A Polynomial Restriction Lemma with Applications

Valentine Kabanets* Daniel M. Kane† Zhenjian Lu‡

February 17, 2017

Abstract

A polynomial threshold function (PTF) of degree d is a boolean function of the form $f = \mathrm{sgn}(p)$, where p is a degree-d polynomial, and sgn is the sign function. The main result of the paper is an almost optimal bound on the probability that a random restriction of a PTF is not close to a constant function, where a boolean function g is called δ-close to constant if, for some $v \in \{1, -1\}$, we have $g(x) = v$ for all but at most δ fraction of inputs. We show for every PTF f of degree $d \geq 1$, and parameters $0 < \delta, r \leq 1/16$, that

$$\mathbf{Pr}_{\rho \sim R_r}[f_\rho \text{ is not } \delta\text{-close to constant}] \leq (\sqrt{r} + \delta) \cdot (\log r^{-1} \cdot \log \delta^{-1})^{O(d^2)},$$

where $\rho \sim R_r$ is a random restriction leaving each variable, independently, free with probability r, and otherwise assigning it 1 or -1 uniformly at random. In fact, we show a more general result for random *block* restrictions: given an arbitrary partitioning of input variables into m blocks, a random block restriction picks a uniformly random block $\ell \in [m]$ and assigns 1 or -1, uniformly at random, to all variable outside the chosen block ℓ. We prove the Block Restriction Lemma saying that a PTF f of degree d becomes δ-close to constant when hit with a random block restriction, except with probability at most $(m^{-1/2} + \delta) \cdot (\log m \cdot \log \delta^{-1})^{O(d^2)}$.

As an application of our Restriction Lemma, we prove lower bounds against constant-depth circuits with PTF gates of any degree $1 \leq d \ll \sqrt{\log n / \log \log n}$, generalizing the recent bounds against constant-depth circuits with linear threshold gates (LTF gates) proved by Kane and Williams (*STOC*, 2016) and Chen, Santhanam, and Srinivasan (*CCC*, 2016). In particular, we show that there is an n-variate boolean function $F_n \in \mathrm{P}$ such that every depth-2 circuit with PTF gates of degree $d \geq 1$ that computes F_n must have at least $\left(n^{\frac{3}{2}+\frac{1}{d}}\right) \cdot (\log n)^{-O(d^2)}$ wires. For constant depths greater than 2, we also show average-case lower bounds for such circuits with super-linear number of wires. These are the first super-linear bounds on the number of wires for circuits with PTF gates. We also give short proofs of the optimal-exponent average sensitivity bound for degree-d PTFs due to Kane (*Computational Complexity*, 2014), and the Littlewood-Offord type anticoncentration bound for degree-d multilinear polynomials due to Meka, Nguyen, and Vu (*Theory of Computing*, 2016).

Finally, we give *derandomized* versions of our Block Restriction Lemma and Littlewood-Offord type anticoncentration bounds, using a pseudorandom generator for PTFs due to Meka and Zuckerman (*SICOMP*, 2013).

*School of Computing Science, Simon Fraser University, Burnaby, BC, Canada; kabanets@sfu.ca

†Department of Mathematics, University of California, San Diego, La Jolla, CA, USA; dakane@ucsd.edu

‡School of Computing Science, Simon Fraser University, Burnaby, BC, Canada; zla54@sfu.ca

Contents

1 Introduction

Random restrictions of boolean functions play an important role in the circuit complexity research. One way to prove that a certain boolean function h is not computable by a class $\mathcal{C}$ of boolean circuits is to show that (1) every boolean function $f \in \mathcal{C}$ becomes "simplified" after a random restriction (which randomly fixes some random subset of variables of f), and (2) the function h "remains hard" after a random restriction. This strategy has been successfully applied to prove circuit lower bounds for explicit boolean functions against (i) de Morgan formulas [Sub61, And87], (ii) AC^0 circuits [Ajt83, FSS84, Yao85, Hås89], and (iii) constant-depth circuits with LTF (linear threshold function) gates [IPS97, CSS16, KW16]. In most of these results, the notion of "simplified" means that a restricted function is (almost) a constant function; the hard function h is the parity function, which is the ultimate example of a function that cannot be made (close to) constant under any restriction that leaves enough variables unrestricted.

Lower bounds against constant-depth circuits with PTF gates. One of the original motivations for the present work was to extend the lower bounds of [CSS16, KW16] to the class of constant-depth circuits consisting of general Polynomial Threshold Function (PTF) gates. Recall that a degree-d PTF is defined to be the sign of a multilinear degree-d polynomial over the reals. Constant-depth PTF circuits are quite powerful. Every n-variate boolean function computable by a polynomial-size $\mathsf{AC}^0[m]$ circuit (constant-depth circuits with AND, OR, NOT, and mod m gates), for any integer $m > 1$, has an equivalent depth-2 circuit of quasipolynomial size with d-degree PTF gates, for $d \leq \mathsf{poly}(\log n)$ [All89, Yao90]. Moreover, every boolean function computable by a polynomial-size AC^0 circuit can be well approximated by a single PTF gate of degree $d \leq \mathsf{poly}(\log n)$ [LMN93].

We prove circuit lower bounds against constant-depth PTF circuits of super-constant degree d as long as $d \ll \sqrt{\log n / \log \log n}$. This is close to the best possible given the current knowledge, as virtually nothing is known for PTFs of degree bigger than $\log n$. We state our circuit lower bounds below in Section 1.2. Our main technical tool is a restriction lemma for PTFs, which we discuss next.

Restriction Lemmas. Let $f(x_1, \ldots, x_n)$ be a boolean function assuming the values $\{1, -1\}$ over the boolean cube $\{-1, 1\}^n$. For a parameter $0 < r < 1$, a random r-restriction is defined to leave each variable free (unrestricted) with probability r, and, otherwise (with probability $1 - r$), fixing the variable to either 1 or -1 uniformly at random. For a parameter $0 < \delta < 1$, we say that a boolean function g is δ-close to constant if, for some $v \in \{-1, 1\}$, we have $g(x) = v$ for all but at most δ fraction of boolean inputs x. We show that a PTF f of degree d is likely to become δ-close to constant after being hit with a random r-restriction: for $\delta \leq r$, the probability that f fails to become δ-close to constant is at most $\sqrt{r} \cdot (\log r^{-1} \cdot \log \delta^{-1})^{O(d^2)}$.

This Restriction Lemma for PTFs is sufficient to derive the aforementioned lower bounds against constant-depth PTF circuits. Moreover, it can also be used to re-derive (as an immediate corollary) the optimal-exponent average sensitivity bound for degree-d PTFs due to Kane [Kan14]. But perhaps more interestingly, this Restriction Lemma for PTFs is a consequence of a more general Restriction Lemma for degree-d polynomials, which has other applications.

First, we generalize our PTF Restriction Lemma to the case of "structured" random restrictions. Suppose that the variables of a given boolean function $f(x_1, \ldots, x_n)$ are partitioned into m disjoint

subsets (blocks) of variables. Given such a block partition, a random block restriction is defined as follows: pick a uniformly random block $\ell \in [m]$, and assign each variable outside the chosen block ℓ a uniformly random value in $\{-1, 1\}$. We show that a degree-d PTF f with an arbitrary block partition into m blocks is likely to become δ-close to constant after being hit with a random block restriction: for $\delta \leq 1/m$, the probability that f fails to become δ-close to constant is at most $m^{-1/2} \cdot (\log m \cdot \log \delta^{-1})^{O(d^2)}$. It is not hard to see that a PTF Restriction Lemma for r-random restrictions is a corollary of the m-block Restriction Lemma when $m = 1/r$.

The PTF Block Restriction Lemma mentioned above is a consequence of the following Block Restriction Lemma for polynomials. If a multilinear degree-d polynomial with a given block partition into m blocks is hit with a random block restriction, it is likely to become "concentrated" in the sense that its standard deviation becomes quite small relative to its expectation. It can be shown that if a polynomial is concentrated, then the sign function of this polynomial (i.e., the corresponding PTF) is close to a constant function. Thus, the PTF Block Restriction Lemma follows. In addition, this structural property of a polynomial becoming "concentrated" under random block restrictions is also useful for proving Littlewood-Offord type anticoncentration results for polynomials.

Littlewood-Offord type anticoncentration bounds. Let p be an arbitrary n-variate degree-d multilinear polynomial containing at least t disjoint maximal monomials (i.e., not contained in other monomials), each with a coefficient at least 1 in magnitude. Meka et al. [MNV16] showed that it is unlikely that, for a random $x \in \{-1, 1\}^n$, the value $p(x)$ will fall in the interval $[0, 1]$: the probability that $p(x) \in [0, 1]$ is at most $(1/\sqrt{t}) \cdot (\log t)^{O(d \log d)} \cdot \exp(d^2 \log d)$. We re-derive this result, in a simple way, from our Block Restriction Lemma for polynomials. Moreover, we also prove a *derandomized* version of this bound, which we discuss next.

Derandomized Block Restriction Lemma and derandomized Littlewood-Offord. A random block restriction chooses a uniformly random block, and then assigns uniformly random values to all variables outside that block. Our Block Restriction Lemma says that such a random block restriction is likely to make an n-variate degree-d multilinear polynomial "concentrated". A derandomized version of this lemma would say that a similar conclusion is true for block restrictions that can be sampled with significantly fewer random bits. We prove such a derandomized version for pseudorandom m-block restrictions that are sampled using about $(\log m)^{O(d^2)} \cdot \log n$ random bits.

We then use this derandomized version of the Block Restriction Lemma to obtain derandomizations of the Littlewood-Offord type bounds. We show that there is an efficient pseudorandom generator for sampling inputs $x \in \{-1, 1\}^n$, using significantly fewer than n random bits, such that the following holds. For any degree-d multilinear polynomial p with many degree-d monomials that have large coefficients, it is unlikely that $p(x) \in [0, 1]$ for these pseudorandom inputs $x \in \{-1, 1\}^n$. No derandomized versions of the Littlewood-Offord type anticoncentration bounds were previously known.

Next we provide more details about our main results and our proof techniques.

1.1 Circuit complexity

One of the main goals of complexity theory is to understand the computational power of efficient nonuniform algorithms (circuits). As the general class of polynomial-size boolean circuits (nonuni-

form P, or $\mathsf{P/poly}$) seems well beyond the currently known methods for proving lower bounds, the focus of circuit complexity research has been on various restricted circuit classes. Particularly successful has been the study of constant-depth circuits (which can be thought of as very efficient parallel nonuniform algorithms).

Different natural sets of gates (elementary logical operations) were considered. For the gates AND, OR, and NOT (where AND and OR have unbounded fanin), the resulting circuit class is AC^0. A milestone in circuit complexity was the proof that the parity function on n bits requires exponential-size AC^0 circuits (with a matching upper bound also known) [Ajt83, FSS84, Yao85, Hås89]. Adding the parity gate to AC^0 circuits, we get the class $\mathsf{AC}^0[2]$ with modulo 2 gates. An exponential lower bound against $\mathsf{AC}^0[2]$ for the n-bit majority function was shown by Razborov [Raz87], and was extended by Smolensky [Smo87] to exponential lower bounds against $\mathsf{AC}^0[p]$, for an arbitrary prime modulus $p > 0$. It is still open (though widely believed) whether the majority function requires exponential size also for the class $\mathsf{AC}^0[m]$ for any composite modulus $m > 1$; significant progress has been recently made by Williams [Wil14] who showed that a boolean function computable in nondeterministic exponential time (NEXP) requires superpolynomial-size $\mathsf{AC}^0[m]$ circuits, for any integer modulus $m > 1$.

Adding the majority gate to AC^0 circuits, we get the class TC^0, for which no superpolynomial circuit lower bounds are known for any explicit function (not even for a function in NEXP), despite serious efforts by complexity researchers over the past thirty years. One reason for our inability to prove strong lower bounds against TC^0 stems from the fact that TC^0 is a powerful circuit class, capable of computing many interesting and useful functions: addition, multiplication, division, and sorting (see [Raz92] and the references therein). Surprisingly, every function computable by a polynomial-size $\mathsf{AC}^0[m]$ circuit, for any integer $m > 1$, has an equivalent depth-3 TC^0 circuit of quasipolynomial size [All89, Yao90]. Moreover, TC^0 is conjectured to be capable of computing cryptographically secure pseudorandom function generators [NR04], which, coupled with arguments in [RR97], means that it is highly unlikely that a "usual" (termed "natural" by Razborov and Rudich [RR97]) lower bound proof method would work against TC^0, as any such "natural" proof would yield an efficient algorithm to break every candidate pseudorandom function generator in TC^0.

As it seems very difficult to prove superpolynomial lower bounds against TC^0, the focus has shifted to proving fixed-polynomial lower bounds (even for a fixed constant depth, say depth 2 or 3). Before discussing these results, let us mention another motivation for studying TC^0. Closely related to the majority function is a Linear Threshold Function (LTF), defined as the sign of a linear (degree 1) polynomial in variables $x_1, \ldots, x_n$; when the variables x_i assume boolean values, the resulting LTF is a boolean function. Note that an LTF may have arbitrarily large coefficients (weights) for the underlying linear polynomial, which makes an LTF provably more powerful than a majority function, or more generally, than an LTF with small (polynomially bounded) weights [MK61]. On the other hand, somewhat surprisingly, an arbitrary LTF can be represented by a polynomial-size depth-2 TC^0 circuit [GHR92].

