2\lambda^2 + y_1\lambda + y_0$ for integers $y_2 = \lfloor a_{ij}/\lambda^2 \rfloor, y_0 \leq \lambda - 1$ and $y_1 \leq \lambda - 1$. Ideally, we would want that for some $q \approx \lambda^4$,

$$p\lambda^2 \equiv 1 \bmod q,\ p\lambda \equiv \lambda^2 \bmod q \text{ and } p \equiv \lambda^3 \bmod q$$

so that $a_{ij}p \equiv y_2 - y_1\lambda + y_0\lambda^2 \bmod q$ and then we can carry out a similar analysis as in the $\Delta = 3$ case. But this is not possible since multiplying the second congruence equation by λ gives $p\lambda^2 \equiv \lambda^3 \bmod q$, which contradicts with the first congruence equation. So we decide to express a_{ij} as $a_{ij} = y_2b_2 + y_1b_1 + b_0$ where b_2, b_1, b_0 are close to $\lambda^2, \lambda, 1$ respectively instead of being precisely equal to these powers of λ. Then we choose $c_2 \approx 1, c_1 \approx -\lambda, c_0 \approx \lambda^2$ and we assign values to p and q such that

$$pb_2 \equiv c_2 \bmod q,\ pb_1 \equiv c_1 \bmod q \text{ and } pb_0 \equiv c_0 \bmod q.$$

It's easy to verify that all these conditions are satisfied if we define
$b_0 = 1, b_1 = \lambda, b_2 = b_1(\lambda - 1) + b_0$; $\quad c_2 = 1, c_1 = -\lambda, c_0 = c_2 - c_1(\lambda - 1)$;
$p = c_0$ and $q = pb_1 - c_1$.
This inspired our construction of the sequences $\{b_m\}$ and $\{c_m\}$ for general product-depth Δ.

Proof overview of Theorem 3. As mentioned before, we would like to find a family of polynomials for which our lower bound is tight. All the same, we want to maintain high relative rank of these polynomials. If we are able to achieve this and find the appropriate small sized formulas for the said polynomials, we will have that the lower bound cannot be improved using the relative rank measure.

The polynomial P we define will be a close variant of the word polynomials from before. This will ensure that the partial derivative matrix has the maximum possible rank for a matrix of its dimension. From the Imbalance property, the relative rank we obtain is $2^{-|w_{[d]}|/2}$ where we have ensured that $w_{[d]}$ is small. We want to construct the formula F for P such that it has a nice inductive structure. That is, we want the polynomials computed by the subformulas of F to also have high relative rank. This will help us construct a formula from its sub formulas while maintaining high relative rank.

Suppose a subformula F' of F is set multilinear with respect to a subtuple $\mathcal{T}$ of the sets of variables $\overline{X}(w)$. Let these sets in $\mathcal{T}$ be indexed by a set $S_{\mathcal{T}} \subseteq [d]$. As we would like high relative rank of $P_{F'}$, the Imbalance property again suggests that $|w_{S_{\mathcal{T}}}|$ be small. And we desire this of every subformula, their subformulas, and so on. So roughly, we want a way to partition our intial index set $[d]$ into some number of index sets $S_1, \ldots, S_r$ such that each $|w_{S_i}|$ is small. Suppose we are then able to create subformulas of rank $2^{-|w_{S_i}|/2}$. It turns out that we will have to add roughly $2^{\sum_i |w_{S_i}|}$ many of them to get a polynomial of high relative rank. So to control the size of the formula, we would like $\sum_i |w_{S_i}|$ to be small as well.

In their Depth Hierarchy section, [19] use Dirichlet's approximation principle [27] to pick these nice index sets $\{S_i\}$. Their procedure only works for the particular two variable-set sizes they choose. We extend this to any two set sizes in Claim 13. Interestingly, we do not use Dirichlet to pick the index sets but rather to obtain a lower bound on the size of the sets that we do eventually pick. We think of picking sets as an investment process: when we pick a set S, we buy the $|S|$ elements in it for a cost of $|w_S|$. Hence the cost per element is $|w_S|/|S|$. At each product-depth, we are only allowed to pick sets of size under a certain threshold and we pick the ones with the lowest cost per element. It turns out that this lowest cost decreases exponentially as the depth increases and helps us build a small formula. The decrease is captured by the Fibonacci numbers and is the reason why they emerge in our lower bound and upper bound.

Making these ideas precise requires extensive notation and we postpone further discussion to Section 5.

4 The lower bound: Proof of Theorem 2

In this section we prove the set-multilinear lower bound of Theorem 2.

Fix the product-depth Δ for which we want to prove the lower bound. Define $G(i) := F(i) - 1$ for all i and $\lambda = \lfloor d^{1/G(\Delta)} \rfloor$. We can assume that $\lambda \geq 2$ because otherwise $d^{\mu(\Delta)} < 2$ and in that case, the lower bound is uninteresting. The lower bound we aim to prove is $n^{\Omega(d^{1/G(\Delta)})}$. We first define the sequences $\{b_m\}$ and $\{c_m\}$ mentioned in the proof overview:

Let $r_m := \lambda^{G(m+1)-G(m)} - 1$ for $0 \leq m \leq \Delta - 2$.

Define $$b_0 := 1, \quad b_1 := \lambda \text{ and } b_m := b_{m-2} + r_{m-1} b_{m-1} \text{ for } 2 \leq m \leq \Delta - 2$$

Define $$c_{\Delta-2} := (-1)^{\Delta-2}, \qquad c_{\Delta-3} := (-1)^{\Delta-3} \lambda^{G(\Delta-1)-G(\Delta-2)} \text{ and}$$
$$c_m := (-1)^m (|c_{m+2}| + r_{m+1}|c_{m+1}|) \text{ for } \Delta - 4 \geq m \geq 0.$$

Note that the sign parity of c_m is $(-1)^m$ for all m.

Thus, $c_{m-2} = (-1)^{m-2}(|c_m| + r_{m-1}|c_{m-1}|) = c_m - r_{m-1}c_{m-1}$ which implies

$$c_m = c_{m-2} + r_{m-1} c_{m-1} \text{ for } 2 \leq m \leq \Delta - 2$$

Each b_m is close to λ^{G_m} and each $|c_m|$ is close to $\lambda^{G(\Delta-1)-G(m+1)}$:

$$\frac{\lambda^{G(m)}}{2} \leq b_m \leq \lambda^{G(m)} \quad \text{and} \quad \frac{\lambda^{G(\Delta-1)-G(m+1)}}{2} \leq |c_m| \leq \lambda^{G(\Delta-1)-G(m+1)} \quad \text{for all } m \quad (1)$$

We prove this as Lemma 17 in Section A.

Define $$p := c_0 \quad \text{and} \quad q := pb_1 - c_1 = c_0(r_0 + 1) - c_1$$

By defining p and q this way, we have ensured that $pb_0 \equiv c_0 \bmod q$ and $pb_1 \equiv c_1 \bmod q$. Hence from the relations $b_m = b_{m-2} + r_{m-1}b_{m-1}$ and $c_m = c_{m-2} + r_{m-1}c_{m-1}$, it inductively follows that

$$pb_m \equiv c_m \bmod q \qquad \text{for } 0 \leq m \leq \Delta - 2 \tag{2}$$

Constructing the word: Define $\alpha = 1 - p/q$. As $\dfrac{p}{q} \leq \dfrac{c_0}{c_0(r_0+1)} = 1/\lambda$, we have $\alpha \geq 1/2$. Since $q = c_0\lambda - c_1$, it implies that

$$q \leq |c_0|\lambda + |c_1| \leq 2\lambda^{G(\Delta-1)} \leq d < \lfloor \log_2 n \rfloor / 2$$

where the second inequality follows from the upper bound on each $|c_m|$ in (1). Therefore, there exists a multiple of q in the interval $\left[\frac{\lfloor \log_2 n \rfloor}{2}, \lfloor \log_2 n \rfloor\right]$. Let k be this multiple of q. Then αk is an integer. We can construct a word w over the alphabet $\{\alpha k, -k\}$ such that w is k-unbiased. This can be done using induction: if $|w_{[i]}| \leq 0$, set $w_{i+1} = \alpha k$, otherwise set $w_{i+1} = -k$.

With these definitions in place, we are ready to prove Theorem 2. Assume the following lemma:

▶ **Lemma 7.** *Let $\delta \leq \Delta$ be an integer and α, k be as defined above. Let w be any word of length d over the alphabet $\{\alpha k, -k\}$. Then any set-multilinear formula C of product-depth δ, degree $D \geq \lambda^{G(\delta)}/8$ and size at most s satisfies*

$$\text{relrk}_w(C) \leq s 2^{-k\lambda/256}.$$

Proof of Theorem 2. By lemma 5, there exists a set-multilinear projection P_w of $\text{IMM}_{2^k,d}$ such that $\text{relrk}_w(P_w) \geq 2^{-k}$. If there is a set-multilinear circuit of size s and product-depth

Δ computing $\mathrm{IMM}_{n,d}$, then we can expand it to a set-multilinear formula of size at most $s^{2\Delta}$ which computes the same polynomial. Hence we will also have a set-multilinear formula of size at most $s^{2\Delta}$ computing P_w. As $d \geq \lambda^{G(\Delta)}/8$, taking the particular case of $\delta = \Delta$ in lemma 7, we obtain $\mathrm{relrk}_w(P_w) \leq s^{2\Delta} 2^{-k\lambda/256}$. This gives the desired lower bound

$$s^{2\Delta} \geq 2^{-k} 2^{k\lambda/256} \geq \left(\frac{n}{4}\right)^{\frac{d^{1/G(\Delta)}}{512}} /n = n^{\Omega(d^{\mu(\Delta)})}.$$

◀

Proof of Lemma 7. We proceed by induction on δ. We can write $C = C_1 + \cdots + C_t$ where each C_i is a subformula of size s_i rooted at a product gate. Because of the subadditivity of relrk_w, it suffices to show that

$$\mathrm{relrk}_w(C_i) \leq s_i 2^{-k\lambda/256} \qquad \text{for all } i.$$

Base case: C has product-depth $\delta = 1$ and degree $D \geq \lambda/8$.
Then C is a product of linear forms. If L is linear form on some variable set $X(w_i)$, then $\mathrm{relrk}_w(L) \leq 2^{-|w_i|/2} \leq 2^{-k/4}$. Therefore by the multiplicativity of relrk_w,

$$\mathrm{relrk}_w(C_i) \leq 2^{-kD/4} \leq 2^{-k\lambda/32}$$

Induction hypothesis: Assume that the lemma is true for all product-depths $\leq \delta - 1$.
Induction step: Let C be a formula of product-depth δ and degree $D \geq \lambda^{G(\delta)}/8$.
We can write $C_i = C_{i,1} \ldots C_{i,t_i}$ where each $C_{i,j}$ is a subformula of product-depth $\delta - 1$.
If C_i has a factor, say $C_{i,1}$, of degree $\geq \lambda^{G(\delta-1)}/8$, then by induction hypothesis,

$$\mathrm{relrk}_w(C_i) \leq \mathrm{relrk}_w(C_{i,1}) \leq s_i 2^{-k\lambda/256}$$