Linear Threshold Functions (LTFs) and circuits with LTF gates have been studied since at least the 1940s in the context of artificial neural networks [MP43] (see [Ant01] and the references therein). Some of the early lower bounds for LTFs are due to Minsky and Papert [MP69], who showed, for example, that the parity function cannot be computed by any LTF. Minsky and Papert [MP69] also considered Polynomial Threshold Functions (PTFs), defined as the sign of an arbitrary degree-d polynomial in $x_1, \ldots, x_n$, and showed that the n-bit parity function requires a PTF of degree n.

Apart from minimizing the degree of a PTF $f = \mathrm{sgn}(p)$, it is also natural to minimize its sparsity, defined as the number of monomials in the polynomial p. This particularly makes sense if we view a PTF as a depth-2 circuit computing an LTF of parities of subsets of input variables (where the parities correspond to monomials of p when the boolean variables are assumed to be from $\{-1, 1\}$). Superpolynomial lower bounds on the PTF sparsity for simple boolean functions were shown by Bruck [Bru90] (see also [BS92] for the complementary upper bounds on the sparsity of PTFs).

For constant-depth LTF circuits, two complexity measures have been considered: the number of gates (excluding the input variables), and the number of wires. The first super-linear wire complexity bound was obtained by Impagliazzo et al. [IPS97], who showed that the n-bit parity function requires depth-D LTF circuits with at least $n^{1+\epsilon_D}$ wires, where $\epsilon_D = \exp(-D)$. They also showed that the n-bit parity function requires depth-D LTF circuits with at least $(n/2)^{1/(2(D-1))}$ gates. Recently, these bounds were generalized to average-case (correlation) bounds by Chen et al. [CSS16]. For depth-2 LTF circuits, Kane and Williams [KW16] have recently proved an $n^{3/2}/\mathsf{poly}(\log n)$ gate complexity bound, and $n^{5/2}/\mathsf{poly}(\log n)$ wire complexity bound for an explicit function in P (Andreev's function [And87]). In a more recent result, Alman et al. [ACW16] showed a fast satisfiability algorithm for $\mathsf{ACC}^0 \circ \mathrm{LTF} \circ \mathrm{LTF}$ circuits with sub-quadratic number of LTF gates on the bottom layer and sub-exponential number of gates on the other layers, and hence obtained lower bounds against these circuits for an explicit function in E^{NP}, using the connection between satisfiability algorithms and circuit lower bounds due to Williams [Wil13, Wil14]. Also, Tamaki [Tam16] has recently showed a fast satisfiability algorithm and lower bounds for depth-2 LTF circuits with sub-quadratic number of gates.

Circuits with PTF gates of degree $d > 1$ were previously studied by Nisan [Nis94]. Using multi-party communication complexity lower bounds, Nisan proved exponential correlation bounds against circuit with $\Omega_d(n^{1-o(1)})$ PTF gates of degree d. We are not aware of any prior work on circuits with PTF gates of degree $d > 1$ that would prove either super-linear wire complexity lower bounds or super-linear gate complexity lower bounds.

1.2 Our contributions

Lower bounds for constant-depth PTF circuits. We generalize the lower bounds of [KW16] and [CSS16] to the case of constant-depth circuits with PTF gates of degree $d \geq 1$. For the case of $d = 1$, our results match those obtained in [KW16, CSS16]. For $d > 1$, these appear to be the first super-linear wire complexity lower bounds against constant-depth circuits with degree-d PTF gates.

The following generalizes the lower bounds against LTF circuits of depth 2 of [KW16].

Theorem 1.1. *There is an n-variate boolean function $F_n \in \mathsf{P}$ such that every depth-2 circuit with PTF gates of degree $d \geq 1$ that computes F_n must have at least $\left(n^{\frac{1}{2}+\frac{1}{d}}\right) \cdot (\log n)^{-O(d^2)}$ gates, and at least $\left(n^{\frac{3}{2}+\frac{1}{d}}\right) \cdot (\log n)^{-O(d^2)}$ wires.*

We also generalize to PTF gates (and somewhat strengthen) a lower bound of [KW16] for depth-3 circuits that have the Majority gate at the top, with depth-2 LTF circuits feeding in.

Theorem 1.2. *There is a polynomial-time computable boolean function B such that the following holds. For any $\frac{1}{\log n} \ll \epsilon < 1$, let C be a majority vote of depth-2 circuits with degree-d PTF gates*

such that the top majority gate has fanin at most 2^{n^ϵ} *and the total fanin of the gates on the bottom layer at most* $w = \left(n^{\frac{3}{2}+\frac{1}{d}}\right) \cdot (n^\epsilon \cdot \log n)^{-c \cdot d^2}$*, where* c *is a constant. Then* C *cannot compute* B*.*

For boolean functions $f, g\colon \{-1,1\}^n \to \{-1,1\}$, define the correlation between f and g as

$$\mathbf{Corr}(f,g) = \left|\mathbf{Exp}_{x\sim\{-1,1\}^n}\left[f(x)\cdot g(x)\right]\right|.$$

Let Par_n denote the n-input parity function. We generalize the correlation bounds of [CSS16], getting the following.

Theorem 1.3. *For any* $D \geq 1$ *and* $1 \leq d \ll \sqrt{\log n/\log\log n}$*, let* C *be any depth-*D *circuit on* n *inputs with degree-*d *PTF gates, of wire complexity at most* $n^{1+\varepsilon_D}$*, where* $\varepsilon_D = B^{-(2D-1)}$*, for some constant* $B > 0$*. Then we have*

$$\mathbf{Corr}(C, \mathrm{Par}_n) \leq O\left(n^{-\varepsilon_D}\right).$$

Theorem 1.4. *There is an* n*-variate boolean function* $G_n \in \mathsf{P}$ *such that the following holds. For any* $D \geq 1$ *and* $1 \leq d \ll (\log n/\log\log n)^{1/(2D-1)}$*, let* C *be any depth-*D *circuit on* n *inputs with degree-*d *PTF gates, of wire complexity at most* $n^{1+\mu_{D,d}}$*, where* $\mu_{D,d} = (E\cdot d)^{-(2D-1)}$*, for some constant* $E > 0$*. Then we have*

$$\mathbf{Corr}(C, G_n) \leq \exp(-n^{\mu_{D,d}/2}).$$

Restriction lemmas for polynomials. Our main tool is the following structural lemma showing that a PTF is likely to become an almost constant function after being hit with a random restriction. Below, we denote by $\rho \sim R_r$ the process of picking a random restriction ρ that leaves a variable free with probability r, and otherwise fixes uniformly at random to 1 or -1. We denote by f_ρ the function f restricted by ρ. We say that a boolean function f is δ-close to constant if, for some value $v \in \{-1,1\}$, we have $f(x) = v$ for all but at most δ fraction of boolean inputs x.

Lemma 1.5 (PTF Restriction Lemma)**.** *For any PTF* $f(x) = \mathrm{sgn}(p(x))$ *of degree* $d \geq 1$*, and any* $0 < \delta$*,* $r \leq 1/16$*, we have*

$$\mathbf{Pr}_{\rho\sim R_r}\left[f_\rho \text{ is not } \delta\text{-close to constant}\right] \leq (\sqrt{r}+\delta)\cdot(\log r^{-1}\cdot\log\delta^{-1})^{O(d^2)}.$$

We note that the bound $r^{1/2}$ in this lemma has an optimal exponent.

The above lemma is a consequence of a more general result, the *Block Restriction Lemma*, which deals with certain structured restrictions that we define next. Suppose the variables of a given function are arbitrarily partitioned into m blocks. For the given block partitioning, a random *block restriction* $\rho \sim B_m$ is defined by picking a block $\ell \in [m]$ uniformly at random, and assigning each variable outside block ℓ the value 1 or -1 uniformly at random. We show that, for an arbitrary partitioning of input variables into m blocks, the probability that a degree-d PTF is not δ-close to constant, after being hit with a random block restriction $\rho \sim B_m$, is at most the same as the bound in the PTF Restriction Lemma above, with $r = 1/m$.

Lemma 1.6 (Block Restriction Lemma: Simplified version)**.** *For any PTF* $f(x) = \mathrm{sgn}(p(x))$ *of degree* $d \geq 1$*, any* $m \geq 16$ *, and any* $0 < \delta \leq 1/16$*, we have*

$$\mathbf{Pr}_{\rho\sim B_m}\left[f_\rho \text{ is not } \delta\text{-close to constant}\right] \leq (m^{-1/2}+\delta)\cdot(\log m\cdot\log\delta^{-1})^{O(d^2)}.$$

Note that a standard random restriction $\rho \sim R_r$ can be obtained by first randomly partitioning the input variables into $m = 1/r$ blocks, and then applying a random block restriction from B_m. So the Block Restriction Lemma implies the PTF Restriction Lemma.

Our actual Block Restriction Lemma (Lemma 5.1) shows something even stronger. If a degree d multilinear polynomial p is hit with a random block restriction, it becomes "concentrated around the expectation" in the sense that its standard deviation becomes quite small relative to its expectation (in particular, implying that the restriction of the PTF $\mathrm{sgn}(p)$ is close to constant).

Other applications. Apart from the aforementioned circuit lower bound applications, our Block Restriction Lemma also immediately implies two other results. We get the average sensitivity bound on degree-d PTFs, with an optimal exponent, first shown by Kane [Kan14] in the context of the Gotsman-Linial conjecture [GL94]; see Theorem 6.13 below.

We also get the following Littlewood-Offord type anticoncentration bound for degree-d multilinear polynomials, due to Meka et al. [MNV16], which is an extension of the classical Littlewood-Offord result for linear polynomials [LO43, Erd45].

Theorem 1.7 ([MNV16])**.** *For any real interval I, and any n-variate degree-d multilinear polynomial p such that there exists a set of t disjoint monomials in p, each of which is maximal (i.e., not contained by any other monomials) and has coefficient at least $|I|$ in magnitude, we have*

$$\mathbf{Pr}[p(A) \in I] \leq t^{-1/2} \cdot (\log t)^{O(d \log d)} \cdot 2^{O\left(d^2 \log d\right)},$$

where A is the uniform distribution over $\{-1,1\}^n$.

Derandomization. We prove a derandomized version of the Block Restriction Lemma mentioned above.

Theorem 1.8 (Derandomized Block Restriction Lemma: Simplified version)**.** *For any $0 < \delta \leq 1/16$ and $0 < \zeta < 1$, there is a polynomial-time algorithm for sampling block restrictions $\rho \in B_m$, for any $m \geq 16$, that uses at most $m^{\zeta} \cdot \log n$ random bits, so that the following holds. For any n-variate degree-d PTF f whose variables are partitioned into m blocks, we have*

$$\mathbf{Pr}_{\rho}\left[p_{\rho} \text{ is not } \delta\text{-concentrated}\right] \leq \left(m^{-1/2} + \delta\right) \cdot \left(\log m \cdot \log \delta^{-1}\right)^{O\left(\zeta^{-1} \cdot d^2\right)}.$$

Our actual version of this lemma (see Theorem 7.1) shows that a degree-d polynomial p is likely to become "concentrated" under a pseudorandom block restriction. This in turn is used to prove the following derandomized versions of Theorem 1.7.

Theorem 1.9. *For any positive integers t and d, and $0 < \zeta < 1$, there exists a distribution D on $\{-1,1\}^n$, samplable in $\mathsf{poly}(n)$ time using $t^{\zeta/d} \cdot \log n$ random bits, such that the following holds. For any real interval I, and any n-variate degree-d multilinear polynomial p that has at least t disjoint degree-d monomials with coefficient at least $|I|$ in magnitude, we have*

$$\mathbf{Pr}\left[p(D) \in I\right] \leq t^{-\frac{1}{2d}} \cdot (\log t)^{O\left(\zeta^{-1} \cdot d^2\right)}.$$

Theorem 1.10. *For any positive integers t and d, and $0 < \zeta < 1$, there exists a distribution D on $\{-1,1\}^n$, samplable in $\mathsf{poly}(n)$ time using $t^{\zeta} \cdot \log n$ random bits, such that the following holds.*

For any real interval I, and any n-variate degree-d multilinear polynomial p with at least $t \cdot n^{d-1}$ degree-d monomials whose coefficients are at least $|I|$ in magnitude, we have

$$\mathbf{Pr}\left[p(D) \in I\right] \leq t^{-\frac{1}{2}} \cdot (\log t)^{O\left(\zeta^{-1} \cdot d^2\right)}.$$

Note that it is possible to use PRGs for PTFs directly to get a derandomized Littlewood-Offord anticoncentration bound. However, using the best currently known PRGs for PTFs, such a derandomization will have large error and seed length. In particular, for dense polynomials with $t = n^{1-o(1)}$, Theorem 1.10 achieves error less than $1/n^{0.49}$ with the seed size at most polylogarithmic in n; such parameters are beyond reach of the best available PRGs for PTFs.