Otherwise every factor of C_i has has degree $< \lambda^{G(\delta-1)}/8$. Let $C_i = C_{i,1} \ldots C_{i,t_i}$ where each $C_{i,j}$ has degree $D_{ij} < \lambda^{G(\delta-1)}/8$. If C is set-multilinear with respect to $(X_l)_{l\in S}$, then let $(S_1, \ldots, S_{t_i})$ be the partition of S such that each $C_{i,j}$ is set-multilinear with respect to $(X_l)_{l\in S_j}$.
For $j \in [t_i]$, let a_{ij} be the number of positive indices in S_j. We have two cases: If $a_{ij} \leq D_{ij}/2$, then

$$w_{S_j} \leq \frac{D_{ij}}{2} \cdot \alpha k + \frac{D_{ij}}{2} \cdot (-k) = -\frac{D_{ij} p}{2q} k \leq -\frac{D_{ij} k}{4\lambda}$$

where the last inequality follows from $\frac{p}{q} \geq \frac{c_0}{2c_0(r_0+1)} = \frac{1}{2\lambda}$. The other case is $a_{ij} > D_{ij}/2$. If we can prove that $|w_{S_j}| \geq a_{ij}k/(8\lambda^{G(\delta)-1})$,then in both of the above cases, we would have $|w_{S_j}| \geq D_{ij}k/(16\lambda^{G(\delta)-1})$. By the multiplicativity and imbalance property of relrk_w, it would follow that

$$\mathrm{relrk}_w(C_i) \leq \prod_{j=1}^{t_i} 2^{-\frac{1}{2}|w_{S_j}|} \leq 2^{-\sum_{j=1}^{t_i} D_{ij}k/(32\lambda^{G(\delta)-1})} = 2^{-Dk/(32\lambda^{G(\delta)-1})} \leq 2^{-k\lambda/256}$$

and we would be done. Thus, we now only have to show that $|w_{S_j}| \geq a_{ij}k/(8\lambda^{G(\delta)-1})$.

$$|w_{S_j}| = |a_{ij} \cdot \alpha k + (D_{ij} - a_{ij}) \cdot (-k)| = \left| a_{ij} \frac{p}{q} - (2a_{ij} - D_{ij}) \right| k \qquad \text{as } \alpha = 1 - p/q$$

$$\geq \left| \frac{a_{ij}p}{q} - \left\lfloor \frac{a_{ij}p}{q} \right\rceil \right| k \qquad \text{where } \lfloor . \rceil \text{ denotes the nearest integer}$$

The fractional part of $\dfrac{a_{ij}p}{q}$ is $\dfrac{a_{ij}p \bmod q}{q}$. Hence in order to prove that $|w_{S_j}|$, it is enough to verify that the following inequality is satisfied:

$$\min\left(\frac{a_{ij}p \bmod q}{q}, 1 - \frac{a_{ij}p \bmod q}{q}\right) \geq \frac{a_{ij}}{8\lambda^{G(\delta)-1}} \tag{3}$$

Showing that the p, q we defined satisfy the inequality (3): We will first find what we call the base $(b_0, \ldots, b_{\Delta-2})$ representation of the number a_{ij}. For $0 \le m \le \Delta - 2$, inductively define y_m to be the integer quotient when $\left(a_{ij} - \sum_{m'=m+1}^{\Delta-2} b_{m'}y_{m'}\right)$ is divided by b_m. Then we can express a_{ij} as $a_{ij} = \sum_{m=0}^{\Delta-2} b_m y_m$. Since $b_m \ge \lambda^{G(m)}/2$ for all m and $a_{ij} \le D_{ij} < \lambda^{G(\delta-1)}/8$, we have the following bounds on the values of y_m:

$$y_m = 0 \text{ for } m \ge \delta - 1, \tag{4}$$

$$y_{\delta-2} = \left\lfloor \frac{a_{ij}}{b_{\delta-2}} \right\rfloor < \frac{\frac{\lambda^{G(\delta-1)}}{8}}{\frac{\lambda^{G(\delta-2)}}{2}} \le \frac{\lambda^{G(\delta-1)-G(\delta-2)} - 1}{2} = \frac{r_{\delta-2}}{2}, \tag{5}$$

$$y_m \le \left\lfloor \frac{b_{m+1}-1}{b_m} \right\rfloor = r_m \text{ for } m < \delta - 2 \tag{6}$$

By (2), $a_{ij}p \equiv \sum_{m=0}^{\Delta-2} c_m y_m \bmod q$. Therefore,

$$\min\left(\frac{a_{ij}p \bmod q}{q}, 1 - \frac{a_{ij}p \bmod q}{q}\right) = \min\left(\left|\sum_{m=0}^{\Delta-2} c_m y_m\right|/q,\ 1 - \left|\sum_{m=0}^{\Delta-2} c_m y_m\right|/q\right) \tag{7}$$

if $\left|\sum_{m=0}^{\Delta-2} c_m y_m\right|/q \le 1$, which is true by the following claim (proved in Section A):

▷ Claim 8. If $0 \le y_m \le r_m$ for all m, then $\left|\sum_{m=0}^{\Delta-2} c_m y_m\right| < q - c_0$.

Now let f be the highest index such that $y_f \ge 1$ (by (4), $f \le \delta - 2$) and e be the smallest index such that $y_e \ge 1$. Then $\left|\sum_{m=0}^{\Delta-2} c_m y_m\right| = \left|\sum_{m=e}^{f} c_m y_m\right|$. We need two more claims whose proofs we postpone to section A.

▷ Claim 9. Let y_m be non-negative integers such that $y_e \ge 1$. Then $\left|\sum_{m=e}^{f} c_m y_m\right| \ge \min\Big(|c_f y_f|, |c_{f-1}| - |c_f y_f|\Big)$.

▷ Claim 10. Let $0 \le e \le f \le \delta - 2$. If $y_f \ge 1$, $y_{\delta-2} = \lfloor \frac{a_{ij}}{b_{\delta-2}} \rfloor \le r_{\delta-2}/2$ and $0 \le y_m \le r_m$ for all $m \le \delta - 2$, then $\min\Big(|c_f y_f|, |c_{f-1}| - |c_f y_f|\Big) \ge |c_{\delta-2}a_{ij}/(2b_{\delta-2})|$.

If $\delta = 2$, then $f = 0$ by (4). Thus $q - \left|\sum_{m=e}^{f} c_m y_m\right| > c_0 r_0 - |c_0 y_0| > c_0 r_0/2 > |c_f y_f|$ where the last two inequalities follow from (5).

Otherwise $\delta > 2$. By Claim 8, $q - \left|\sum_{m=e}^{f} c_m y_m\right| > c_0$. From the definition of the sequence $\{c_m\}$, we have $c_0 \ge |c_f r_f| \ge |c_f y_f|$ when $f > 0$. But when $f = 0$, it follows that $y_{\delta-2} = 0$ implying $a_{ij} < b_{\delta-2}$. This further implies $c_0 \ge |c_{\delta-2}| \ge |c_{\delta-2}a_{ij}/b_{\delta-2}|$.

From the analysis of the two cases above and by Claims 9 and 10, we get that

$$\min\left(\left|\sum_{m=e}^{f} c_m y_m\right|,\ q - \left|\sum_{m=e}^{f} c_m y_m\right|\right)/q \ge \left|\frac{c_{\delta-2}a_{ij}}{2b_{\delta-2}q}\right|.$$

The bounds on each b_m and $|c_m|$ given in (1) imply the following:

$$|c_{\delta-2}| \ge \lambda^{G(\Delta-1)-G(\delta-1)}/2, \quad b_{\delta-2} \le \lambda^{G(\delta-2)}, \quad q \le |c_0|\lambda + |c_1| \le 2\lambda^{G(\Delta-1)}$$

Hence $\min\left(\left|\sum_{m=e}^{f} c_m y_m\right|/q,\ 1 - \left|\sum_{m=e}^{f} c_m y_m\right|/q\right) \ge \frac{a_{ij}}{8\lambda^{G(\delta-1)+G(\delta-2)}} = \frac{a_{ij}}{8\lambda^{G(\delta)-1}}$ which together with (7) implies (3). ◀

5 Limitations on improving the bounds: Proof of Theorem 3

We will show here that the techniques used by [19] cannot hope to prove much stronger lower bounds. We do this by constructing polynomials for which the lower bound we proved earlier is tight. We begin by showing this in the case of two different set sizes. We can normalize with respect to the bigger set size to assume that the weights are $-k$ and αk ($\alpha \in [0,1]$) without loss of generality. Clearly, $k \leq \log n$.

▶ **Lemma 11.** *Let n, d, Δ be such that $d \leq n$. For any $\alpha \in [0,1]$ let $w \in \{-k, \alpha k\}^d$ be a word. There is a polynomial $P_\Delta \in \mathbb{F}_{sm}[\overline{X}(w)]$ which is computable by a set-multilinear formula of size at most $n^{O(\Delta d^{\mu(\Delta)})}$ and has the maximum possible relative rank.*

▶ Remark. We can replace αk with $\lfloor \alpha k \rceil$ and assume that the weights in w are integers. It can be shown that this will not change the arguments in any significant way (Claim 21).

We will need the extensive notation from [19]. We restate it here.

Notation.