1.3 Related work

Random restrictions. The concept of random restrictions for boolean functions was introduced by Subbotovskaya [Sub61], who applied it to show that the n-bit parity function requires de Morgan formulas[1] of size $\Omega(n^{1.5})$ (later improved to the optimal bound $\Omega(n^2)$ by [Khr71]). Andreev [And87] combined random restrictions with a counting argument to show a stronger lower bound against de Morgan formulas for a function in P (resulting in the $n^3/\mathsf{poly}(\log n)$ bound, when using Håstad's improved restriction lemma for de Morgan formulas [Hås98]). Random restrictions were also used for showing the aforementioned exponential lower bounds against AC^0 circuits computing the parity function [Ajt83, FSS84, Yao85, Hås89], as well as for the lower bounds against constant-depth LTF circuits by [IPS97, CSS16, KW16]. A common feature in all of these lower bound proofs is a structural result showing how "easy" boolean functions (of appropriately small formula or circuit complexity) become much "simpler" (e.g., become almost constant) after being hit with random restrictions. In contrast, the parity function is the ultimate "restriction-resistant" function that does not simplify under random restrictions, but rather stays the parity function (albeit on a smaller number of variables).

Two classical examples of restriction lemmas are the Shrinkage Lemma for de Morgan formulas [Sub61, IN93, PZ93, Hås98, Tal14], and Håstad's Switching Lemma for AC^0 circuits [Hås89] (see also [RST15, Hås16]). The Shrinkage Lemma says that a de Morgan formula of size s is expected to shrink to size about $r^2 \cdot s$, after being hit with a random restriction $\rho \sim R_r$ that leaves each variable free with probability r, and otherwise fixes the variable to a uniform bit. Håstad's Switching Lemma says that any given k-cnf formula (the conjunction of clauses of size at most k each) is very likely to become expressible as a k-dnf (the disjunction of size-k terms), after being hit with a random restriction $\rho \sim R_r$ for $r = O(1/k)$. By repeatedly applying this Switching Lemma to a given AC^0 circuit (of not too large size), level by level, we can merge the adjacent levels, thereby collapsing the original circuit to depth at most 2. Once the original AC^0 circuit is thus "simplified", one can argue directly that the new circuit is too weak to compute the restriction of the original function (e.g., the parity function).

A similar strategy was used by Impagliazzo et al. [IPS97] to show that the parity function is hard for constant-depth LTF circuits. The main technical result of [IPS97] shows that a depth d LTF circuit (of not too large size) can be reduced to a depth $d-1$ LTF circuit by fixing not too many input variables. This is argued by showing that there exists a particular restriction of input variables, chosen adaptively, that will make all LTF gates at the bottom level of the circuit to be constants (or depend on at most one input). To extend the worst-case lower bounds of [IPS97] to

[1] De Morgan formulas are built using AND, OR, and NOT gates.

the average-case correlation bounds, Chen et al. [CSS16] extended the restriction lemma of [IPS97] to the setting of truly random, non-adaptive restrictions. So too did Kane and Williams [KW16] to get a lower bound against depth-2 LTF circuits for Andreev's function; their restriction lemma is for certain "block-structured" random restrictions, as required by Andreev's original argument.

Our restriction lemma, Lemma 1.5, can be used to re-derive the same lower bounds (up to polylogarithmic factors) for LTF circuits as in [CSS16, KW16]. Moreover, it extends these lower bounds to the case of PTF gates of any degree $d \ll \sqrt{\log n / \log \log n}$.

Comparison with [KW16], [CSS16] and [Nis94]. Kane and Williams [KW16] prove an LTF Restriction Lemma with similar parameters to our PTF Restriction Lemma (for $d = 1$), for certain random block restrictions and for the case of the restricted LTF becoming a constant function (rather than close to constant). For the proof, they rely on the Littlewood-Offord lemma from additive combinatorics [LO43, Erd45]. It is not clear how to extend such a proof to the case of higher degree PTFs.

Chen et al. [CSS16] obtain a quantitatively weaker version of the LTF Restriction Lemma, using proof techniques similar to ours, but with worse parameters. We get better (almost optimal) parameters for both LTFs and higher degree PTFs, by using the more refined proof techniques developed in [Kan14].

Nisan [Nis94] obtains an almost linear $\Omega_d(n^{1-o(1)})$ gate complexity lower bound against circuits with degree-d PTF gates of any depth. The technique used by Nisan is based on communication complexity and is quite different from the technique in [KW16], [CSS16] and this work. It is not clear how such techniques can be used to obtain super-linear lower bounds in the setting of either wire complexity or gate complexity. For degree $d = 1$, our lower bound for depth-2 PTF circuits matches the super-linear $n^{1.5-o(1)}$ gate complexity lower bounds against depth-2 LTF circuits first shown in [KW16]. For higher degrees, this result cannot give super-linear lower bounds and does not match Nisan's lower bounds. However, both our results for depth-2 and higher constant depth PTF circuits give super-linear wire complexity lower bounds, which is not implied by [Nis94] or prior work.

Lower bounds against TC^0. For depth-2 circuits with majority gates (equivalently, LTF gates with polynomially small weights), Hajnal et al. [HMP+93] showed an exponential size lower bound for the Inner Product modulo 2 (IP2) function. For the parity function, Paturi and Saks [PS94] showed a nearly optimal $\tilde{\Omega}(n)$ gate complexity lower bound against depth-2 majority circuits. This was extended by Siu et al. [SRK94] to depth-D such circuits, showing that n-bit Parity requires at least $\tilde{\Omega}(D \cdot n^{1/(D-1)})$ gates; they also showed a matching upper bound of $O(D \cdot n^{1/(D-1)})$ gates.

Goldmann et al. [GHR92] (improving upon [SB91]) proved a surprising result that any general LTF circuit of constant depth D has an equivalent majority circuit of polynomially related size and depth $D + 1$. Thus any superpolynomial lower bound against majority circuits of constant depth D would immediately yield a superpolynomial lower bounds against general LTF circuits of depth $D - 1$. (This connection may explain the lack of any strong lower bounds even for depth-3 majority circuits.) Allender and Koucký [AK10] show that proving superpolynomial circuit lower bounds against TC^0 circuits (for an NC^1-complete function) is equivalent to proving super-linear, $n^{1+\epsilon}$, lower bounds for every depth $D \geq 2$, where $\epsilon > 0$ is independent of the depth D.

PTFs. PTFs have also been studied in the context of learning [STT12, DOSW11, DSTW14], pseudorandomness [DGJ+10, DKN10, Kan11, Kan12, MZ13, Kan14, Kan15], approximate counting [DDS14, DS14], and extremal combinatorics [Sak93, GL94, OS08, DRST14, Kan14].

1.4 Our proof techniques

Block Restriction Lemma. The proof of our Block Restriction Lemma (Lemma 1.6) relies on the techniques in [Kan14]. An oversimplified proof sketch is as follows. We first show that if a degree-d multilinear polynomial is not "concentrated" (i.e., has the standard deviation much larger that the expectation), then it is expected to have a relatively large directional derivative compared to its actual value. We then use anticoncentration bounds for polynomials to argue that it is unlikely that a random restriction of a degree-d multilinear polynomial will have such a property.

One issue is that strong enough anticoncentration bounds for polynomials (e.g., the Carbery-Wright bound [CW01], or the bound from [Kan14] that we will actually use) are true only under the Gaussian measure rather than the uniform distribution over the boolean cube. To use these anticoncentration results, we thus need to move from the Bernoulli distribution to the Gaussian distribution over the inputs of polynomials. Such change of the probability measure is possible thanks to the celebrated Invariance Principle of [MOO10]. It applies to "regular" polynomials only, but fortunately there is a "regularity lemma" of [DSTW14] (or a variant from [Kan14]) that allows one to reduce the analysis of arbitrary polynomials to the case of regular ones, at a small cost.

The next problem is that the Invariance Principle incurs significant (and unavoidable) losses that have a bad dependence on the degree d of the polynomial in question. To mitigate such losses, we apply a random block restriction ρ in a series of few steps, viewing ρ as a composition of t restrictions $\rho_1 \circ \rho_2 \circ \cdots \circ \rho_t$ (for not too large value $t \geq 1$), where each ρ_i is on relatively small number of blocks m_i. This allows us to ensure that the loss from the Invariance Principle at each step i is "absorbed" by the parameter m_i.

Thus we get a recursive proof, where in each step we apply the regularity lemma, the Invariance Principle, and the anticoncentration bound. Carrying out such a proof directly, we get a weak version of the Block Restriction Lemma (see Lemma 4.1). By a more careful recursive analysis (using a "soft" measure of "non-concentration" for polynomials), we get the stronger version stated in Lemma 5.1.

Derandomized Block Restriction Lemma. To prove a derandomized version of the Block Restriction Lemma (Theorem 1.8), we first observe that a block restriction ρ makes a given degree-d polynomial "concentrated" if and only if a certain PTF of degree $2d$ evaluates to -1 on ρ. Thus finding a good-restriction ρ is reduced to the task of fooling degree-$2d$ PTFs. For the latter, we can use known constructions of pseudorandom generators (PRGs) for PTFs, e.g., the construction due to Meka and Zuckerman [MZ13]. Unfortunately, the parameters of the known PRGs for PTFs are far from optimal. Using such a PRG in a single step would yield a derandomized Block Restriction Lemma with very poor parameters. Instead, we use a recursive strategy similar to the recursive proof of Lemma 1.6 above. We build a pseudorandom block restriction in a sequence of steps, where in each single step we are facing a block partition on a relatively small number of blocks, and so can afford the relatively poor parameters of the PRG construction from [MZ13].

Derandomized Littlewood-Offord. To prove Theorem 1.9 and Theorem 1.10, we first argue that bounded-wise independent hash functions can be used to produce a block partition of input

variables such that, with high probability, every polynomial p satisfying the assumptions of these theorems will contain within each block a high-degree monomial with a large coefficient. Once we have a good partition, we can use our derandomized Block Restriction Lemma to generate the required pseudorandom inputs x for the polynomial p so that $p(x)$ is unlikely to be contained within a small interval.

Remainder of the paper. We give the necessary background in Section 2. In Section 3, we prove a simpler Block Restriction Lemma as a warm-up. In Section 4, we prove another Block Restriction Lemma that achieves the optimal exponent in the parameter m (the number of blocks). It is a weaker version of our final Block Restriction Lemma, which illustrates our proof techniques. The stronger version is then proved in Section 5. In Section 6, we give our applications of the Block Restriction Lemma: we prove Theorems 1.1–1.4; re-derive Kane's average sensitivity bound for degree-d PTFs in Section 6.4; and show a Littlewood-Offord type anticoncentration bound for degree-d multilinear polynomials in Section 6.5. We prove our derandomized restriction lemma and derandomized Littlewood-Offord type anticoncentration bounds in Section 7. Section 8 contains some open problems.

2 Preliminaries

Here we present some definitions and results on polynomials and polynomial threshold functions, and introduce some basic notions in the analysis of boolean functions. For more details on these (and related) topics, the reader is referred to [O'D14].

2.1 Notation

We will denote by X, Y, Z standard multidimensional Gaussian random variables. That is, for dimension n, we have $X \sim N(0,1)^n$, where $X = (X_1, \ldots, X_n)$ and all components $X_i \sim N(0,1)$ are independent Gaussians. Similarly, we denote by A, B, C multidimensional Bernoulli variables, where, for dimension n, $A = (A_1, \ldots, A_n)$, and all components $A_i \sim \{-1, 1\}$ are independent fair coin flips. Occasionally, for results that hold for both Gaussian and Bernoulli distributions, we use I, J to denote distributions that may be either standard n-dimensional Gaussian, or Bernoulli distributions.

2.2 Boolean functions and polynomial threshold functions

We think of an n-variate boolean function f as $f\colon \{-1,1\}^n \to \{-1,1\}$. For two boolean functions $f, g\colon \{-1,1\}^n \to \{-1,1\}$ and a parameter $\delta \in [0,1]$, we say that f and g are *δ-close* if $\mathbf{Pr}_{x\in\{-1,1\}^n}[f(x) \neq g(x)] \leq \delta$. We say that a Boolean function f is δ-close to constant if there is a constant function v, where $v = -1$ or $v = 1$, such that f and v are δ-close.

Definition 2.1. *A* degree-d polynomial threshold function (PTF) *is a function $f\colon \{-1,1\}^n \to \{-1,1\}$ of the form $f = \mathrm{sgn}(p)$, where $p\colon \mathbb{R}^n \to \mathbb{R}$ is a multilinear polynomial of degree at most d, and $\mathrm{sgn}\colon \mathbb{R} \to \{-1,1\}$ is the sign function defined to be 1 on all positive inputs, and -1 on all negative inputs and on 0.*[2]

[2]Without loss of generality, we may assume for every PTF $f = \mathrm{sgn}(p)$ that $p(x) \neq 0$ for all $x \in \{-1,1\}^n$. The reason is that we may always change p to a new polynomial $\tilde{p} = p + \eta$, for a small constant $\eta \in \mathbb{R}$, so that, for every

2.3 Concentration and anticoncentration for polynomials

Definition 2.2 (L^t norm)**.** *For $f\colon \mathbb{R}^n \to \mathbb{R}$ and a real number $t \geq 1$, the Gaussian (Bernoulli) L^t norm of f is defined as*

$$\|f\|_t = \left(\mathbf{Exp}[|f(I)|^t]\right)^{1/t},$$

where I is an n-dimensional Gaussian (Bernoulli) random variable.

It is easy to see that the Gaussian and Bernoulli L^2 norms are the same for any multilinear polynomial p, i.e.,

$$\mathbf{Exp}[|p(X)|^2] = \mathbf{Exp}[|p(A)|^2].$$

We denote by $\|p\|_2$ the L^2 norm of a multilinear polynomial p under Gaussian (or Bernoulli) distribution. For multilinear polynomials p, we also have

$$\mathbf{Exp}[p(X)] = \mathbf{Exp}[p(A)].$$

Hence the variance of p is the same under Gaussian and Bernoulli measures, and we will denote this variance by $\mathbf{Var}[p]$.

The hypercontractivity results of [Bon70] relate the L^t norm of a polynomial to its L^2 norm, both for Gaussian and Bernoulli measures. For multilinear d-degree polynomials p, the relevant hypercontractive inequality is

$$\|p\|_t \leq (t-1)^{d/2} \cdot \|p\|_2, \tag{1}$$

where the L^t norm on the left-hand side may be either Gaussian or Bernoulli.

The following strong concentration bound for polynomials is an immediate consequence of Equation (1) and the Markov inequality.

Theorem 2.3 (Concentration bound)**.** *For every d-degree multilinear polynomial $p\colon \mathbb{R}^n \to \mathbb{R}$, and for every $K \geq 2^d$, we have*

$$\mathbf{Pr}\left[|p(I)| \geq K \cdot \|p\|_2\right] \leq \exp\left(-(1/4) \cdot K^{2/d}\right),$$

where I is an n-dimensional Gaussian or Bernoulli random variable.

The following weak anticoncentration result for polynomials is also an immediate consequence of Equation (1) (for $t = 4$) and the Paley-Zygmund inequality (applied to p^2).

Theorem 2.4 (Weak anticoncentration bound)**.** *For every d-degree multilinear polynomial $p\colon \mathbb{R}^n \to \mathbb{R}$, we have*

$$\mathbf{Pr}\left[|p(I)| \geq (1/2) \cdot \|p\|_2\right] \geq (1/2) \cdot 9^{-d},$$

where I is an n-dimensional Gaussian or Bernoulli random variable.

A stronger anticoncentration result for polynomial with respect to the Gaussian measure, due to Carbery and Wright [CW01], shows that $|p(X)|$ is likely to exceed $\epsilon \cdot \|p\|_2$.

Theorem 2.5 (Anticoncentration bound [CW01])**.** *For any non-zero degree-d polynomial p and any $\epsilon > 0$, we have*

$$\mathbf{Pr}\left[|p(X)| \leq \epsilon \cdot \|p\|_2\right] = O(d \cdot \epsilon^{1/d}).$$

$x \in \{-1,1\}^n$, we have both $\mathrm{sgn}(p(x)) = \mathrm{sgn}(\tilde{p}(x))$ and $\tilde{p}(x) \neq 0$.

The anticoncentration bound above has poor dependence on the degree d. For an improved dependence on d, we use a version of the anticoncentration result due to Kane [Kan14], where $|p(X)|$ is compared to the directional derivative of p, rather than the norm of p.

Definition 2.6 (Directional derivative)**.** *For an n-variate function $g(x_1, \ldots, x_n)$ from $\mathbb{R}^n$ to $\mathbb{R}$, and $u, v \in \mathbb{R}^n$, the* directional derivative of g at u in the direction v, *denoted $\mathsf{D}_v\, g(u)$, is defined as*

$$\mathsf{D}_v\, g(u) = v \cdot \nabla g(u),$$

where

$$\nabla g = \left(\frac{\partial g}{\partial x_1}, \ldots, \frac{\partial g}{\partial x_n}\right)$$

is the gradient of g, and "$\cdot$" denotes the usual inner product of vectors.

Theorem 2.7 (Strong anticoncentration bound [Kan14])**.** *For any non-zero polynomial p of degree d and any $\epsilon > 0$, we have*

$$\mathbf{Pr}\left[|p(X)| \leq \epsilon \cdot |\mathsf{D}_Y\, p(X)|\right] = O\left(d^2 \cdot \epsilon\right).$$

This strong anticoncentration bound will be useful to us thanks to the following (easily provable) identity:

$$\mathbf{Exp}\left[|\mathsf{D}_J\, p(I)|^2\right] = \mathbf{Exp}\left[\|\nabla p(I)\|_2^2\right], \tag{2}$$

where I and J are either independent Gaussians, or independent Bernoulli distributions. In turn, the quantity on the right-hand of Equation (2) can be related to the variance $\mathbf{Var}[p]$, via the notion of *influence*, to be discussed in the next subsection.

We conclude this subsection with the following useful relationship between the directional derivative and the gradient.

Lemma 2.8. *For any non-negative integer k and any degree-d multilinear n-variate polynomial p such that $p(x) \neq 0$ for all $x \in \{-1, 1\}^n$, we have*

$$\mathbf{Exp}_{A,B}\left[\min\left\{k, \frac{|\mathsf{D}_B\, p(A)|^2}{|p(A)|^2}\right\}\right] \leq \mathbf{Exp}_A\left[\min\left\{k, \frac{\|\nabla p(A)\|_2^2}{|p(A)|^2}\right\}\right], \tag{3}$$

$$\mathbf{Exp}_A\left[\min\left\{k, \frac{\|\nabla p(A)\|_2^2}{|p(A)|^2}\right\}\right] \leq 72 \cdot \mathbf{Exp}_{A,B}\left[\min\left\{k, \frac{|\mathsf{D}_B\, p(A)|^2}{|p(A)|^2}\right\}\right]. \tag{4}$$

Proof. Note that, for any random variable Z, we have $\mathbf{Exp}[\min\{k, Z\}] \leq k$ and $\mathbf{Exp}[\min\{k, Z\}] \leq \mathbf{Exp}[Z]$. Thus,

$$\mathbf{Exp}_{A,B}\left[\min\left\{k, \frac{|\mathsf{D}_B\, p(A)|^2}{|p(A)|^2}\right\}\right] \leq \mathbf{Exp}_A\left[\min\left\{k, \mathbf{Exp}_B\left[\frac{|\mathsf{D}_B\, p(A)|^2}{|p(A)|^2}\right]\right\}\right],$$

which implies Equation (3).

We now prove Equation (4). For a fixed A, let

$$q_A(B) = B \cdot \frac{\nabla p(A)}{|p(A)|} = \frac{\mathsf{D}_B\, p(A)}{|p(A)|}.$$

Note that q_A is a degree-1 polynomial in the variables B with $\|q_A\|_2^2 = \frac{\|\nabla p(A)\|_2^2}{|p(A)|^2}$. Then by Theorem 2.4, we have for every A that

$$\Pr_B\left[\frac{|\mathsf{D}_B\, p(A)|^2}{|p(A)|^2} \geq (1/4)\cdot\frac{\|\nabla p(A)\|_2^2}{|p(A)|^2}\right] = \Pr_B\left[|q_A(B)|^2 \geq (1/4)\cdot\|q_A\|_2^2\right] \geq 1/18. \tag{5}$$

Now let

$$X = X(A) = \frac{\|\nabla p(A)\|_2^2}{|p(A)|^2},$$

and

$$Y = Y(A,B) = \frac{|\mathsf{D}_B\, p(A)|^2}{|p(A)|^2}.$$

By Equation (5), $Y \geq X/4$ with probability at least $1/18$. Next we have

$$\begin{aligned}
\mathbf{Exp}_{A,B}\left[\min\{k,Y\}\right] &\geq \frac{1}{18}\cdot\mathbf{Exp}_A\left[\mathbf{Exp}_B\left[\min\{k,Y\} \mid Y \geq X/4\right]\right] \\
&\geq \frac{1}{18}\cdot\mathbf{Exp}_A\left[\mathbf{Exp}_B\left[\min\{k,X/4\} \mid Y \geq X/4\right]\right] \\
&\geq \frac{1}{72}\cdot\mathbf{Exp}_A\left[\min\{k,X\}\right],
\end{aligned}$$

where in the last step we dropped the expectation over B since X does not depend on B. □

2.4 Invariance principle for polynomials

For a boolean function $f\colon\{-1,1\}^n \to \{-1,1\}$ and $i \in [n]$, the *influence of coordinate i on f*, denoted $\mathbf{Inf}_i[f]$, is defined as

$$\mathbf{Inf}_i[f] = \mathbf{Pr}_{x\sim\{-1,1\}^n}[f(x) \neq f(x^{\oplus i})],$$

where $x^{\oplus i}$ is x with the ith coordinate x_i replaced with $-x_i$. The *total influence* (also known as average sensitivity) of $f\colon\{-1,1\}^n \to \{-1,1\}$, denoted $\mathbf{Inf}[f]$, is defined as

$$\mathbf{Inf}[f] = \sum_{i=1}^{n}\mathbf{Inf}_i[f].$$

For a function $f\colon\{-1,1\}^n \to \mathbb{R}$, the definition of influence becomes:

$$\mathbf{Inf}_i[f] = \frac{1}{4}\cdot\mathbf{Exp}_{x\sim\{-1,1\}^n}\left[\left|f(x) - f(x^{\oplus i})\right|^2\right].$$

For the case of multilinear polynomials $p\colon\mathbb{R}^n \to \mathbb{R}$, it can be equivalently expressed as follows:

$$\mathbf{Inf}_i[p] = \left\|\frac{\partial p}{\partial x_i}\right\|_2^2, \tag{6}$$

yielding

$$\mathbf{Inf}[p] = \mathbf{Exp}\left[\|\nabla p(A)\|_2^2\right]. \tag{7}$$

The following is a well-known fact about influence; we sketch the proof for completeness.

Theorem 2.9. *For every d-degree multilinear polynomial $p\colon \mathbb{R}^n \to \mathbb{R}$, we have*

$$\mathbf{Var}[p] \le \mathbf{Inf}[p] \le d \cdot \mathbf{Var}[p].$$

Proof. For a multilinear polynomial $p(x_1, \ldots, x_n)$ and a set $S \subseteq [n]$, denote by $\hat{p}(S)$ the coefficient of p at the monomial $\prod_{i \in S} x_i$ (the Fourier coefficient of p at S). The proof is obtained from the following (easily verifiable) identities: $\mathbf{Inf}_i[p] = \sum_{S \ni i} \hat{p}(S)^2$, and $\mathbf{Var}[p] = \sum_{\emptyset \neq S \subseteq [n]} \hat{p}(S)^2$. □

Definition 2.10 (τ-regular)**.** *We say that a polynomial p is τ-regular if for all i.*

$$\mathbf{Inf}_i[p] \le \tau \cdot \mathbf{Var}[p].$$

A polynomial threshold function $f(x) = \mathrm{sgn}(p(x))$ is ϵ-regular if p is τ-regular.

The following result shows that, for every PTF f, there exists a partitioning of the boolean cube $\{-1, 1\}^n$ into few sub-cubes so that, on most of these sub-cubes, the PTF f restricted to the sub-cube is either regular or has small variance relative to its L^2 norm.

Theorem 2.11 ([Kan14])**.** *For all $1/4 > \tau, \delta, \epsilon > 0$ and $\gamma > 0$, every degree-d multilinear polynomial p can be expressed as a decision tree of depth at most*

$$\tau^{-1} \cdot (d \cdot \log \tau^{-1} \cdot \log \delta^{-1})^{O(\gamma \cdot d)} \cdot \log \epsilon^{-1},$$

so that, with probability at least $1 - \epsilon$, a random leaf ω (reached from the root of the tree by branching uniformly at random at each internal node) defines a restricted polynomial p_ω (obtained from p by setting the variables on the branch leading to ω to the values specified by the branch) such that the polynomial $p_\omega(y)$ either is τ-regular or satisfies $\mathbf{Var}[p_\omega] \le \left(\log \delta^{-1}\right)^{-\gamma \cdot d} \cdot \|p_\omega\|_2^2$.

The following is a version of the Invariance Principle of Mossel et al. [MOO10] in the form that will be convenient for us.

Theorem 2.12 (Invariance principle [Kan14])**.** *Let p and q be two polynomials such that for some $\tau > 0$, $\mathbf{Inf}_i[p], \mathbf{Inf}_i[q] \le \tau$ for all i and that $\|p + q\|_2, \|p - q\|_2 \ge 1$. Then*

$$\mathbf{Pr}[|p(A)| \le |q(A)|] = \mathbf{Pr}[|p(X)| \le |q(Y)|] + O\left(d \cdot \tau^{-1/8d}\right).$$

Corollary 2.13. *For any d-degree τ-regular non-constant multilinear polynomial p, and any $\epsilon > 0$, we have*

$$\mathbf{Pr}[|p(A)| \le \epsilon \cdot |\mathsf{D}_B\, p(A)|] = O\left(d^2 \cdot \epsilon + d \cdot \tau^{-1/8d}\right).$$

Proof. The idea is to apply the Invariance Principle of Theorem 2.12, and then the strong anticoncentration bound of Theorem 2.7. To this end, we need to argue that the assumptions of these two theorems are satisfied. First, we normalize our polynomial p so that the new polynomial has all influences at most τ.