- As in Section 2 and from the remark above, we assume $|X(w_i)| = 2^{|w_i|}$ and that the variables are indexed by binary strings $\{0,1\}^{|w_i|}$.
- Given any subset $S \subseteq [d]$, we denote by $S_+ = \{i \in S \mid w_i > 0\}$ the positive indices of S and similarly by S_-, the negative indices.
- We let $K = \sum_{i\in[d]} |w_i|$, $k_+ = \sum_{i\in S_+} |w_i|$ and $k_- = \sum_{i\in S_-} |w_i|$. We say S is $\mathcal{P}$-heavy if $k_+ \geq k_-$ and $\mathcal{N}$-heavy otherwise.
- Setting $I = [K]$, we partition the index set $I = I_1 \cup \dots I_d$ where I_j is an interval of length $|w_j|$ that starts at $\sum_{i<j} |w_j| + 1$. Given a $T \subseteq [d]$, we let $I(T) = \bigcup_{j\in T} I_j$.
- Let $m = m_+ m_- \in \mathcal{M}_w^S$ be any monomial. The indices of the string associated with the positive monomial, $\sigma(m_+)$ can be thought of as labelled by elements of $I(S_+)$ in the natural way - $\sigma(m_+) : I(S_+) \to \{0,1\}$. Similarly for $\sigma(m_-)$.

Given a set S, we define a sequence of polynomials that we will later show to have small size set multilinear formulas but large rank.

Fix $J_+ \subseteq I(S_+)$ and $J_- \subseteq I(S_-)$ such that $|J_+| = |J_-| = \min\{k_+, k_-\}$. Let π be a bijection from J_+ to J_-. Such a tuple (S, J_+, J_-, π) is called valid. Fix a valid (S, J_+, J_-, π).

A string $\tau \in \{0,1\}^{|k_+ - k_-|}$ defines a map $I(S_+) \setminus J_+ \to \{0,1\}$ if S is $\mathcal{P}$-heavy and a map $I(S_-) \setminus J_- \to \{0,1\}$ if S is $\mathcal{N}$-heavy.

The polynomial $P_{(S,J_+,J_-,\pi,\tau)}$ is the sum of all monomials m such that

1. $\sigma(m_+)(j) = \sigma(m_-)(\pi(j))$ for all $j \in J_+$, and
2. $\sigma(m_+)(j) = \tau(j)$ for all $j \in I(S_+) \setminus J_+$ if S is $\mathcal{P}$-heavy or $\sigma(m_-)(j) = \tau(j)$ for all $j \in I(S_-) \setminus J_-$ if S is $\mathcal{N}$-heavy.

These polynomials have high relative rank and a few more useful properties (as observed in [19]) that help us in building formulas for these polynomials inductively (See Section B).

To proceed, we will need a few notions that help make the ideas in the proof overview above precise. Fix Δ as in Claim 12. We define the *fractional cost* fc. Set $\mathrm{fc}(0) = 1$ and

$$\mathrm{fc}(\delta) := \min_{q < d^{\mu(\Delta)}/\mathrm{fc}(\delta-1)} |q\alpha - \lfloor q\alpha \rceil|/q \qquad \text{for } 1 \leq \delta \leq \Delta - 1$$

The quantity $|q\alpha - \lfloor q\alpha \rceil|$ is the distance to the nearest integer from $q\alpha$. For $1 \leq \delta \leq \Delta - 1$, we denote by p_δ the (least) value of q for which the above expression attains the minimum. We also denote by $n_\delta := \lfloor p_\delta \alpha \rceil$ the nearest integer to $p_\delta \alpha$. Finally, we set $p_\Delta := |\mathcal{P}_w|$ (total number of positive sets) and $n_\Delta := |\mathcal{N}_w|$ (total number of negative sets).

We state a few properties of the terms defined above (See Section B for the proof)

(C1) (Exponential decline) The fractional cost falls exponentially with depth i.e., $\mathrm{fc}(\delta) \leq 1/(d^{\mu(\Delta)})^{F(\delta+1)-2}$ for $1 \leq \delta \leq \Delta - 1$. This exponential decline causes $\mathrm{fc}(\Delta-1)$ to be very small: $\mathrm{fc}(\Delta - 1) \leq 2d^{\mu(\Delta)}/p_\Delta$. (Claim 19).

(C2) (Monotonicity) Let $\Delta' \leq \Delta - 1$ be the smallest integer for which $\mathrm{fc}(\Delta') \leq 2d^{\mu(\Delta)}/p_\Delta$ holds (such a Δ' exists from the second part of (C1)). Redefine $p_{\Delta'+1} \coloneqq p_\Delta$ and $n_{\Delta'+1} \coloneqq n_\Delta$. We have that $p_{\delta-1} \leq p_\delta$ and $n_{\delta-1} \leq n_\delta$ for all $\delta \leq \Delta' + 1$ (Claim 20).

With the notation in place ,we can now state the following central claim that constructs the polynomial needed for Lemma 11:

▷ Claim 12. Let $S \subseteq [d]$ such that $|w_S| \leq k$. Then, there exist J_+, J_-, π such that (S, J_+, J_-, π) is valid and for any integer $\delta \leq \Delta' + 1$ and for all $\tau \in \{0,1\}^{|k_+ - k_-|}$, the polynomial $P_{(S,J_+,J_-,\pi,\tau)})$ can be computed by a set-multilinear formula of product-depth δ and size at most $|S|^\delta 2^{5k\delta d^{\mu(\Delta)}}$.

We finish the proof of Lemma 11 assuming the above claim:

Proof of Lemma 11. As $w_{[d]} \leq k$, applying Claim 12 to $S = [d]$ and $\delta = \Delta' + 1$, gives a polynomial $P_{\Delta'+1} \in \mathbb{F}_{sm}[\overline{X}(w)]$ with $\mathrm{relrk}_w(P_{\Delta'+1}) = 2^{-|w_{[d]}|/2}$ (using property (P1) in Section B), and it is computable by a set-multilinear formula of product-depth at most Δ of size at most $d^\Delta 2^{10k\Delta d^{\mu(\Delta)}} \leq n^{O\left(\Delta d^{\mu(\Delta)}\right)}$, since $\Delta' + 1 \leq \Delta$ by definition. ◀

The following claim is the main technical result that helps in proving Claim 12. It is in the same spirit as [19, Claim 28], but we show the existence of a better partition with a more careful analysis. Our analysis holds for any $\alpha \in [0,1]$.

▷ Claim 13. Fix $\delta \leq \Delta' + 1$. Let $S \subseteq [d]$ with $|w_S| \leq k$ such that $|S_+| \leq p_\delta$ and $|S_-| \leq n_\delta$. Then there exists a partition of S as $S_1 \cup S_2 \cup \ldots S_r$ where the following conditions hold:

1. $|S_{i,+}| \leq p_{\delta-1}$ and $|S_{i,-}| \leq n_{\delta-1}$
2. $\sum_{i=1}^{r} |w_{S_i}| \leq 5kd^{\mu(\Delta)}$
3. $|w_{S_i}| \leq k$ for all $i \in [r]$

Proof of Claim 13. As long as possible, pick sets S_i with $|S_{i,+}| = p_{\delta-1}$ positive indices and $|S_{i,-}| = n_{\delta-1}$ negative indices. For all such sets picked, we have

$$|w_{S_i}| = \left|\sum\nolimits_{j \in S_i} w_j\right| = k \cdot |p_{\delta-1}\alpha - n_{\delta-1}| = k \cdot |p_{\delta-1}\alpha - n_{\delta-1}| \leq k \tag{8}$$

Suppose the sets chosen after the procedure are $S_1, \ldots, S_m$, where $m = \min\left\{\left\lfloor \frac{|S_+|}{p_{\delta-1}} \right\rfloor, \left\lfloor \frac{|S_-|}{n_{\delta-1}} \right\rfloor\right\}$ and we are left with the set S'. Since we cannot pick the sets any more, we must have that $|S'_+| < p_{\delta-1}$ or $|S'_-| < n_{\delta-1}$ (or both). We analyze one case, others being analogous.

Say $m = \left\lfloor \frac{|S_+|}{p_{\delta-1}} \right\rfloor$ (i.e. $|S'_+| < p_{\delta-1}$). Also suppose $|S'_-| > n_{\delta-1}$. We pick a set S_{m+1} with $|S'_+|$ positive indices and $p \leq (|S_-| - m \cdot n_{\delta-1})$ negative indices such that

$$|w_{S_{m+1}}| = k\left|\alpha|S'_+| - p\right| = k\left|\alpha(|S_+| - m \cdot p_{\delta-1}) - p\right| \leq k \tag{9}$$

Note that we can always choose $\alpha|S'_+| - 1 \leq p \leq \alpha|S'_+| + 1$ to satisfy the desired constraints. This follows from noting that $|p_{\delta-1}\alpha - n_{\delta-1}| \leq 1$ which gives $p_{\delta-1}\alpha - 1 \leq n_{\delta-1} \leq p_{\delta-1}\alpha + 1$. Now use the fact that $|S'_-| > n_{\delta-1}$.

The remaining set $T = S' \setminus S_{m+1}$ has only negative values which we split into singletons $S_{m+2}, \ldots, S_r$ (there are $(|S_-| - mn_{\delta-1} - p)$ of these sets). As these are singletons, for $m + 2 \leq j \leq r$ we trivially have $|w_{S_j}| \leq k$.

We also note that since $(|S_-| - m \cdot n_{\delta-1} - p)$ is positive, it is equal to $|m \cdot n_{\delta-1} + p - |S_-||$, which can be rewritten as $|(\alpha|S_+| - |S_-|) - (m(p_{\delta-1}\alpha - n_{\delta-1})) - (\alpha(|S_+| - m \cdot p_{\delta-1}) - p)|$.

Using the triangle inequality, we have that this quantity is at most the sum of $|\alpha|S_+| - |S_-||$, $|m(p_{\delta-1}\alpha - n_{\delta-1})|$ and $|\alpha(|S_+| - mp_{\delta-1}) - p|$. The first term is less than 1 since $|w_S| \leq k$ and the last term is less than 1 from (9). Putting it all together, we have

$$(|S_-| - m \cdot n_{\delta-1} - p) \leq |m(p_{\delta-1}\alpha - n_{\delta-1})| + 2 \tag{10}$$

Finally, we get

$$\begin{aligned}
\sum_{i=1}^{r} |w_{S_i}| &= \sum_{i=1}^{m} |w_{S_i}| + |w_{S_{m+1}}| + \sum_{i=m+2}^{r} |w_{S_i}| \\
&\leq km|p_{\delta-1}\alpha - n_{\delta-1}| + k + k(|S_-| - m \cdot n_{\delta-1} - p) \\
&\leq km|p_{\delta-1}\alpha - n_{\delta-1}| + k + k\,|m(p_{\delta-1}\alpha - n_{\delta-1})| + 2k \qquad \text{(using (10))} \\
&\leq k\left(2\left\lfloor \frac{|S_+|}{p_{\delta-1}} \right\rfloor |p_{\delta-1}\alpha - n_{\delta-1}| + 3\right) \leq k\left(2|S_+|\frac{|p_{\delta-1}\alpha - n_{\delta-1}|}{p_{\delta-1}} + 3\right) \\
&\leq k\left(2p_\delta \cdot \mathrm{fc}(\delta-1) + 3\right) \qquad \text{(By definition of fc)} \\
&\leq 5kd^{\mu(\Delta)}
\end{aligned}$$

where the last inequality is true because $\mathrm{fc}(\delta-1) \leq 2d^{\mu(\Delta)}/p_\delta$ holds for $\delta \leq \Delta'$ by the definition of fc and p_δ; it also holds for $\delta = \Delta' + 1$ by the definition of Δ'. ◀

Armed with all this, the proof of Claim 12 becomes quite similar to the proof of Claim 27 in [19]. See Section B for details.

Handling more than two weights. To handle the case when there are multiple weights, we partition the index set $[d]$ into sets $\{S_i\}$ such that the sub-word indexed by each S_i contains at most two distinct weights (See Section B for the proof). We can assume without loss of generality that all entries of w are integers as before.

▶ **Lemma 14.** *Let $w \in \{\alpha_1, \ldots, \alpha_\gamma\}^d$ ($|\alpha_i| \leq k$ for all i) be a word with $\gamma \leq d$ different weights and $|w_{[d]}| \leq k$. Then, the index set $[d]$ can be partitioned as $S_1 \cup \ldots \cup S_\eta$ with $\eta \leq 6\gamma$ such that for all $i \in [\eta]$, the sub-word $w_{|S_i}$ has at most two distinct weights and $|w_{S_i}| \leq k$.*

We can now use Claim 12 to construct polynomials with small set-multilinear formula size but large rank, even when the number of distinct set sizes is not two.

▷ Claim 15. Let $S \subseteq [d]$ and let $w \in \{\alpha_1, \ldots, \alpha_\gamma\}^d$ ($|\alpha_i| \leq k$ for all i) be a word with $\gamma \leq d$ different weights and $|w_S| \leq k$. Then, there exist $(J_+, J_-, \pi), (J'_+, J'_-, \pi')$ such that (S, J_+, J_-, π) and (S, J'_+, J'_-, π') are valid. For any fixed integer Δ and for all $\tau \in \{0,1\}^{|k_+ - k_-|}$, the polynomial $P_{(S,J_+,J_-,\pi,\tau)}$ can be computed by a set-multilinear formula of product-depth Δ and size at most $|S|^\Delta 2^{30k\gamma\Delta d^{\mu(\Delta)}}$ while the polynomial $P_{(S,J'_+,J'_-,\pi',\tau)}$ can be computed by a set-multilinear formula of product-depth Δ and size at most $|S|^\Delta 2^{5k\Delta d^{\mu(\Delta-1)}+6\gamma k}$.

The proof of Claim 15 is quite similar to that of Claim 12 and we prove it in Section B. Assuming it, we can finally prove Theorem 3:

Proof of Theorem 3. As $w_{[d]} \leq k$, applying Claim 15 to $S = [d]$, gives polynomials $P_\Delta, Q_\Delta \in \mathbb{F}_{sm}[\overline{X}(w)]$ with relative rank $\mathrm{relrk}_w(P_\Delta) = \mathrm{relrk}_w(Q_\Delta) = 2^{-|w_{[d]}|/2}$ (using property (P1) in Section B). Hence the lower bound measures are maximum.

The polynomial P_Δ has product-depth Δ set-multilinear formula size of at most

$$d^\Delta 2^{30k\gamma\Delta d^{\mu(\Delta)}} \leq n^{O\left(\gamma\Delta d^{\mu(\Delta)}\right)}$$

The polynomial Q_Δ has product-depth Δ set-multilinear formula size of at most

$$d^\Delta 2^{5k\Delta d^{\mu(\Delta-1)}+6\gamma k} \leq n^{O\left(\Delta d^{\mu(\Delta-1)}+\gamma\right)}$$

◀

References

1 Manindra Agrawal and V. Vinay. Arithmetic Circuits: A Chasm at Depth Four. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 67–75, 2008. 2

2 Walter Baur and Volker Strassen. The complexity of partial derivatives. *Theoret. Comput. Sci.*, 22(3):317–330, 1983. 2

3 Peter Bürgisser, Michael Clausen, and M. Amin Shokrollahi. *Algebraic complexity theory*, volume 315 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1997. With the collaboration of Thomas Lickteig. 1

4 Xi Chen, Neeraj Kayal, and Avi Wigderson. Partial derivatives in arithmetic complexity and beyond. *Found. Trends Theor. Comput. Sci.*, 6(1-2):front matter, 1–138 (2011), 2010. 2

5 Suryajith Chillara, Nutan Limaye, and Srikanth Srinivasan. Small-depth multilinear formula lower bounds for iterated matrix multiplication with applications. *SIAM J. Comput.*, 48(1):70–92, 2019. 2

6 Hervé Fournier, Nutan Limaye, Guillaume Malod, and Srikanth Srinivasan. Lower bounds for depth-4 formulas computing iterated matrix multiplication. *SIAM J. Comput.*, 44(5):1173–1201, 2015. 2

7 Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Approaching the chasm at depth four. In *2013 IEEE Conference on Computational Complexity—CCC 2013*, pages 65–73. IEEE Computer Soc., Los Alamitos, CA, 2013. 2

8 Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic circuits: a chasm at depth 3. *SIAM J. Comput.*, 45(3):1064–1079, 2016. 2

9 Nikhil Gupta, Chandan Saha, and Bhargav Thankey. A super-quadratic lower bound for depth four arithmetic circuits. In *35th Computational Complexity Conference*, volume 169 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. 23, 31. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2020. 2

10 K. A. Kalorkoti. A lower bound for the formula size of rational functions. *SIAM J. Comput.*, 14(3):678–687, 1985. 2

11 Neeraj Kayal. An exponential lower bound for the sum of powers of bounded degree polynomials. 2012. 2

12 Neeraj Kayal, Nutan Limaye, Chandan Saha, and Srikanth Srinivasan. An exponential lower bound for homogeneous depth four arithmetic formulas. *SIAM J. Comput.*, 46(1):307–335, 2017. 2

13 Neeraj Kayal, Chandan Saha, and Ramprasad Saptharishi. A super-polynomial lower bound for regular arithmetic formulas. In *Symposium on Theory of Computing (STOC)*. ACM - Association for Computing Machinery, June 2014. 2

14 Neeraj Kayal, Chandan Saha, and Sébastien Tavenas. An almost cubic lower bound for depth three arithmetic circuits. In *43rd International Colloquium on Automata, Languages, and Programming*, volume 55 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 33, 15. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016. 2

15 Neeraj Kayal, Chandan Saha, and Sébastien Tavenas. On the size of homogeneous and of depth-four formulas with low individual degree. *Theory Comput.*, 14:Paper No. 16, 46, 2018. 2

16 Pascal Koiran. Arithmetic circuits: the chasm at depth four gets wider. *Theoret. Comput. Sci.*, 448:56–65, 2012. 2

17 Mrinal Kumar and Shubhangi Saraf. On the power of homogeneous depth 4 arithmetic circuits. In *55th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2014*, pages 364–373. IEEE Computer Soc., Los Alamitos, CA, 2014. 2

18 Mrinal Kumar and Shubhangi Saraf. The limits of depth reduction for arithmetic formulas: it's all about the top fan-in. *SIAM J. Comput.*, 44(6):1601–1625, 2015. 2

19 Nutan Limaye, Srikanth Srinivasan, and Sébastien Tavenas. Superpolynomial Lower Bounds Against Low-Depth Algebraic Circuits. to appear in FOCS, 2021. 2, 3, 4, 5, 6, 10, 11, 12

20 Meena Mahajan. Algebraic complexity classes. In *Perspectives in computational complexity*, volume 26 of *Progr. Comput. Sci. Appl. Logic*, pages 51–75. Birkhäuser/Springer, Cham, 2014. 1
21 D. S. Mitrinović. *Analytic inequalities*. Die Grundlehren der mathematischen Wissenschaften, Band 165. Springer-Verlag, New York-Berlin, 1970. In cooperation with P. M. Vasić. 14
22 Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Comput. Complexity*, 6(3):217–234, 1995. 2
23 Ran Raz. Separation of multilinear circuit and formula size. *Theory Comput.*, 2:121–135, 2006. 2
24 Ran Raz. Multi-linear formulas for permanent and determinant are of super-polynomial size. *J. ACM*, 56(2):Art. 8, 17, 2009. 2
25 Ran Raz. Elusive functions and lower bounds for arithmetic circuits. *Theory Comput.*, 6:135–177, 2010. 2
26 Ramprasad Saptharishi. A survey of lower bounds in arithmetic circuit complexity. *Github Survey*, 2015. 2
27 Wolfgang M. Schmidt. *Diophantine approximations and Diophantine equations*, volume 1467 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1991. 6, 19
28 Victor Shoup and Roman Smolensky. Lower bounds for polynomial evaluation and interpolation problems. *Comput. Complexity*, 6(4):301–311, 1996/97. 2
29 Amir Shpilka and Avi Wigderson. Depth-3 arithmetic circuits over fields of characteristic zero. *Comput. Complexity*, 10(1):1–27, 2001. 2
30 Amir Shpilka and Amir Yehudayoff. Arithmetic circuits: a survey of recent results and open questions. *Found. Trends Theor. Comput. Sci.*, 5(3-4):207–388 (2010), 2009. 1
31 Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. In *Mathematical foundations of computer science 2013*, volume 8087 of *Lecture Notes in Comput. Sci.*, pages 813–824. Springer, Heidelberg, 2013. 2
32 Sébastien Tavenas, Nutan Limaye, and Srikanth Srinivasan. Set-multilinear and non-commutative formula lower bounds for iterated matrix multiplication. to appear in STOC 2022. 2
33 L. G. Valiant. Completeness classes in algebra. In *Conference Record of the Eleventh Annual ACM Symposium on Theory of Computing (Atlanta, Ga., 1979)*, pages 249–261. ACM, New York, 1979. 1
34 L. G. Valiant, S. Skyum, S. Berkowitz, and C. Rackoff. Fast parallel computation of polynomials using few processors. *SIAM J. Comput.*, 12(4):641–644, 1983. 2

A Proofs of Section 4: Lower bound

In the following lemmas, let the sequences $\{b_m\}, \{c_m\}, \{r_m\}$ be as defined in section 4.