For the given multilinear polynomial $p(x_1, \ldots, x_n)$, define

$$q(x_1, \ldots, x_n, y_1, \ldots, y_n) = \sum_{i=1}^{n} y_i \cdot \frac{\partial p}{\partial x_i}, \tag{8}$$

which is easily seen to be a multilinear polynomial of degree at most d. Let $\sigma = \sqrt{\mathbf{Var}[p]}$. Since p is a non-constant function, we have $\sigma \neq 0$. Define the normalized polynomial $p' = p/\sigma$. We have $\mathbf{Var}[p'] = 1$, and, by the definition of influence in Equation (6), we also have for all $i \in [n]$ that

$$\begin{aligned}\mathbf{Inf}_i[p'] &= \mathbf{Inf}_i[p]/\mathbf{Var}[p] \\ &\leq \tau, \end{aligned} \tag{9}$$

since $\mathbf{Inf}_i[p] \leq \tau \cdot \mathbf{Var}[p]$ for all $i \in [n]$ by assumption. By the linearity of differentiation, we also get that $q' = q/\sigma$ is the directional derivative of p'. Thus our task is reduced to upper-bounding the probability

$$\mathbf{Pr}[|p'(A)| \leq \epsilon \cdot |\mathsf{D}_B\, p'(A)|]. \tag{10}$$

To apply the Invariance Principle of Theorem 2.12 to Equation (10), we need to upper-bound the influences of q'. For every $i \in [n]$, we get from Equations (8) and (9) that

$$\begin{aligned}\left\| \frac{\partial q'}{\partial y_i} \right\|_2^2 &= \left\| \frac{\partial p'}{\partial x_i} \right\|_2^2 \\ &= \mathbf{Inf}_i[p'] \\ &\leq \tau. \end{aligned}$$

We also get

$$\begin{aligned}\left\| \frac{\partial q'}{\partial x_i} \right\|_2^2 &= \mathbf{Exp}\left[\left| \mathsf{D}_B \frac{\partial p'}{\partial x_i}(A) \right|^2 \right] \\ &= \mathbf{Exp}\left[\left\| \nabla \frac{\partial p'}{\partial x_i}(A) \right\|_2^2 \right] && \text{(by Equation (2))} \\ &= \mathbf{Inf}\left[\frac{\partial p'}{\partial x_i} \right] && \text{(Equation (7))} \\ &\leq d \cdot \mathbf{Var}\left[\frac{\partial p'}{\partial x_i} \right] && \text{(by Theorem 2.9)} \\ &\leq d \cdot \left\| \frac{\partial p'}{\partial x_i} \right\|_2^2 \\ &= d \cdot \mathbf{Inf}_i[p'] && \text{(by Equation (6))} \\ &\leq d\tau. && \text{(by Equation (9))} \end{aligned}$$

Thus all of the influences of p' and q' are at most $d\tau$.

Finally, for every $\lambda \in \mathbb{R}$, we have for independent n-dimensional standard Gaussians A and B that

$$\begin{aligned}\left\| p' + \lambda \cdot q' \right\|_2^2 &= \mathbf{Exp}\left[|p'(A) + \lambda \cdot q'(A,B)|^2 \right] \\ &= \mathbf{Exp}[|p'(A)|^2] + \lambda^2 \cdot \mathbf{Exp}[|q'(A,B)|^2] + (2\lambda) \cdot \mathbf{Exp}[p'(A) \cdot q'(A,B)]. \end{aligned} \tag{11}$$

By Equation (8), we get that $\mathbf{Exp}[p'(A) \cdot q'(A,B)] = 0$. Hence, we get from Equation (11) that $\|p' + \lambda \cdot q'\|_2^2 \geq \|p'\|_2^2 \geq \mathbf{Var}[p'] = 1$.

Thus p' and q' satisfy all assumptions of Theorem 2.12, with the influences bounded by $d\tau$. Applying Theorem 2.12 and then Theorem 2.7, we get the required upper bound on the probability in Equation (10), concluding the proof. □

2.5 Random block restrictions and concentrated polynomials

Definition 2.14 (Random block restriction)**.** *Suppose the variables of a polynomial are arbitrarily partitioned into m blocks. A random block restriction ρ is obtained by the following process:*

1. *Uniformly at random pick a block $\ell \in [m]$.*
2. *Assign each variable that is outside the chosen block ℓ a uniformly random value in $\{-1, 1\}$, independently.*

We use B_m to denote the distribution over all possible restrictions ρ generated by the above process.

We need the following notion of "concentration" for polynomials.

Definition 2.15 (δ-concentrated polynomials)**.** *Let p be a degree-d multilinear polynomials and $f = \mathrm{sgn}(p)$. For a universal constant $L = 192$, and parameters $0 < \delta \le 1/2$ and $\gamma > 0$, we call p (and f) (δ, γ)-*concentrated *if*

$$\mathbf{Var}[p] \le \left(L \cdot \log \delta^{-1}\right)^{-\gamma \cdot d} \cdot \|p\|_2^2 .$$

We refer to $(\delta, 1)$-concentrated polynomials as δ-concentrated.

A useful property of concentrated PTFs is that they are close to constant.

Lemma 2.16. *For every degree PTF $f = \mathrm{sgn}(p)$ and every $0 < \delta \le 1/2$, if p is δ-concentrated, then f is δ^2-close to constant.*

Proof. Let $p' = p - \mu$, where $\mu = \mathbf{Exp}[p(A)]$, and let $\nu = (L \cdot \log \delta^{-1})^d$ for a constant $L > 0$ to be determined. Since p is δ-concentrated and $\|p\|_2^2 = \mu^2 + \mathbf{Var}[p]$, we get

$$\mu^2 \ge (\nu - 1) \cdot \mathbf{Var}[p] \ge \frac{\nu}{4} \cdot \mathbf{Var}[p],$$

for $L \ge 4/3$. Thus we have

$$|\mu| \ge \frac{\sqrt{\nu}}{2} \cdot \left\|p'\right\|_2 . \tag{12}$$

Note that for all points $x \in \{-1, 1\}^n$ where $|p'(x)| < |\mu|$, we have $\mathrm{sgn}(p(x)) = \mathrm{sgn}(\mu)$. Therefore,

$$\begin{aligned}
\mathbf{Pr}[\mathrm{sgn}(p(A)) \ne \mathrm{sgn}(\mu)] &\le \mathbf{Pr}[|p'(x)| \ge |\mu|] \\
&\le \mathbf{Pr}\left[|p'(x)| \ge \frac{\sqrt{\nu}}{2} \cdot \left\|p'\right\|_2\right] && \text{(by Equation (12))} \\
&\le \delta^2, && \text{(by Theorem 2.3)}
\end{aligned}$$

where the last inequality holds if we choose $L \ge 32$. □

3 Block Restriction Lemma: A simple bound

As a warm-up, we first prove a simpler bound on the probability that under random block restrictions, a degree-d multilinear polynomial does not become concentrated.

Lemma 3.1 (Block Restriction Lemma: Simple Bound)**.** *For any degree-d multilinear polynomial p, and any $m \ge 16$, $\gamma \ge 1$, $0 < \delta \le 1/16$, we have*

$$\mathbf{Pr}_{\rho \sim B_m}\left[p_\rho \text{ is not } (\delta, \gamma)\text{-concentrated}\right] \le m^{-\frac{1}{8d+1}} \cdot \left(d \cdot \log m \cdot \log \delta^{-1}\right)^{O(\gamma \cdot d)} + 2\delta.$$

3.1 Regularization

In this section, we show that it suffices to consider only regular polynomials. We start with the following definition.

Definition 3.2. *Let $\mathcal{P}(d, m, \delta, \gamma)$ be the supremum, over all degree-d multilinear polynomials p and all possible partitions of the variables into m blocks, of the probabilities*

$$\mathbf{Pr}_{\rho \sim B_m}\left[p_\rho \text{ is not } (\delta, \gamma)\text{-concentrated}\right].$$

Let $\mathcal{P}_{\mathrm{reg}}(d, m, \delta, \gamma, \tau)$ be the same as $\mathcal{P}$ but only for τ-regular polynomials. We will use $\mathcal{P}(d, m, \delta)$ (resp. $\mathcal{P}_{\mathrm{reg}}(d, m, \delta, \tau)$) for $\mathcal{P}(d, m, \delta, 1)$ (resp. $\mathcal{P}_{\mathrm{reg}}(d, m, \delta, 1, \tau)$).

Claim 3.3. *For any $m \geq 1/16$, $\gamma > 0$, and $0 < \delta, \tau \leq 1/4$, we have*

$$\mathcal{P}(d, m, \delta, \gamma) \leq \mathcal{P}_{\mathrm{reg}}(d, r, \delta, \gamma, \tau) + \frac{1}{m} \cdot \tau^{-1} \cdot \left(d \cdot \log \tau^{-1} \cdot \log \delta^{-1}\right)^{O(\gamma \cdot d)} + 2\delta.$$

Proof. Let p be a degree-d multilinear polynomial with its variables petitioned into m blocks. By Theorem 2.11, there exists a decision tree of depth at most

$$H = \tau^{-1} \cdot (d \cdot \log \tau^{-1} \cdot \log \delta^{-1})^{O(\gamma \cdot d)}$$

such that for a random leaf ω of the tree, the restricted polynomial p_ω (obtained from p by fixing the variables on the branch leading to ω, as specified by the branch) is either τ-regular or $(\delta, \gamma+1)$-concentrated, with probability at least $1 - \delta$. The given decision tree partitions the boolean cube $\{-1, 1\}^n$ into disjoint regions (sub-cubes) according to the partial restrictions labelling the branches of the tree. Let us call a random restriction ρ *partition-respecting* if it is consistent with some partial restriction labelling one of the branches of the decision tree (i.e., the restriction does not select a block containing any of the variables appearing on the branch, and the assignment to those variables agrees with their corresponding values on the branch). We claim that the probability that a random restriction $\rho \sim B_m$ is *not* partition-respecting is at most H/m.

Indeed, first note that choosing a random restriction $\rho \sim B_m$ is equivalent to first picking a uniformly random assignment to all variables, and then un-assigning the variables in a uniformly random block $i \in [m]$. Picking a uniformly random assignment to all variables is equivalent to picking a random branch in our decision tree (setting some of the variables to constants), and then randomly assigning the remaining variables (not appearing on the branch). For each fixed variable on the branch, the probability that its corresponding block is chosen when we pick a uniformly random block $i \in [m]$ is $1/m$. It follows by the union bound that the overall probability that a random block $i \in [m]$ contains some variable from the given branch is at most H/m.

Thus, at the expense of the additive error term

$$\frac{H}{m} = \frac{1}{m} \cdot \tau^{-1} \cdot \left(d \cdot \log \tau^{-1} \cdot \log \delta^{-1}\right)^{O(\gamma \cdot d)},$$

it suffices to upper-bound the probability

$$\mathbf{Pr}_{\rho \sim B_m}[p_\rho \text{ is not } (\delta, \gamma)\text{-concentrated}]$$

only for partition-respecting restrictions ρ. This probability can be expressed as the expectation over random leaves ω of the decision tree for f of the probability

$$\mathbf{Pr}_{\rho' \sim B_m}[(p_\omega)_{\rho'} \text{ is not } (\delta, \gamma)\text{-concentrated}],$$

where f_ω is the restriction of f to the leaf ω, and ρ' is a random restriction on the variables of f_ω (those not fixed by the branch leading to ω).

Finally, as p_ω is neither τ-regular nor $(\delta, \gamma+1)$-concentrated with probability at most δ over random leaves ω, and a polynomial that is $(\delta, 2)$-concentrated will stay $(\delta, 1)$-concentrated with probability at least $1-\delta$ by Lemma 4.2, we get that

$$\mathcal{P}(d, m, \delta, \gamma) \leq \mathcal{P}_{\mathrm{reg}}(d, m, \delta, \gamma, \tau) + \frac{1}{m} \cdot \tau^{-1} \cdot \left(d \cdot \log \tau^{-1} \cdot \log \delta^{-1}\right)^{O(\gamma \cdot d)} + 2\delta,$$

as required. □

3.2 Proof of the simple bound

Given Claim 3.3, it remains to upper-bound the quantity $\mathcal{P}_{\mathrm{reg}}(d, m, \delta, \gamma, \tau)$. Here we show the following.

Lemma 3.4. *There is a constant $c > 0$ such that, for any $m \geq 1/16$, $\gamma \geq 1$, and $0 < \delta, \tau \leq 1/4$, we have*

$$\mathcal{P}_{\mathrm{reg}}(d, m, \delta, \gamma, \tau) \leq O\left(d^2\right) \cdot (c \cdot \log \delta^{-1})^{\gamma \cdot d} \cdot \left(m^{-1/2} + \tau^{1/(8d)}\right).$$

First we prove the following property of non-concentrated polynomials.

Lemma 3.5. *For any $0 < \delta \leq 1/4$ and $\gamma \geq 1$, if a degree-d multilinear polynomial p is not (δ, γ)-concentrated, then*

$$\mathbf{Pr}\left[|\mathsf{D}_B\, p(A)|^2 \geq (1/16) \cdot \left((9L) \cdot \log \delta^{-1}\right)^{-\gamma \cdot d} \cdot |p(A)|^2\right] \geq (1/4) \cdot 9^{-d},$$

where $L > 0$ is the constant from Definition 2.15.