To prove bounds on each b_m and $|c_m|$, we use a generalized version of the well known Bernoulli's inequality [21, Section 2.4]:

▷ Claim 16 (Bernoulli's inequality). Let $x_1, \ldots, x_r$ be real numbers all greater than -1 and all with the same sign. Then,

$$(1 + x_1)(1 + x_2) \ldots (1 + x_r) \geq 1 + x_1 + \ldots + x_r \,.$$

▶ **Lemma 17.** *Let $\lambda \geq 2$ be as defined in Section 4. Then For $0 \leq m \leq \Delta - 2$, $\frac{\lambda^{G(m)}}{2} \leq b_m \leq \lambda^{G(m)}$ and $\frac{\lambda^{G(\Delta-1)-G(m+1)}}{2} \leq |c_m| \leq \lambda^{G(\Delta-1)-G(m+1)}$.*

Proof. Clearly, b_m satisfies the bounds when $m = 0$ or 1. For $m \geq 2$,

$$b_m = (\lambda^{G(m)-G(m-1)} - 1)b_{m-1} + b_{m-2}$$

$$
\begin{aligned}
&\le \lambda^{G(m)-G(m-1)} b_{m-1} \\
&\le \lambda^{G(m)-G(m-1)} . \lambda^{G(m-1)-G(m-2)} \dots \lambda^{G(2)-G(1)} b_1 \\
&= \lambda^{G(m)}
\end{aligned}
$$

$$
\begin{aligned}
b_m &= (\lambda^{G(m)-G(m-1)} - 1) b_{m-1} + b_{m-2} \\
&\ge (\lambda^{G(m)-G(m-1)} - 1) b_{m-1} \\
&\ge (\lambda^{G(m)-G(m-1)} - 1).(\lambda^{G(m-1)-G(m-2)} - 1) \dots (\lambda^{G(2)-G(1)} - 1) b_1 \\
&= \lambda^{G(m)-G(1)} b_1 . \left(1 - \frac{1}{\lambda^{G(m)-G(m-1)}}\right)\left(1 - \frac{1}{\lambda^{G(m-1)-G(m-2)}}\right) \dots \left(1 - \frac{1}{\lambda^{G(2)-G(1)}}\right) \\
&\ge \lambda^{G(m)} . \left(1 - \frac{1}{\lambda^{G(m)-G(m-1)}} - \frac{1}{\lambda^{G(m-1)-G(m-2)}} - \dots - \frac{1}{\lambda^{G(2)-G(1)}}\right) \quad \text{[By Claim 16]} \\
&\ge \lambda^{G(m)} . \left(1 - \frac{1}{\lambda^{m-1}} - \frac{1}{\lambda^{m-2}} - \dots - \frac{1}{\lambda}\right) \\
&= \lambda^{G(m)} . \left(1 - \frac{1}{\lambda - 1}\left(1 - \frac{1}{\lambda^{m-1}}\right)\right) \ge \frac{\lambda^{G(m)}}{2}
\end{aligned}
$$

Clearly, $|c_m|$ satisfies the bounds when $m = \Delta - 2$ or $\Delta - 3$. For $m \le \Delta - 4$,

$$
\begin{aligned}
|c_m| &= (\lambda^{G(m+2)-G(m+1)} - 1)|c_{m+1}| + |c_{m+2}| \\
&\le \lambda^{G(m+2)-G(m+1)} |c_{m+1}| \\
&\le \lambda^{G(m+2)-G(m+1)} \cdot \lambda^{G(m+3)-G(m+2)} \dots \lambda^{G(\Delta-2)-G(\Delta-3)} |c_{\Delta-3}| \\
&= \lambda^{G(\Delta-2)-G(m+1)} \cdot \lambda^{G(\Delta-1)-G(\Delta-2)} = \lambda^{G(\Delta-1)-G(m+1)}
\end{aligned}
$$

$$
\begin{aligned}
|c_m| &= (\lambda^{G(m+2)-G(m+1)} - 1)|c_{m+1}| + |c_{m+2}| \\
&\ge (\lambda^{G(m+2)-G(m+1)} - 1)|c_{m+1}| \\
&\ge (\lambda^{G(m+2)-G(m+1)} - 1) \cdot (\lambda^{G(m+3)-G(m+2)} - 1) \dots (\lambda^{G(\Delta-2)-G(\Delta-3)} - 1)|c_{\Delta-3}| \\
&= \lambda^{G(\Delta-2)-G(m+1)} |c_{\Delta-3}| \cdot \left(1 - \frac{1}{\lambda^{G(m+2)-G(m+1)}}\right)\left(1 - \frac{1}{\lambda^{G(m+3)-G(m+2)}}\right) \dots \\
&\qquad \dots \left(1 - \frac{1}{\lambda^{G(\Delta-2)-G(\Delta-3)}}\right) \\
&\ge \lambda^{G(\Delta-2)-G(m+1)} |c_{\Delta-3}| . \left(1 - \frac{1}{\lambda^{G(m+2)-G(m+1)}} - \dots - \frac{1}{\lambda^{G(\Delta-2)-G(\Delta-3)}}\right) \\
&\qquad \text{[By Claim 16]} \\
&\ge \lambda^{G(\Delta-2)-G(m+1)} |c_{\Delta-3}| . \left(1 - \frac{1}{\lambda^{m+1}} - \frac{1}{\lambda^{m+2}} - \dots - \frac{1}{\lambda^{\Delta-3}}\right) \\
&= \lambda^{G(\Delta-1)-G(m+1)} . \left(1 - \frac{1}{\lambda^m(\lambda - 1)}\left(1 - \frac{1}{\lambda^{\Delta-3-m}}\right)\right) \ge \frac{\lambda^{G(\Delta-1)-G(m+1)}}{2}
\end{aligned}
$$

◀

▷ Claim 8. If $0 \le y_m \le r_m$ for all m, then $\left|\sum_{m=0}^{\Delta-2} c_m y_m\right| < q - c_0$.

Proof.

$$
\sum_{m=0}^{\Delta-2} c_m y_m = \sum_{m=0}^{\lfloor \frac{\Delta-2}{2} \rfloor} c_{2m} y_{2m} + \sum_{m=1}^{\lceil \frac{\Delta-2}{2} \rceil} c_{2m-1} y_{2m-1}
$$

where the first summand is ≥ 0 and the second summand is ≤ 0 as c_i takes positive values at even indices and negative values at odd indices. Hence $\left|\sum_{m=0}^{\Delta-2} c_m y_m\right|$ is upper bounded by the maximum of the absolute values of these two summands.

$$\left|\sum_{m=0}^{\lfloor\frac{\Delta-2}{2}\rfloor} c_{2m}y_{2m}\right| \leq \left|\sum_{m=0}^{\lfloor\frac{\Delta-2}{2}\rfloor} c_{2m}r_{2m}\right| = \left|c_0r_0 - c_1 + \left(c_1 + \sum_{m=1}^{\lfloor\frac{\Delta-2}{2}\rfloor} c_{2m}r_{2m}\right)\right|$$

$$\text{and } \left|\sum_{m=1}^{\lceil\frac{\Delta-2}{2}\rceil} c_{2m-1}y_{2m-1}\right| \leq \left|\sum_{m=1}^{\lceil\frac{\Delta-2}{2}\rceil} c_{2m-1}r_{2m-1}\right| = \left|-c_0 + \left(c_0 + \sum_{m=1}^{\lceil\frac{\Delta-2}{2}\rceil} c_{2m-1}r_{2m-1}\right)\right|$$

By repeated substitution of the form $c_m + c_{m+1}r_{m+1} = c_{m+2}$, the first equation becomes equal to $(c_0r_0 - c_1) + c_{2\lfloor\frac{\Delta-2}{2}\rfloor+1}$ and the second equation becomes equal to $\left|-c_0 + c_{2\lceil\frac{\Delta-2}{2}\rceil}\right| = c_0 - c_{2\lceil\frac{\Delta-2}{2}\rceil}$ [We might need to define $c_{\Delta-1} := c_{\Delta-2}r_{\Delta-2} + c_{\Delta-3}$ for this as we have not defined it earlier. It's easy to see that the sign parity of $c_{\Delta-1}$ will be $(-1)^{\Delta-1}$].
Finally,

$$(c_0r_0 - c_1) + c_{2\lfloor\frac{\Delta-2}{2}\rfloor+1} < q - c_0 \qquad \text{as } q - c_0 = c_0r_0 - c_1 \text{ and } c_{2\lfloor\frac{\Delta-2}{2}\rfloor+1} \text{ is negative;}$$

$$c_0 - c_{2\lceil\frac{\Delta-2}{2}\rceil} < q - c_0 \qquad \text{as } q - c_0 = c_0r_0 - c_1 > c_0r_0 > c_0 \text{ and } c_{2\lceil\frac{\Delta-2}{2}\rceil} \text{ is positive.}$$

◀

We will need the following lemma for proving Claim 9.

▶ **Lemma 18.** *Let $z_e, \ldots, z_f$ be integers with $0 \leq z_m \leq r_m$ $\forall m$ and $f \geq e+2$. Also let Y be an integer of the same sign as c_e such that $|Y| \geq |c_e|$. Then there exists an integer Y' of the same sign as c_{e+2} such that $|Y'| \geq |c_{e+2}|$ and*

$$|Y + c_ez_e + \sum_{m=e+1}^{f} c_mz_m| = |Y' + c_{e+2}z_{e+2} + \sum_{m=e+3}^{f} c_mz_m|$$

Proof.

$$\begin{aligned}
&|Y + c_ez_e + \sum_{m=e+1}^{f} c_mz_m| \\
=&|(Y - c_e) + c_ez_e + (c_e + c_{e+1}r_{e+1}) - c_{e+1}(r_{e+1} - z_{e+1}) + \sum_{m=e+2}^{f} c_mz_m| \\
=&|(Y - c_e) + c_ez_e + c_{e+2} - c_{e+1}(r_{e+1} - z_{e+1}) + \sum_{m=e+2}^{f} c_mz_m| \\
=&|Y' + c_{e+2}z_{e+2} + \sum_{m=e+3}^{f} c_mz_m| \qquad \text{where } Y' = (Y - c_e) + c_ez_e + c_{e+2} - c_{e+1}(r_{e+1} - z_{e+1})
\end{aligned}$$

Each of the terms $(Y - c_e)$, c_ez_e, c_{e+2} and $-c_{e+1}(r_{e+1} - z_{e+1})$ is either zero or has the same sign as c_{e+2} because

1. Y and c_e are of the same sign and $|Y| \geq |c_e|$
2. $z_{e+1} \leq r_{e+1}$
3. $c_e, -c_{e+1}$ and c_{e+2} have the same sign

Hence $Y' = (Y - c_e) + c_ez_e + c_{e+2} - c_{e+1}(r_{e+1} - z_{e+1})$ has the same sign as c_{e+2} and

$$|Y'| = |Y - c_e| + |c_ez_e| + |c_{e+2}| + |-c_{e+1}(r_{e+1} - z_{e+1})| \geq |c_{e+2}|.$$

◀

▷ Claim 9. Let y_m be non-negative integers such that $y_e \geq 1$. Then $\left|\sum_{m=e}^{f} c_m y_m\right| \geq \min\left(|c_f y_f|, |c_{f-1}| - |c_f y_f|\right)$.

Proof.

- If $e = f$, then

$$\left|\sum_{m=e}^{f} c_m y_m\right| = |c_f y_f|$$

- If $e = f - 1$, then

$$\begin{aligned}\left|\sum_{m=e}^{f} c_m y_m\right| &= |c_f y_f + c_{f-1} y_{f-1}| \geq |c_{f-1} y_{f-1}| - |c_f y_f| \\ &\geq |c_{f-1}| - |c_f y_f| \qquad [\because y_{f-1} = y_e \geq 1]\end{aligned}$$

- If $f - e \geq 2$ and $f - e$ is even, then

$$\begin{aligned}\left|\sum_{m=e}^{f} c_m y_m\right| &= \left|Y + c_e(y_e - 1) + \sum_{m=e+1}^{f} c_m y_m\right| \text{ where } Y = c_e \\ &= |Y' + c_f y_f| \text{ where } Y' \text{ has the same sign as } c_f \\ &\qquad \text{[By repeated application of Lemma 18]} \\ &\geq |c_f y_f|\end{aligned}$$

- If $f - e \geq 2$ and $f - e$ is odd, then

$$\begin{aligned}\left|\sum_{m=e}^{f} c_m y_m\right| &= \left|Y + c_e(y_e - 1) + \sum_{m=e+1}^{f} c_m y_m\right| \text{ where } Y = c_e \\ &= |Y' + c_{f-1} y_{f-1} + c_f y_f| \text{ where } Y' \text{ has the same sign as } c_{f-1} \\ &\qquad \text{and } |Y'| \geq |c_{f-1}| \\ &\qquad \text{[By repeated application of Lemma 18]} \\ &\geq |Y' + c_{f-1} y_{f-1}| - |c_f y_f| \\ &\geq |Y'| - |c_f y_f| \\ &\geq |c_{f-1}| - |c_f y_f|\end{aligned}$$

Hence in all four cases, $\left|\sum_{m=e}^{f} c_m y_m\right| \geq \min\left(|c_f y_f|, |c_{f-1}| - |c_f y_f|\right)$. ◀

▷ Claim 10. Let $0 \leq e \leq f \leq \delta - 2$. If $y_f \geq 1$, $y_{\delta-2} = \lfloor \frac{a_{ij}}{b_{\delta-2}} \rfloor \leq r_{\delta-2}/2$ and $0 \leq y_m \leq r_m$ for all $m \leq \delta - 2$, then $\min\left(|c_f y_f|, |c_{f-1}| - |c_f y_f|\right) \geq |c_{\delta-2} a_{ij}/(2b_{\delta-2})|$.

Proof. If $f = \delta - 2$ i.e. $y_{\delta-2} \geq 1$, then

$$\begin{aligned}|c_f y_f| &= |c_{\delta-2} y_{\delta-2}| \text{ and} \\ |c_{f-1}| - |c_f y_f| = |c_{\delta-3}| - |c_{\delta-2} y_{\delta-2}| &\geq |c_{\delta-3}| - \left|c_{\delta-2} \frac{r_{\delta-2}}{2}\right| \geq \left|c_{\delta-2} \frac{r_{\delta-2}}{2}\right| \geq |c_{\delta-2} y_{\delta-2}|\end{aligned}$$

where the the second inquality follows from the definition of the sequence $\{c_m\}$. As $y_{\delta-2} \geq 1$, we obtain $|c_{\delta-2} y_{\delta-2}| = \left|c_{\delta-2} \left\lfloor \frac{a_{ij}}{b_{\delta-2}} \right\rfloor\right| \geq \left|\frac{c_{\delta-2} a_{ij}}{2b_{\delta-2}}\right|$.

Otherwise if $f < \delta - 2$ i.e. $y_{\delta-2} = 0$ i.e. $a_{ij} < b_{\delta-2}$, then

$|c_f y_f| \geq |c_f| \geq |c_{\delta-2}|$ and

$|c_{f-1}| - |c_f y_f| \geq |c_{f-1}| - |c_f r_f| = |c_{f+1}| \geq |c_{\delta-2}|$

As $a_{ij} < b_{\delta-2}$, we get $|c_{\delta-2}| > \left|\frac{c_{\delta-2}a_{ij}}{b_{\delta-2}}\right|$.

Hence in both the cases, $\min\Big(|c_f y_f|, |c_{f-1}| - |c_f y_f|\Big) \geq |c_{\delta-2}a_{ij}/(2b_{\delta-2})|$. ◀

B Proofs of Section 5: Upper bound

We state some properties of the polynomials we defined in Section 5.

(P1) For any valid (S, J_+, J_-, π) and any $\tau \in \{0,1\}^{|k_+ - k_-|}$ the matrix $M_{w_{|S}}(P_{(S,J_+,J_-,\pi,\tau)})$ has the maximum possible rank for a matrix with its dimensions:

$$\mathrm{rank}(M_{w_{|S}}(P_{(S,J_+,J_-,\pi,\tau)})) = \min\{|\ \mathcal{M}_w^{\mathcal{P}\cap S}\ |, |\ \mathcal{M}_w^{\mathcal{N}\cap S}\ |\} = 2^{\min\{k_+,k_-\}}$$

(P2) Let $(S_i, J_{i,+}, J_{i,-}, \pi_i)$ $(i \in [r])$ be valid tuples with $S_i (i \in [r])$ being all $\mathcal{P}$-heavy and pairwise disjoint. Also assume that we have $\tau_i \in \{0,1\}^{k_{i,+} - k_{i,-}}$ where $k_{i,+} = \sum_{j \in I(S_{i,+})} w_j$. We can construct a new polynomial using these. Let $S = \bigcup_i S_i$ (also $\mathcal{P}$-heavy by definition), $J_+ = \bigcup_i J_{i,+}$, $J_- = \bigcup_i J_{i,-}$, $\pi = \bigcup_i \pi_i$ and $\tau = \bigcup_i \tau_i$. Then, (S, J_+, J_-, π) is a valid tuple and moreover

$$P_{(S,J_+,J_-,\pi,\tau)} = \prod_{i=1}^{r} P_{(S_i,J_{i,+},J_{i,-},\pi_i,\tau_i)}$$

If each S_i is $\mathcal{N}$-heavy, an analogous fact can be shown to hold.

(P3) Say S', S'' are disjoint sets where S' is $\mathcal{P}$-heavy and S'' is $\mathcal{N}$-heavy. Also fix any valid (S', J'_+, J'_-, π') and $(S'', J''_+, J''_-, \pi'')$.
Assume that $S = S' \cup S''$ is $\mathcal{P}$-heavy. Let $J_- = I(S_-)$ and $J_+ = J'_+ \cup J''_+ \cup J'''$ where $J''' \subseteq I(S'_+)$ is any set of size $|I(S''_-)| - |I(S''_+)|$ disjoint from $J'_+ \cup J''_+$ (As S is $\mathcal{P}$-heavy, a set like this exists). Fix any bijection $\pi''' : J''' \to I(S''_-) \setminus J''_-$ Assume $\pi : J_+ \to J_-$ is defined to be $(\pi \cup \pi'' \cup \pi''')(j)$ for $j \in J'_+ \cup J''_+ \cup J'''$
Also, fix any $\tau : I(S_+) \setminus J_+ \to \{0,1\}$. Any $\tau' : I(S'_+) \setminus J'_+ \to \{0,1\}$ is said to extend τ if τ' restricts to τ on the set $I(S_+) \setminus J_+$ (note that J_+ contains $J''_+ = I(S''_+)$ and hence $I(S_+) \setminus J_+ \subseteq I(S'_+) \setminus J'_+$, so this definition makes sense). We denote by $\tau' \setminus \tau$ the restriction of τ' to the set J'''. We thus obtain

$$P_{(S,J_+,J_-,\pi,\tau)} = \sum_{\tau' \text{ extends } \tau} P_{(S',J'_+,J'_-,\pi',\tau')} \cdot P_{(S'',J''_+,J''_-,\pi'',(\tau'\setminus\tau)\circ\pi'''^{-1})}$$

The size of this sum is $2^{|J'''|} - 2^{k''_- - k''_+}$. An analogous identity holds in the case that S is $\mathcal{N}$-heavy.

For the rest of this section, Δ will refer to the same integer as in that Section 5. We now prove some properties of the notions introduced in that section.

▷ Claim 19. (Property (C1)) The fractional cost falls exponentially with depth i.e., $\mathrm{fc}(\delta) \leq 1/(d^{\mu(\Delta)})^{F(\delta+1)-2}$ for $1 \leq \delta \leq \Delta - 1$. Also, $\mathrm{fc}(\Delta - 1) \leq 2d^{\mu(\Delta)}/p_\Delta$.

Proof. The second part of the claim is true for all $\delta < \Delta - 1$ by definition. We show it for $\Delta - 1$.

For any $\delta \le \Delta$, using Dirichlet's approximation principle ([27]), we get that there exists an integer $q' \le d^{\mu(\Delta)}/\mathrm{fc}(\delta-1)$ such that

$$|q'\alpha - \lfloor q'\alpha \rceil| < \mathrm{fc}(\delta-1)/d^{\mu(\Delta)} \tag{11}$$

We claim that the q' obtained from Dirichlet isn't too small:

$$q' \ge d^{\mu(\Delta)}/\mathrm{fc}(\delta-2) \tag{12}$$

Indeed if not, then

$$\begin{aligned} a_{\delta-1} &= \min_{q<d^{\mu(\Delta)}/\mathrm{fc}(\delta-2)} \frac{|q\alpha - \lfloor q\alpha \rceil|}{q} \\ &\le \mathrm{fc}(\delta-1)/d^{\mu(\Delta)}. && \text{from (11) and } q' \text{ is now a candidate} \end{aligned}$$

This leads to a contradiction since $d^{\mu(\Delta)} > 1$. So $q' \ge d^{\mu(\Delta)}/\mathrm{fc}(\delta-2)$ and we obtain the following bound on $\mathrm{fc}(\delta)$ using (11) and (12):

$$\mathrm{fc}(\delta) \le \frac{|q'\alpha - \lfloor q'\alpha \rceil|}{q'} \le \frac{\mathrm{fc}(\delta-1)}{d^{\mu(\Delta)}} \cdot \frac{\mathrm{fc}(\delta-2)}{d^{\mu(\Delta)}} \tag{13}$$

Solving (13) readily gives

$$\mathrm{fc}(\delta) \le \frac{1}{d^{f(\delta)\mu(\Delta)}} \qquad \text{where } f(i) \ge f(i-1) + f(i-2) + 2 \tag{14}$$

Rearranging, we have $f(i)+1 \ge (f(i-1)+1) + (f(i-2)+1) + 1$ whence we see that setting $f(i-1)+1 \coloneqq F(i)-1$ satisfies the required constraints.