Proof. For the given multilinear polynomial $p(x_1, \ldots, x_n)$, define

$$q(x_1, \ldots, x_n, y_1, \ldots, y_n) = \sum_{i=1}^{n} y_i \cdot \frac{\partial p}{\partial x_i},$$

which is easily seen to be a multilinear polynomial of degree at most d. Applying Theorem 2.4 to q, we get that

$$\mathbf{Pr}\left[|q(C)|^2 \geq (1/4) \cdot \|q\|_2^2\right] \geq (1/2) \cdot 9^{-d}. \tag{13}$$

Next we relate $\|q\|_2^2$ to $\mathbf{Var}[p]$ as follows:

$$\begin{aligned}
\|q\|_2^2 &= \mathbf{Exp}\left[|\mathsf{D}_B\, p(A)|^2\right] \\
&= \mathbf{Exp}\left[\|\nabla p(A)\|_2^2\right] && \text{(by Equation (2))} \\
&= \mathbf{Inf}[p] && \text{(by Equation (7))} \\
&\geq \mathbf{Var}[p]. && \text{(by Theorem 2.9)}
\end{aligned}$$

Together with Equation (13), this implies that

$$\mathbf{Pr}\left[|\mathsf{D}_B\, p(A)|^2 \geq (1/4) \cdot \mathbf{Var}[p]\right] \geq (1/2) \cdot 9^{-d}. \tag{14}$$

Applying the Markov inequality to p^2, we get

$$\mathbf{Pr}\left[|p(A)|^2 \geq (4 \cdot 9^d) \cdot \|p\|_2^2\right] \leq (1/4) \cdot 9^{-d}. \tag{15}$$

We conclude from Equations (14) and (15) that, with probability at least $(1/4) \cdot 9^{-d}$, we have both

$$|\mathsf{D}_B\, p(A)|^2 \geq (1/4) \cdot \mathbf{Var}[p] \tag{16}$$

and

$$|p(A)|^2 \leq (4 \cdot 9^d) \cdot \|p\|_2^2 . \tag{17}$$

As p is assumed to be (δ, γ)-concentrated, we also have

$$\mathbf{Var}[p] \geq \left(L \cdot \log \delta^{-1}\right)^{-\gamma \cdot d} \cdot \|p\|_2^2 . \tag{18}$$

Combining Equations (16) to (18) yields the required claim. □

Definition 3.6. *For p a non-zero polynomial, we define*

$$\alpha(p) \colon = \mathbf{Exp}\left[\min\left\{1, \frac{|\mathsf{D}_B\, p(A)|^2}{|p(A)|^2}\right\}\right].$$

By Lemma 3.5, we get the following.

Corollary 3.7. *There is a constant $c > 0$ such that, for any $0 < \delta \leq 1/4$ and $\gamma \geq 1$, if a degree-d multilinear polynomial p is not (δ, γ)-concentrated, then*

$$\alpha(p) \geq \left(c \cdot \log \delta^{-1}\right)^{-\gamma \cdot d}.$$

We are now ready to prove Lemma 3.4.

Proof of Lemma 3.4. For a block, ℓ, we let A_ℓ denote the random assignment to the variables in ℓ and let $A_{\bar{\ell}}$ denote the random assignment to the variables that are not in ℓ. Then by the definition of random block restriction, we have

$$\begin{aligned}
\mathbf{Pr}_{\rho \sim B_m}[p_\rho \text{ is not } (\delta,\gamma)\text{-concentrated}] &= \frac{1}{m} \cdot \sum_\ell \mathbf{Pr}_{A_{\bar{\ell}}}[f_{A_{\bar{\ell}}} \text{ is not } (\delta,\gamma)\text{-concentrated}] \\
&\leq \frac{1}{m} \cdot \sum_\ell \mathbf{Pr}_{A_{\bar{\ell}}}\left[\left(c \cdot \log \delta^{-1}\right)^{\gamma \cdot d} \cdot \alpha(p_{A_{\bar{\ell}}}) \geq 1\right] \\
&\qquad\qquad \text{(by Corollary 3.7)} \\
&\leq \frac{1}{m} \cdot \sum_\ell \mathbf{Exp}_{A_{\bar{\ell}}}\left[\left(c \cdot \log \delta^{-1}\right)^{\gamma \cdot d} \cdot \alpha(p_{A_{\bar{\ell}}})\right] \\
&= \frac{1}{m} \cdot \left(c \cdot \log \delta^{-1}\right)^{\gamma \cdot d} \cdot \sum_\ell \mathbf{Exp}_{A_{\bar{\ell}}}\left[\alpha\left(p_{A_{\bar{\ell}}}\right)\right]. \qquad (19)
\end{aligned}$$

We upper-bound the sum $\sum_\ell \mathbf{Exp}_{A_{\bar{\ell}}}\left[\alpha\left(p_{A_{\bar{\ell}}}\right)\right]$ in Equation (19) as follows:

$$\begin{aligned}
\sum_\ell \mathbf{Exp}\left[\min\left\{1, \frac{|\mathsf{D}_B\, p_{A_{\bar{\ell}}}(A_\ell)|^2}{|p_{A_{\bar{\ell}}}(A_\ell)|^2}\right\}\right] &\le \sum_\ell \mathbf{Exp}\left[\min\left\{1, \frac{\left\|\nabla p_{A_{\bar{\ell}}}(A_\ell)\right\|^2}{|p_{A_{\bar{\ell}}}(A_\ell)|^2}\right\}\right] && \text{(by Lemma 2.8)}\\
&\le \mathbf{Exp}\left[\min\left\{m, \frac{\|\nabla p(A)\|^2}{|p(A)|^2}\right\}\right]\\
&\le 72 \cdot \mathbf{Exp}\left[\min\left\{m, \frac{|\mathsf{D}_B\, p(A)|^2}{|p(A)|^2}\right\}\right] && \text{(by Lemma 2.8)}
\end{aligned}$$

If p is a constant function, then its directional derivative is always 0, and hence the expectation above becomes 0. Otherwise, for a non-constant p, we upper-bound this expectation by

$$\begin{aligned}
&\mathbf{Exp}\left[\min\left\{1, \frac{|\mathsf{D}_B\, p(A)|^2}{|p(A)|^2}\right\}\right] + \int_1^m \mathbf{Pr}\left[|p(A)| \le t^{-1/2} \cdot |\mathsf{D}_B\, p(A)|\right] dt\\
&\le 1 + \int_1^m O\left(d^2 t^{-1/2} + d\tau^{1/(8d)}\right) dt && \text{(by Corollary 2.13)}\\
&\le O(d^2) \cdot \left(\sqrt{m} + m \cdot \tau^{1/(8d)}\right). && (20)
\end{aligned}$$

Combining Equations (19) and (20), we conclude that

$$\mathbf{Pr}_{\rho\sim B_m}[p_\rho \text{ is not } (\delta,\gamma)\text{-concentrated}] \le O\left(d^2\right) \cdot (c \cdot \log \delta^{-1})^{\gamma\cdot d} \cdot \left(m^{-1/2} + \tau^{1/(8d)}\right),$$

as required. □

We can now finish the proof of Lemma 3.1.

Proof of Lemma 3.1. Combining Claim 3.3 and Lemma 3.4, we get

$$\begin{aligned}
\mathcal{P}(d, m, \delta, \gamma) \le\ & O\left(d^2\right) \cdot (c \cdot \log \delta^{-1})^{\gamma\cdot d} \cdot \left(m^{-1/2} + \tau^{1/(8d)}\right)\\
&+ 2\delta + \frac{1}{m} \cdot \tau^{-1} \cdot \left(d \cdot \log \tau^{-1} \cdot \log \delta^{-1}\right)^{O(\gamma\cdot d)}.
\end{aligned}$$

Setting $\tau = m^{-\frac{8d}{8d+1}}$, we get the desired bound. □

4 Block Restriction Lemma with Optimal Exponent: Weak version

The bound in Lemma 3.1 has an undesirable dependence on the degree d in the exponent of the interested parameter m. In this section, we prove a better bound that achieves the optimal exponent in m. To illustrate the ideas of the proof techniques, we first prove the following weaker version.

Lemma 4.1 (Block Restriction Lemma: Weak Version)**.** *For any degree-d multilinear polynomial p, and any $m \ge 16$, $\gamma \ge 1$, and $0 < \delta \le 1/16$, we have*

$$\mathbf{Pr}_{\rho\sim B_m}\left[p_\rho \textit{ is not } (\delta,\gamma)\textit{-concentrated}\right] \le \left(m^{-1/2} + \delta\right) \cdot O\left(d \cdot \log m \cdot \log \delta^{-1}\right)^{O\left(\gamma\cdot d^2 \cdot \log\log m\right)}.$$

Our road map for the proof is as follows. We set up a recurrence (in Section 4.1), reducing the analysis of random restrictions from B_m to that of random restrictions from $B_{m/b}$, for a parameter $b > 0$. Solving this recurrence (in Section 4.2) will conclude the proof of Lemma 4.1.

The reason for the recursive analysis is to be able to control the error coming from an application of the invariance principle, Theorem 2.12. That error (of the form $\tau^{1/(8d)}$) has an undesirable dependence on the degree d in the exponent, which would be overwhelming if we try to apply a random block restriction $\rho \sim B_m$ in a single step, as in the proof of Lemma 3.1 in Section 3. However, by viewing ρ as a two-step process, where we first apply a random block restriction $\rho_1 \sim B_b$, and then apply another random block restriction $\rho_2 \sim B_{m/b}$, we only need to ensure that the error coming from the invariance principle is small relative to the value of $1/b$ (more precisely, $b^{-1/2}$). By choosing b so that $b^{-1/2}$ is equal to $\tau^{1/(8d)}$, we ensure that the error from the invariance principle is not overwhelming when we reduce from the case of B_m to the case of $B_{m/b}$. Then we repeat this recursive process enough times to get the final bound.

For simplicity, we only prove Lemma 4.1 for $\gamma = 1$. It is easy to modify the proof for any γ.

4.1 Setting up the recurrence

By Claim 3.3, we have

$$\begin{aligned}\mathbf{Pr}_{\rho\sim B_m}[p_\rho \text{ is not } \delta\text{-concentrated}] \leq \mathbf{Pr}_{\rho\sim B_m}[q_\rho \text{ is not } \delta\text{-concentrated}] \\ + \frac{1}{m}\cdot\tau^{-1}\cdot\left(d\cdot\log\tau^{-1}\cdot\log\delta^{-1}\right)^{O(d)} + 2\delta,\end{aligned}$$

where q is some τ-regular polynomial of degree at most d.

We now upper bound

$$\mathbf{Pr}_{\rho\sim B_m}[q_\rho \text{ is not } \delta\text{-concentrated}].$$

Consider the following equivalent way of choosing a random block restriction $\rho \sim B_m$. Let $0 < b \leq m$ be an integer parameter.

1. Partition the m blocks of variables of p into b disjoint super-blocks, where each super-block has m/b blocks.
2. Uniformly at random pick a super-block $\ell \in [b]$, and assign each variable that is outside the chosen super-block ℓ a uniformly random value in $\{-1, 1\}$, independently.
3. Uniformly at random pick a block within super-block ℓ, and assign each variable that is outside the chosen block a uniformly random value in $\{-1, 1\}$, independently.

To avoid some technicalities due to divisibility that can be overcome easily by adding dummy blocks, we assume here that m is divisible by b.

Note that step 2 above is an application of random block restriction on b blocks, and step 3 is an application of random block restriction on m/b blocks. Then we have

$$\mathbf{Pr}_{\rho\sim B_m}[q_\rho \text{ is not } \delta\text{-concentrated}] = \mathbf{Pr}_{\rho_1\sim B_b,\rho_2\sim B_{m/b}}[(q_{\rho_1})_{\rho_2} \text{ is not } \delta\text{-concentrated}].$$

Let $E(\rho_1)$ denote the random event that q_{ρ_1} is $(\delta, 2)$-concentrated. By conditioning on this event, we get that the probability above equals

$$\begin{aligned}&\mathbf{Pr}_{\rho_1,\rho_2}[(q_{\rho_1})_{\rho_2} \text{ is not } \delta\text{-concentrated} \mid E(\rho_1)]\cdot\mathbf{Pr}_{\rho_1}[E(\rho_1)] \\ &+ \mathbf{Pr}_{\rho_1,\rho_2}[(q_{\rho_1})_{\rho_2} \text{ is not } \delta\text{-concentrated} \mid \neg E(\rho_1)]\cdot\mathbf{Pr}_{\rho_1}[\neg E(\rho_1)]. \qquad (21)\end{aligned}$$

The first summand in Equation (21) contains the quantity

$$\mathbf{Pr}_{\rho_1,\rho_2}[(q_{\rho_1})_{\rho_2} \text{ is not } \delta\text{-concentrated} \mid q_{\rho_1} \text{ is } (\delta,2)\text{-concentrated}]. \tag{22}$$

To bound this quantity, we use the following which says a concentrated polynomial is likely to remain concentrated under random block restrictions.