This also proves the first part of the claim.

As $\mu(\Delta) = \frac{1}{F(\Delta)-1} = \frac{1}{f(\Delta-1)+1}$ this implies $f(\Delta-1)\mu(\Delta) \ge 1-\mu(\Delta)$ from which we obtain

$$\mathrm{fc}(\Delta-1) \le 1/d^{f(\Delta-1)\mu(\Delta)} \le 1/d^{1-\mu(\Delta)} = d^{\mu(\Delta)}/d \le 2d^{\mu(\Delta)}/p_\Delta \tag{15}$$

where the last inequality follows since $d \ge p_\Delta/2$. Hence the first part of the claim holds for $\Delta - 1$ as well. ◀

▷ Claim 20. (Property (C2)) For all $\delta \le \Delta' + 1$, $p_{\delta-1} \le p_\delta$ and $n_{\delta-1} \le n_\delta$.

Proof. Consider any $\delta < \Delta'$. From the definition, we know that

$$p_\delta < d^{\mu(\Delta)}/\mathrm{fc}(\delta-1) \qquad \text{and} \qquad p_{\delta-1} < d^{\mu(\Delta)}/\mathrm{fc}(\delta-2)$$

Using Dirichlet, we get an integer $\frac{d^{\mu(\Delta)}}{\mathrm{fc}(\delta-2)} \le q' < \frac{d^{\mu(\Delta)}}{\mathrm{fc}(\delta-1)}$ such that $|q'\alpha - \lfloor q'\alpha \rceil| \le \mathrm{fc}(\delta-1)/d^{\mu(\Delta)}$.

We claim that $p_\delta \ge d^{\mu(\Delta)}/\mathrm{fc}(\delta-2)$. When $p_\delta \ge q'$, this follows from above.

If $p_\delta < q'$, we claim that $|p_\delta\alpha - \lfloor p_\delta\alpha \rceil| \le \mathrm{fc}(\delta-1)/d^{\mu(\Delta)}$. Suppose not. We have,

$$\frac{|p_\delta\alpha - \lfloor p_\delta\alpha \rceil|}{p_\delta} > \frac{\mathrm{fc}(\delta-1)}{d^{\mu(\Delta)}p_\delta}.$$

But then

$$\frac{|q'\alpha - \lfloor q'\alpha \rceil|}{q'} \le \frac{\mathrm{fc}(\delta-1)}{d^{\mu(\Delta)}q'} \le \frac{\mathrm{fc}(\delta-1)}{d^{\mu(\Delta)}p_\delta} < \frac{|p_\delta\alpha - \lfloor p_\delta\alpha \rceil|}{p_\delta}$$

which is a contradiction to the definition of p_δ.

Now, if $p_\delta < d^{\mu(\Delta)}/\mathrm{fc}(\delta-2)$

$$\mathrm{fc}(\delta-1) = \min_{q<d^{\mu(\Delta)}/\mathrm{fc}(\delta-2)} \frac{|q\alpha - \lfloor q\alpha \rceil|}{q} \le \frac{|p_\delta\alpha - \lfloor p_\delta\alpha \rceil|}{p_\delta} \le \mathrm{fc}(\delta-1)/d^{\mu(\Delta)}$$

which is a contradiction.

In either case, $p_\delta \geq d^{\mu(\Delta)}/\mathrm{fc}(\delta-2) > p_{\delta-1}$. Thus we also have $p_{\delta-1}\alpha \leq p_\delta\alpha$ which implies $n_{\delta-1} = \lfloor p_{\delta-1}\alpha\rceil \leq \lfloor p_\delta\alpha\rceil = n_\delta$.

Now consider the case when $\delta = \Delta'+1$: We have that $p_{\Delta'+1} = p_\Delta$ and $n_{\Delta'+1} = n_\Delta$ We know that $\mathrm{fc}(\Delta'-1) > 2d^{\mu(\Delta)}/p_\Delta$ which implies

$$p_{\Delta'} < d^{\mu(\Delta)}/\mathrm{fc}(\Delta'-1) < p_\Delta/2 = p_{\Delta'+1}/2 \tag{16}$$

This means $|p_{\Delta'+1}\alpha - p_{\Delta'}\alpha| \geq p_{\Delta'+1}\alpha/2$.

Suppose there was an integer between $p_\Delta\alpha$ and $p_{\Delta'}$. As $p_{\Delta'}\alpha < p_\Delta\alpha$, this forces $n_{\Delta'} \leq n_{\Delta'+1} = n_\Delta$ and we're done.

But if $|p_\Delta\alpha - p_{\Delta'}\alpha| \leq 1$, along with (16), we get that $p_{\Delta'+1}\alpha/2 \leq 1$. So we have $p_\Delta\alpha \leq 2$ and also $n_\Delta \leq 3$ since $|p_\Delta\alpha - n_\Delta| \leq 1$. The total monomials in the original polynomial then is $n^{p_\Delta\alpha+n_\Delta} \leq n^5$ which is not the case as it would already have a small sized formula. ◀

The following claim shows that when the entries of the word w are not integers, we can still take a word w' with integer entries such that the small sized formula maximizing the relative rank for w' also nearly maximizes it for w. By "nearly maximizes", we mean that it differs from the maximum attainable relative rank by at most a factor of 2^d, which isn't much since $d = o(\log n)$.

▷ Claim 21. Let $S \subseteq [d]$ and let $w \in \{\alpha_1 k, \ldots, \alpha_\gamma k, -\beta_1 k, \ldots, -\beta_{\gamma'} k\}^d$,($|\alpha_i|, |\beta_i| \leq 1$ for all i) be a word with $\gamma \leq d$ different weights. Consider the word w' where every $\alpha_i k$ of w is replaced by $\lfloor k\alpha_i \rfloor$ and every $-\beta_j k$ of w is replaced by $-\lfloor \beta_j k \rfloor$. Let P' be the polynomial obtained using Claim 15 for the word w'. Then, $\mathrm{relrk}_w(P') \geq 2^{-d} 2^{-|w_{[d]}|/2}$.

Proof. From the definition of w', we have $|w'_i| \leq |w_i| \leq |w'_i| + 1$. Hence $\sum_i(|w_i| - |w'_i|) \leq d$. Using the definition of relative rank and noting that $\mathrm{rank}(\mathcal{M}_w(P')) = \mathrm{rank}(\mathcal{M}_{w'}(P'))$,

$$\mathrm{relrk}_w(P')/\mathrm{relrk}_{w'}(P') = \ = \frac{1}{2^{\sum_i(|w_i|-|w'_i|)/2}} \geq 2^{-d/2}.$$

As P' is the polynomial obtained using Claim 15 for the word w', we have

$$\mathrm{relrk}_{w'}(P') = 2^{-|w'_{[d]}|/2}.$$

Thus it suffices to show that $|w'_{[d]}| \leq |w_{[d]}| + d$.

By triangle inequality, $|\sum_i w'_i| \leq |\sum_i w_i| + |\sum_i w'_i - w_i|$ which implies

$$|w'_{[d]}| \leq |w_{[d]}| + \left|\sum_i w_i - w'_i\right| \leq |w_{[d]}| + \sum_i |w_i| - |w'_i| \leq |w_{[d]}| + d$$

where the second inequality holds because $|w_i| \geq |w'_i|$ for all i. ◀

We now prove the claims that build the required polynomials for Section 5.

▷ Claim 12. Let $S \subseteq [d]$ such that $|w_S| \leq k$. Then, there exist J_+, J_-, π such that (S, J_+, J_-, π) is valid and for any integer $\delta \leq \Delta'+1$ and for all $\tau \in \{0,1\}^{|k_+-k_-|}$, the polynomial $P_{(S,J_+,J_-,\pi,\tau)})$ can be computed by a set-multilinear formula of product-depth δ and size at most $|S|^\delta 2^{5k\delta d^{\mu(\Delta)}}$.

Proof. The proof is by induction on the product-depth δ for all $\delta \leq \Delta'+1$ where $\Delta'+1$ is as defined in property (C2) above.

- **Base Case**: When $\delta = 1$, we use the trivial expression for $P_{(S,J_+,J_-,\pi,\tau)}$ as sum of monomials. This is a product-depth one $\sum\prod$ set-multilinear formula of size at most $2^{kd} + 1 \leq |S|2^{5kd}$. So the claim is true in the base case.

- **Induction step**: Consider some $\delta > 1$. Let $k_+ := |I(S_+)|$ and $k_- := |I(S_-)|$. Without loss of generality, we can assume S is $\mathcal{P}$-heavy. Using Claim 13, we obtain a partition of $S = S_1 \cup \ldots \cup S_r$ where for all $i \in [r]$, $|w_{S_i}| \leq k$ and

$$\sum_{i=1}^{r} |w_{S_i}| \leq 5kd^{\mu(\Delta)} \tag{17}$$

By induction hypothesis, there exist $J_{i,+}, J_{i,-}, \pi_i$ such that $(S_i, J_{i,+}, J_{i,-}, \pi_i)$ are valid tuples and for each $\tau_i \in \{0,1\}^{|k_{i,+}-k_{i,-}|}$, the polynomial $P_{(S_i,J_{i,+},J_{i,-},\pi_i,\tau_i)}$ has a set-multilinear formula F_{i,τ_i} of product-depth $\delta - 1$ and size $s_i \leq |S_i|^{\delta-1} 2^{5k(\delta-1)d^{\mu(\Delta)}}$.
We can assume that $S_1, \ldots, S_\gamma$ are $\mathcal{P}$-heavy and $S_{\gamma+1}, \ldots, S_r$ are $\mathcal{N}$-heavy. Using (P2) above, we get that

$$P_{(S',J'_+,J'_-,\pi',\tau')} = \prod_{i=1}^{\gamma} P_{(S_i,J_{i,+},J_{i,-},\pi_i,\tau_i)}\ ,\ P_{(S'',J''_+,J''_-,\pi'',\tau'')} = \prod_{i=\gamma+1}^{r} P_{(S_i,J_{i,+},J_{i,-},\pi_i,\tau_i)} \tag{18}$$

where

$$(S', J'_+, J'_-, \pi') = \left(\bigcup_{i\in[\gamma]} S_i, \bigcup_{i\in[\gamma]} J_{i,+}, \bigcup_{i\in[\gamma]} J_{i,-}, \bigcup_{i\in[\gamma]} \pi_i \right)$$
$$(S'', J''_+, J''_-, \pi'') = \left(\bigcup_{i=\gamma+1}^{r} S_i, \bigcup_{i=\gamma+1}^{r} J_{i,+}, \bigcup_{i=\gamma+1}^{r} J_{i,-}, \bigcup_{i=\gamma+1}^{r} \pi_i \right)$$

and for $i \in [\gamma]$, each τ_i is a restriction of τ' to $I(S_{i,+}) \setminus J_{i,+}$ whereas for $i \in \{\gamma+1, \ldots, r\}$, each τ_i is a restriction of τ'' to $I(S_{i,-}) \setminus J_{i,+}$.
Note that both these tuples are valid and S' is $\mathcal{P}$-heavy and S'' is $\mathcal{N}$-heavy. Then using (P3), we construct the polynomial

$$\begin{aligned} P_{(S,J_+,J_-,\pi,\tau)} &= \sum_{\tau' \text{ extends } \tau} P_{(S',J'_+,J'_-,\pi',\tau')} \cdot P_{(S'',J''_+,J''_-,\pi'',\tau'')} \\ &= \sum_{\tau' \text{ extends } \tau} \prod_{i=1}^{r} P_{(S_i,J_{i,+},J_{i,-},\pi_i,\tau_i)} \end{aligned} \tag{19}$$

where (S', J'_+, J'_-, π') and $(S'', J''_+, J''_-, \pi'')$ are constructed as in (P3). We can now use the formulas F_{i,τ_i} we had before from induction and construct a set-multilinear product-depth δ formula for $P_{(S,J_+,J_-,\pi,\tau)}$ of size at most

$$\begin{aligned} r \cdot 2^{|k''_- - k''_+|} \cdot \max_{i\in[r]} s_i &\leq |S| \cdot 2^{\sum_i |w_{S_i}|} \cdot |S_i|^{\delta-1} 2^{5k(\delta-1)d^{\mu(\Delta)}} \\ &\leq |S| \cdot 2^{5kd^{\mu(\Delta)}} \cdot |S|^{\delta-1} 2^{5k(\delta-1)d^{\mu(\Delta)}} \\ &\leq |S|^{\delta} 2^{5k\delta d^{\mu(\Delta)}} \end{aligned} \tag{20}$$

where the second inequality follows from Lemma 13.

◀

▶ **Lemma 14.** *Let $w \in \{\alpha_1, \ldots, \alpha_\gamma\}^d$ ($|\alpha_i| \leq k$ for all i) be a word with $\gamma \leq d$ different weights and $|w_{[d]}| \leq k$. Then, the index set $[d]$ can be partitioned as $S_1 \cup \ldots \cup S_\eta$ with $\eta \leq 6\gamma$ such that for all $i \in [\eta]$, the sub-word $w_{|S_i}$ has at most two distinct weights and $|w_{S_i}| \leq k$.*

Proof. Let $\{T_1, \ldots, T_\gamma\}$ be a partition of $[d]$ where every set T_j in the partition corresponds to one weight (i.e. for every $i \in T_j$, $w_i = \alpha_j$). We give an algorithm to obtain the desired partition of $[d]$.

1. Initialize $j = 1$. Initialize $\pi := \{T_1, \ldots, T_\gamma\}$. Repeat the following steps until π is empty.
2. If possible, pick some set T_p and some set T_n from π such that α_p is positive and α_n is negative .
3. If $|T_p|\alpha_p + |T_n|\alpha_n \leq 0$, then it's easy to see that we can pick a subset $T'_n \subseteq T_n$ such that $\Big||T_p|\alpha_p + |T_n|\alpha_n\Big| \leq k$ as $|\alpha_p|, |\alpha_n| \leq k$.
4. Set $S_j := T_p \cup T'_n$. We have $|w_{S_j}| = \Big||T_p|\alpha_p + |T_n|\alpha_n\Big| \leq k$ as required. Set $T_n := T_n \setminus T'_n$. Drop T_p from π. If $|T_p|\alpha_p + |T_n|\alpha_n \geq 0$, we proceed analogously.
5. If we can't pick two sets T_p and T_n as above, it means that for the remaining sets in π, either their corresponding weights are all positive or all negative. We consider the case when they are all positive (the other case can be dealt with analogously).
   a. If there exists a set T_p such that $|T_p|\alpha_p \leq k$, then set $S_j := T_p$ and drop T_p from π.
   b. Otherwise consider any remaining set T_p. We have $|T_p|\alpha_p > k$. Since $\alpha_p \leq k$, there exist $T'_p \subseteq T'_p \cup \{q\} \subseteq T_p$ such that $|T'_p|\alpha_p \leq k$ and $(|T'_p| + 1)\alpha_p > k$. Set $S_j := T'_p$, $S_{j+1} = \{q\}$ and $T_p := T_p \setminus (T'_p \cup \{q\})$. Increment $j = j + 1$.
6. Increment $j = j + 1$ and continue.

We have ensured that $|w_{S_i}| \leq k$ for all i. It suffices to show that the steps 2-6 are repeated at most 3γ times. Every time step 4 or step $5.a$ is executed, the size of π reduces by at least 1. Hence they can be repeated at most γ times in total. When step $5.b$ is executed for the first time, we know that the remaining collection of sets is $\pi = \{T_1, \ldots, T_\beta\}$ where each T_j corresponds to a positive weight. Let us denote the weight of this collection by $w_\pi = \sum_{j=1}^{\beta} w_{T_j} = \sum_{j=1}^{\beta} |T_j|\alpha_j$. Suppose till now we have picked the sets $S_1, \ldots, S_{\beta'}$ for some $\beta' \leq \gamma$. Then $w_\pi = w_S - \sum_{i=1}^{\beta'} w_{S_i}$. Using triangle inequality, $w_\pi \leq |w_S| + \sum_{i=1}^{\beta'} |w_{S_i}| \leq k + \gamma k$. Every time we remove two sets $S_j = T'_p$ and $S_{j+1} = \{q\}$ as in step $5.b$, the value of w_π reduces by $(|T'_p| + 1)\alpha_p > k$. Hence this can be repeated at most $\gamma + 1$ times. ◀

▷ Claim 15. Let $S \subseteq [d]$ and let $w \in \{\alpha_1, \ldots, \alpha_\gamma\}^d$ ($|\alpha_i| \leq k$ for all i) be a word with $\gamma \leq d$ different weights and $|w_S| \leq k$. Then, there exist $(J_+, J_-, \pi), (J'_+, J'_-, \pi')$ such that (S, J_+, J_-, π) and (S, J'_+, J'_-, π') are valid. For any fixed integer Δ and for all $\tau \in \{0,1\}^{|k_+ - k_-|}$, the polynomial $P_{(S,J_+,J_-,\pi,\tau)})$ can be computed by a set-multilinear formula of product-depth Δ and size at most $|S|^\Delta 2^{30k\gamma\Delta d^{\mu(\Delta)}}$ while the polynomial $P_{(S,J'_+,J'_-,\pi',\tau)})$ can be computed by a set-multilinear formula of product-depth Δ and size at most $|S|^\Delta 2^{5k\Delta d^{\mu(\Delta-1)}+6\gamma k}$.

Proof. As $|w_{[d]}| \leq k$, by Lemma 14, we get a partition of the index set $[d]$ into sets $S_1, \ldots, S_\eta$ ($\eta \leq 6\gamma$) such that the sub-word corresponding to each S_i contains at most two weights and $|w_{S_i}| \leq k$.

- **Constructing** P_Δ: We apply Claim 13 to each S_i to get a partition $S_i = S_{i,1}, \ldots, S_{i,r_i}$ where $\sum_{j \in r_i} |w_{S_{i,j}}| \leq 5kd^{\mu(\Delta)}$. We club all the $\mathcal{P}$-heavy sets together and all the $\mathcal{N}$-heavy sets together across all S_is. We obtain depth $\Delta - 1$ formulas for each $S_{i,j}$ with size at most

$$|S_{ij}|^{\Delta-1} 2^{5k(\Delta-1)d^{\mu(\Delta)}}$$

Using the exact same construction as in the proof of Claim 12, we obtain the polynomial $P_\Delta \coloneqq P_{([d],J_+,J_-,\pi,\tau)}$ of product-depth Δ and size at most

$$\sum_i r_i \cdot 2^{k''_- - k''_+} \cdot \max_{j \in \sum_i r_i} s_j \leq d \cdot 2^{\sum_{i\in[\eta],j\in[r_i]} |w_{S_{i,j}}|} \cdot d^{\Delta-1} 2^{5k(\Delta-1)d^{\mu(\Delta)}}$$
$$\leq d^\Delta 2^{30k\gamma\Delta d^{\mu(\Delta)}}$$

- **Constructing Q_Δ:** We can now apply Lemma 11 to each of these S_is where we set the product depth to $\Delta - 1$. For all $i \in [\eta]$, we obtain polynomials $P_{(S_i,J_{i,+},J_{i,-},\pi_i,\tau_i)}$ with formulas of size
$$|S_i|^{\Delta-1} 2^{5k(\Delta-1)d^{\mu(\Delta-1)}}$$
and product depth $\Delta - 1$.
Using the exact same construction as in the proof of Claim 12, we obtain the polynomial $Q_\Delta \coloneqq P'_{([d],J_+,J_-,\pi,\tau)}$ of product-depth Δ and size at most
$$\eta \cdot 2^{k''_- - k''_+} \cdot \max_{i\in[\eta]} s_i \leq d \cdot 2^{\sum_{i\in[\eta]} |w_{S_i}|} \cdot d^{\Delta-1} 2^{5k(\Delta-1)d^{\mu(\Delta-1)}}$$
$$\leq d^\Delta 2^{5k(\Delta-1)d^{\mu(\Delta-1)}+6\gamma k}$$

◀