Lemma 4.2. *For any $m > 0$, if a degree-d multilinear polynomial p is $(\delta, \gamma+1)$-concentrated, then*

$$\mathbf{Pr}_{\rho\sim B_m}[p_\rho \textit{ is not } (\delta,\gamma)\textit{-concentrated}] \le \delta.$$

Proof. Fix an arbitrary block $\ell \in [m]$. Let S be the set of variables in block ℓ and $\bar{S}$ be the set of variables outside block ℓ (i.e., S is the set of unrestricted variables and $\bar{S}$ is the set of restricted variables). Then we can write

$$p(A_S, A_{\bar{S}}) = q(A_S, A_{\bar{S}}) + r(A_{\bar{S}}) + \mu,$$

where r contains all the monomials in p that only depend on variables in $\bar{S}$, and $\mu = \mathbf{Exp}[p(A)]$. Also, for $\rho \in \{-1,1\}^{|\bar{S}|}$, let $\mu'(\rho) = r(\rho) + \mu$, and define $Q(\rho) = \|q(A_S,\rho)\|_2^2 = \mathbf{Var}[p_\rho]$. It can be shown (see, e.g., [DSTW10, proof of Lemma 6]) that Q is a degree-$2d$ polynomial with

$$\|Q\|_2 \le 3^d \cdot \sum_{i\in S} \mathbf{Inf}_i[p].$$

Thus, we have

$$\mathbf{Inf}[p] \ge \frac{1}{3^d} \cdot \|Q\|_2 . \tag{23}$$

Now let $\nu = \left(L \cdot \log \delta^{-1}\right)^d$, where $L > 0$ is the constant from Definition 2.15. Then we want to show

$$\mathbf{Pr}_\rho \left[\mathbf{Var}[p_\rho] \ge \nu^{-\gamma} \cdot \|p_\rho\|_2^2\right] \le \delta. \tag{24}$$

Note that to show Equation (24), it suffices to show

$$\mathbf{Pr}_\rho \left[Q(\rho) \ge \nu^{-\gamma} \cdot |\mu'(\rho)|^2\right] \le \delta. \tag{25}$$

We first prove the following claim.

Claim 4.3. *We have*

$$\mathbf{Pr}_\rho \left[|\mu'(\rho)|^2 < \frac{\nu^{\gamma+1}}{8\cdot d\cdot 3^d} \cdot \|Q\|_2\right] \le \delta/2.$$

Proof. We have for all ρ,

$$\begin{aligned} |\mu'(\rho)| &= |\mu + r(\rho)| \\ &\ge |\mu| - |r(\rho)| \\ &\ge \sqrt{(\nu^{\gamma+1}-1)\cdot \mathbf{Var}[p]} - |r(\rho)| \\ &\ge \sqrt{\frac{\nu^{\gamma+1}}{2}\cdot \mathbf{Var}[p]} - |r(\rho)|, \end{aligned} \tag{26}$$

where the third line above is by the assumption that p is $(\delta, \gamma+1)$-concentrated. Also, by Theorem 2.3, we have

$$\begin{aligned}\mathbf{Pr}_\rho\left[|r(\rho)| \geq \sqrt{\frac{\nu^{\gamma+1}}{8}} \cdot \|r\|_2\right] &\leq \mathbf{Pr}_\rho\left[|r(\rho)| \geq \sqrt{\frac{\nu}{8}} \cdot \|r\|_2\right] \\ &\leq \exp\left(-(1/4) \cdot \left(\frac{\nu}{8}\right)^{1/d}\right) \\ &\leq \delta/2. \end{aligned} \tag{27}$$

Combining Equation (26) and Equation (27), we get that, with probability at least $1 - \delta/2$,

$$\begin{aligned}|\mu'(\rho)| &\geq \sqrt{\frac{\nu^{\gamma+1}}{2} \cdot \mathbf{Var}[p]} - \sqrt{\frac{\nu^{\gamma+1}}{8}} \cdot \|r\|_2 \\ &= \sqrt{\frac{\nu^{\gamma+1}}{8}} \cdot \left(2 \cdot \sqrt{\mathbf{Var}[p]} - \|r\|_2\right) \\ &\geq \sqrt{\frac{\nu^{\gamma+1}}{8} \cdot \mathbf{Var}[p]} \\ &\geq \sqrt{\frac{\nu^{\gamma+1}}{8 \cdot d} \cdot \mathbf{Inf}[p]} && \text{(by Theorem 2.9)} \\ &\geq \sqrt{\frac{\nu^{\gamma+1}}{8 \cdot d \cdot 3^d} \cdot \|Q\|_2}, && \text{(by Equation (23))} \end{aligned}$$

as desired. □

Therefore, we have

$$\begin{aligned}&\mathbf{Pr}_\rho\left[Q(\rho) \geq \nu^{-\gamma} \cdot |\mu'(\rho)|^2\right] \\ &\leq \mathbf{Pr}_\rho\left[Q(\rho) \geq \frac{\nu^{-\gamma} \cdot \nu^{\gamma+1}}{8 \cdot d \cdot 3^d} \cdot \|Q\|_2\right] + \delta/2 && \text{(Claim 4.3)} \\ &= \mathbf{Pr}_\rho\left[Q(\rho) \geq \frac{\nu}{8 \cdot d \cdot 3^d} \cdot \|Q\|_2\right] + \delta/2 \\ &\leq \delta, && \text{(Theorem 2.3)} \end{aligned}$$

which completes the proof of Equation (24) and hence the lemma. □

Now by Lemma 4.2, the quantity in Equation (22) is at most δ. The second summand in Equation (21) is the product of two probabilities, the first of which is the same as for the original problem but with the restriction parameter m/b instead of m, and so can be analyzed inductively. By our arguments above, we get the recurrence:

$$\mathcal{P}(d, m, \delta) \leq \mathcal{P}(d, m/b, \delta) \cdot \mathcal{P}_{\mathrm{reg}}(d, b, \delta, 2, \tau) + 3\delta + \frac{1}{m} \cdot \tau^{-1} \cdot \left(d \cdot \log \tau^{-1} \cdot \log \delta^{-1}\right)^{O(d)}. \tag{28}$$

Therefore, to reduce from $\mathcal{P}(d, m, \delta)$ to $\mathcal{P}(d, m/b, \delta)$, it remains to bound

$$\mathcal{P}_{\mathrm{reg}}(d, b, \delta, 2, \tau).$$

However, by Lemma 3.4, we know that

$$\mathcal{P}_{\mathrm{reg}}(d, b, \delta, 2, \tau) \leq O\left(d^2\right) \cdot (c \cdot \log \delta^{-1})^{2d} \cdot \left(b^{-1/2} + \tau^{1/(8d)}\right). \tag{29}$$

4.2 Solving the recurrence

We are now ready to finish the proof of Lemma 4.1.

Proof of Lemma 4.1. Let $b = \lceil m^{1/(8d)} \rceil$ and $\tau = m^{-1/2}$. Note that $b^{-1/2} \leq m^{-1/(16d)}$. Then by Equation (28) and Equation (29) we get that

$$\begin{aligned}\mathcal{P}(d, m, \delta) \leq\ & m^{-1/2} \cdot \left(d \cdot \log m \cdot \log \delta^{-1}\right)^{O(d)} \\ & + 3\delta + O\left(d^2\right) \cdot \left(c \cdot \log \delta^{-1}\right)^{2d} \cdot m^{-1/(16d)} \cdot \mathcal{P}(d, m/b, \delta). \end{aligned} \tag{30}$$

We now show the following:

$$\mathcal{P}(d, m, \delta) \leq (m^{-1/2} + \delta) \cdot (d \cdot \log m \cdot \log \delta^{-1})^{16Ed^2 \log\log m}, \tag{31}$$

where E is a sufficiently large constant.

We proceed by induction on m. The base case is $m \leq 2^d$. In this case, the right hand side of Equation (31) is greater than 1 when E is sufficiently large and Equation (31) holds trivially. Now suppose Equation (31) holds for all smaller values of m. Let $M = d \cdot \log m \cdot \log \delta^{-1}$. By Equation (30), we obtain the recurrence

$$\mathcal{P}(d, m, \delta) \leq (m^{-1/2} + \delta) \cdot M^{E \cdot d} + m^{-1/(16d)} \cdot \mathcal{P}(d, m/b, \delta) \cdot M^{E \cdot d}, \tag{32}$$

for a sufficiently large constant E. Then by the induction hypothesis, we get

$$\begin{aligned} & m^{-1/(16d)} \cdot \mathcal{P}(d, m/b, \delta) \cdot M^{E \cdot d} \\ \leq\ & m^{-1/(16d)} \cdot (2 \cdot m^{-1/2+1/(16d)} + \delta) \cdot (d \cdot \log(m/b) \cdot \log \delta^{-1})^{16Ed^2 \log\log(m/b)} \cdot M^{E \cdot d} \\ \leq\ & 2 \cdot (m^{-1/2} + \delta) \cdot M^{16Ed^2 \log\log(m/b)} \cdot M^{E \cdot d}, \end{aligned} \tag{33}$$

where the first inequality above uses the fact that $(m/b)^{-1/2} \leq 2 \cdot m^{-1/2+1/(16d)}$ for $m > 2^d$. Combining Equation (32) and Equation (33), we have

$$\mathcal{P}(d, m, \delta) \leq (m^{-1/2} + \delta) \cdot 3 \cdot M^{16Ed^2 \log\log(m/b) + E \cdot d}.$$

Note that

$$\begin{aligned} \log\log(m/b) &\leq \log\log(m^{1-1/(8d)}) \\ &= \log(1 - 1/(8d)) + \log\log m \\ &\leq -1/(8d) + \log\log m. \end{aligned}$$

Therefore, when E is sufficiently large, we have

$$3 \cdot M^{16Ed^2 \log\log(m/b) + E \cdot d} \leq 3 \cdot M^{16Ed^2 \log\log(m) - E \cdot d} \leq M^{16Ed^2 \log\log(m)},$$

as required. □

5 Block Restriction Lemma with Optimal Exponent: Strong version

In this section, we prove a stronger version of Lemma 4.1. Note that for $d = 1$, the notation $O(d \cdot \log d)$ below should be interpreted as $O(1)$ (rather than 0).

Lemma 5.1 (Block Restriction Lemma: Strong Version)**.** *For any degree-d multilinear polynomial p, any $m \geq 16$, $\gamma \geq 1$, and $0 < \delta \leq 1/16$, we have*

$$\mathbf{Pr}_{\rho \sim B_m}\left[p_\rho \text{ is not } (\delta, \gamma)\text{-concentrated}\right] \leq (m^{-1/2} + \delta) \cdot (\log m)^{O(\gamma \cdot d \cdot \log d)} \cdot (\log \delta^{-1})^{O(\gamma \cdot d^2)}.$$

We follow a strategy similar to that in the proof of Lemma 4.1, except we do not bound the number of blocks where the polynomial p restricted to that block has its α function (see Definition 3.6) greater than some fixed threshold $\left(c \cdot \log \delta^{-1}\right)^{-d}$. Using such a rigid threshold for declaring a polynomial not δ-concentrated results in significant losses at each iteration of the recursion. To get an improved analysis, we instead keep track of an upper bound on the expected value of the function $\alpha(p)$, throughout the recursion. Such an upper bound provides a soft measure of the likelihood that the current function is still not δ-concentrated (cf. Corollary 3.7).

Thus, our proof of Lemma 5.1 will be as follows. We first argue (in Section 5.1) that it suffices to consider regular polynomials. Then we set up a recurrence (in Section 5.2), reducing the case of restrictions from B_m to that of restrictions from $B_{m/b}$, for a parameter $b > 0$. Finally, we solve the recurrence (in Section 5.3) to conclude the proof of Lemma 5.1.

As in the previous section, we only show for $\gamma = 1$ and note that the proof works for any γ.

5.1 Regularization

We shall modify our earlier definition of $\mathcal{P}$ and $\mathcal{P}_{\mathrm{reg}}$, using the function α from Definition 3.6.

Definition 5.2. *Let $\mathcal{P}(d, m, \delta, a)$ be the supremum, over all degree-d polynomials p with $\alpha(p) \leq a$ and all possible partitions of the variables into m blocks, of the probabilities*

$$\mathbf{Pr}_{\rho \sim B_m}\left[p_\rho \text{ is not } \delta\text{-concentrated}\right].$$

Let $\mathcal{P}_{\mathrm{reg}}(d, m, \delta, a, \tau)$ be the same as $\mathcal{P}$ but only for τ-regular polynomials.

We show that the analysis of $\mathcal{P}$ can be reduced to that of $\mathcal{P}_{\mathrm{reg}}$.

Lemma 5.3. *For any real $0 < \tau, \delta < 1/4$ and $a > 0$, integer $m > 4$ and $d, b \geq 1$, we have*

$$\mathcal{P}(d, m, \delta, a) \leq \frac{1}{m} \cdot \tau^{-1} \cdot (d \cdot \log \tau^{-1} \cdot \log \delta^{-1})^{O(d)} + 2\delta + \mathbf{Exp}_{\aleph}[\mathcal{P}_{\mathrm{reg}}(d, m, \delta, \aleph, \tau)],$$

where $\aleph$ is a non-negative random variable with $\mathbf{Exp}[\aleph] \leq O(a)$.

Proof. Let p be a degree-d multilinear polynomial with its variables petitioned into m blocks and $\alpha(p) \leq a$. Consider the decision tree given by Theorem 2.11 with $\epsilon = \delta$ and $\gamma = 2$. Note that the depth of this decision tree is

$$H = \tau^{-1} \cdot (d \cdot \log \tau^{-1} \cdot \log \delta^{-1})^{O(d)}.$$

We view ρ as $\rho = (\ell, \lambda)$, where ℓ is the selected block and λ is an uniform retrictoin to the variables outside block ℓ. For each leaf ω, let R_ω be the set of random restrictions consistent with the branch leading to ω. As observed above, the probability ξ of choosing a restriction from the complement of $\cup_\omega R_\omega$ is at most H/m. We get

$$\mathbf{Pr}_{\rho\sim B_m}[p_\rho \text{ is not } \delta\text{-concentrated}] \leq (1-\xi)\cdot \mathbf{Pr}_{\rho\in\cup_\omega R_\omega}[p_\rho \text{ is not } \delta\text{-concentrated}] + \xi. \tag{34}$$

By conditioning on $\rho \in R_\omega$, we get that $\mathbf{Pr}_{\rho\in\cup_\omega R_\omega}[p_\rho \text{ is not } \delta\text{-concentrated}]$ equals to

$$\sum_\omega \mathbf{Pr}_\rho[p_\rho \text{ is not } \delta\text{-concentrated} \mid \rho \in R_\omega]\cdot \mathbf{Pr}[\rho \in R_\omega \mid \rho \in \cup_\omega R_\omega].$$

Note that the probability of choosing $\rho \in R_\omega$ conditioned on $\rho \in \cup_\omega R_\omega$ is $2^{-\ell_\omega}\cdot(1-\xi)^{-1}$, where ℓ_ω is the length of the branch leading to ω. Hence, the right-hand side of Equation (34) is at most

$$\mathbf{Exp}_\omega[\mathbf{Pr}_{\rho\in R_\omega}[p_\rho \text{ is not } \delta\text{-concentrated}]] + \xi.$$

Each restriction $\rho = (\ell, \lambda) \in R_\omega$ can be viewed as a restriction of the variables on the branch leading to ω (as specified by the branch) plus an uniform restriction λ' to the remaining variables outside block ℓ. So we can express p_ρ as $(p_\omega)_{\rho'}$, where $\rho' = (\ell, \lambda')$.

Note that ρ' is a random block restriction on m blocks, which comes from the set of those restrictions that chose block ℓ outside at most H blocks containing the variables on the branch leading to ω. The set of all such restrictions ρ' has the probability mass at least $1 - H/m$ within the set of all random block restrictions ρ_ω (which pick block ℓ uniformly at random from the set of all m blocks). Therefore, we can upperbound the expression in Equation (34) by

$$\begin{aligned}
&\mathbf{Exp}_\omega[\mathbf{Pr}_{\rho'}[(p_\omega)_{\rho'} \text{ is not } \delta\text{-concentrated}]] + \xi \\
&\quad\leq (1-H/m)^{-1}\cdot \mathbf{Exp}_\omega[\mathbf{Pr}_{\rho_\omega}[(p_\omega)_{\rho_\omega} \text{ is not } \delta\text{-concentrated}]] + \xi \\
&\quad\leq \mathbf{Exp}_\omega[\mathbf{Pr}_{\rho_\omega}[(p_\omega)_{\rho_\omega} \text{ is not } \delta\text{-concentrated}]] + \xi + 2(H/m),
\end{aligned}$$

where the last inequality uses the fact that $(1-x)^{-1} \leq 1 + 2x$ whenever $0 < x \leq 1/2$.

Thus, we have

$$\mathbf{Pr}_{\rho\sim B_m}[p_\rho \text{ is not } \delta\text{-concentrated}] \leq \mathbf{Exp}_\omega\left[\mathbf{Pr}_{\rho_\omega\sim B_m}[(p_\omega)_{\rho_\omega} \text{ is not } \delta\text{-concentrated}]\right] + \frac{3H}{m}. \tag{35}$$

Note that a leaf ω can be in one of the three cases.

1. The polynomial restricted by ω is neither τ-regular nor $(\delta, 2)$-concentrated.
2. The polynomial restricted by ω is $(\delta, 2)$-concentrated.
3. The polynomial restricted by ω is not $(\delta, 2)$-concentrated but τ-regular.

Then the contribution from the ω's in case $i \in [3]$ to the expected value in Equation (35) is

$$\sum_{\omega \text{ in case } i} \mathbf{Pr}_{\rho_\omega\sim B_m}[(p_\omega)_{\rho_\omega} \text{ is not } \delta\text{-concentrated}]\cdot \mathbf{Pr}[\omega].$$

By Theorem 2.11, the contribution form those ω's in the first case at most $\epsilon = \delta$. If ω is in the second case, then by Lemma 4.2, the probability over ρ_ω that $(p_\omega)_{\rho_\omega}$ is not δ-concentrated is at

most δ, so those ω's contribute at most δ. Finally, the contribution from those ω's in the third case is at most

$$\mathbf{Exp}_\omega \left[\mathcal{P}_{\mathrm{reg}}(d, m, \delta, \alpha(p_\omega), \tau)\right].$$

Therefore, we have

$$\mathcal{P}(d, m, \delta, a) \leq \frac{1}{m} \cdot \tau^{-1} \cdot (d \cdot \log \tau^{-1} \cdot \log \delta^{-1})^{O(d)} + 2\delta + \mathbf{Exp}_\omega[\mathcal{P}_{\mathrm{reg}}(d, m, \delta, \alpha(p_\omega), \tau)].$$

To complete the proof, we need to show that

$$\mathbf{Exp}_\omega[\alpha(p_\omega)] \leq O\left(a\right).$$

We have

$$\begin{aligned}
\mathbf{Exp}_\omega[\alpha(p_\omega)] &= \mathbf{Exp}_\omega\left[\mathbf{Exp}_{A,B}\left[\min\left\{1, \frac{|\mathsf{D}_B\, p_\omega(A)|^2}{|p_\omega(A)|^2}\right\}\right]\right] \\
&\leq \mathbf{Exp}_\omega\left[\mathbf{Exp}_{A}\left[\min\left\{1, \frac{\|\nabla p_\omega(A)\|^2}{|p_\omega(A)|^2}\right\}\right]\right] && \text{(by Lemma 2.8)} \\
&\leq \mathbf{Exp}\left[\min\left\{1, \frac{\|\nabla p(A)\|^2}{|p(A)|^2}\right\}\right] \\
&\leq 72 \cdot \mathbf{Exp}\left[\min\left\{1, \frac{|\mathsf{D}_B\, p(A)|^2}{|p(A)|^2}\right\}\right] && \text{(by Lemma 2.8)} \\
&\leq 72 \cdot a,
\end{aligned}$$

as required. □

5.2 Setting up the recurrence

We show the following recurrence relation for regular polynomials.

Lemma 5.4. *For any real $0 < \tau, \delta < 1/4$ and $a > 0$, integer $m > 4$ and $d, b \geq 1$, we have*

$$\mathcal{P}_{\mathrm{reg}}(d, m, \delta, a, \tau) \leq \mathbf{Exp}_\aleph[\mathcal{P}(d, m/b, \delta, \aleph)],$$

where $\aleph$ is a non-negative random variable with $\mathbf{Exp}[\aleph] = O\left(d^3 a b^{-1/2} + d^4 \tau^{1/(8d)}\right)$.

We shall need the following analogue of the function $\alpha(p)$ for the Gaussian case.

Definition 5.5. *For p a non-zero polynomial, we define*

$$\beta(p) \colon = \mathbf{Exp}\left[\min\left\{1, \frac{|\mathsf{D}_Y\, p(X)|^2}{|p(X)|^2}\right\}\right].$$

The functions $\alpha(p)$ and $\beta(p)$ are related in the following way.

Lemma 5.6. *Let p be a degree-d, τ-regular, non-zero polynomial. Then*

$$\beta(p) = \alpha(p) + O\left(d \cdot \tau^{1/(8d)}\right).$$

Proof. We have

$$\begin{aligned}
\beta(p) &= \mathbf{Exp}\left[\min\left\{1, \frac{|\mathsf{D}_Y\, p(X)|^2}{|p(X)|^2}\right\}\right] \\
&= \int_0^1 \mathbf{Pr}\left[|p(X)| \le t^{-1/2} \cdot |\mathsf{D}_Y\, p(X)|\right] dt \\
&= \int_0^1 \mathbf{Pr}\left[|p(A)| \le t^{-1/2} \cdot |\mathsf{D}_B\, p(A)|\right] dt + O\left(d \cdot \tau^{1/(8d)}\right) \qquad \text{(by Theorem 2.12)} \\
&= \alpha(p) + O\left(d \cdot \tau^{1/(8d)}\right),
\end{aligned}$$

as required. □

In Lemma 5.4, we want to keep track of $\alpha(p)$ in the recurrence, and so we need a version of the anticoncentration bound that takes $\alpha(p)$ into account. This is achieved by the following version of Theorem 2.7 that takes $\beta(p)$ into account.

Theorem 5.7 ([Kan14])**.** *For any d-degree polynomial p and any $0 < \epsilon < 1$, we have*

$$\mathbf{Pr}\left[|p(X)| \le \epsilon \cdot |\mathsf{D}_Y\, p(X)|\right] = O\left(d^3 \beta(p) \epsilon\right).$$

We get the following.

Corollary 5.8. *For any d-degree τ-regular non-constant multilinear polynomial p, and any $\epsilon > 0$, we have*

$$\mathbf{Pr}[|p(A)| \le \epsilon \cdot |\mathsf{D}_B\, p(A)|] = O\left(d^3 \cdot \epsilon \cdot \beta(p) + d \cdot \tau^{1/(8d)}\right).$$

Proof. The proof is the same as that of Corollary 2.13, with Theorem 5.7 replacing Theorem 2.7. □

We are now ready to prove Lemma 5.4.

Proof of Lemma 5.4. Let p be a τ-regular, degree-d, multilinear polynomial with $\mathbf{Var}[p(x)] = 1$ and $\alpha(p) \le a$. Consider the way of choosing a random block restriction as described in Section 4.1. Recall that ρ_1 is a random block restriction on b blocks and ρ_2 is a random block restriction on m/b blocks. Then for any arbitrary partition of the variables into m blocks, we have

$$\begin{aligned}
\mathbf{Pr}_{\rho \sim B_m}[p_\rho \text{ is not } \delta\text{-concentrated}] &= \mathbf{Pr}_{\rho_1 \sim B_b, \rho_2 \sim B_{m/b}}[(p_{\rho_1})_{\rho_2} \text{ is not } \delta\text{-concentrated}] \\
&= \mathbf{Exp}_{\rho_1}\left[\mathbf{Pr}_{\rho_2}[(p_{\rho_1})_{\rho_2} \text{ is not } \delta\text{-concentrated}]\right] \\
&\le \mathbf{Exp}_{\rho_1}\left[\mathcal{P}(d, m/b, \delta, \alpha(p_{\rho_1}))\right].
\end{aligned}$$

Therefore, it suffices to show that

$$\mathbf{Exp}_{\rho_1 \sim B_b}[\alpha(p_{\rho_1})] = O\left(d^3 a b^{-1/2} + d^4 \tau^{1/(8d)}\right). \tag{36}$$

Note that

$$\mathbf{Exp}_{\rho_1 \sim B_b}[\alpha(p_{\rho_1})] = \frac{1}{b} \cdot \sum_{\ell} \mathbf{Exp}_{A_{\bar{\ell}}}[\alpha(p_{A_{\bar{\ell}}})],$$

where $A_{\bar{\ell}}$ is a random assignment to the variables that are not in block ℓ. From the calculation in Lemma 3.4 (Equation (20)), we have

$$\begin{aligned}
\sum_{\ell} \mathbf{Exp}_{A_{\bar{\ell}}}[\alpha(p_{A_{\bar{\ell}}})] &= \mathbf{Exp}\left[\min\left\{1, \frac{|\mathsf{D}_B\, p(A)|^2}{|p(A)|^2}\right\}\right] + \int_1^b \mathbf{Pr}\left[|p(A)| \le t^{-1/2} \cdot |\mathsf{D}_B\, p(A)|\right] dt \\
&= \alpha(p) + \int_1^b O\left(d^3 t \beta(p) t^{-1/2} + d\tau^{1/(8d)}\right) dt && \text{(by Corollary 5.8)} \\
&= \alpha(p) + O\left(d^3 \beta(p)\sqrt{b} + bd\tau^{1/(8d)}\right) \\
&= O\left(d^3 \alpha(p)\sqrt{b} + bd^4\tau^{1/(8d)}\right) && \text{(by Lemma 5.6)} \\
&= O\left(d^3 a\sqrt{b} + bd^4\tau^{1/(8d)}\right),
\end{aligned}$$

as required. □

5.3 Solving the recurrence

Since $\alpha(p) \le 1$ by definition, we have $\mathcal{P}(d, m, \delta) = \mathcal{P}(d, m, \delta, 1)$. Thus, to prove Lemma 5.1, it suffices to prove the following stronger result, and apply it with $a = 1$.

Theorem 5.9. *There is a constant $B > 0$ such that, for any $d > 0$, $m \ge 16$, $0 < \delta \le 1/16$ and $0 < a \le 1$, we have*

$$\mathcal{P}(d, m, \delta, a) \le (a \cdot m^{-1/2} + \delta) \cdot (\log m)^{B \cdot d \cdot \log d} \cdot (\log \delta^{-1})^{B \cdot d^2}. \tag{37}$$

Proof. First we argue that, for a sufficiently large constant $B > 0$, we may assume that a and m are relatively large.

Claim 5.10. *For a sufficiently large constant $B > 0$, we may assume that both*

$$a \ge (c \cdot \log \delta^{-1})^{-2d}, \tag{38}$$

and

$$m^{1/(32 \cdot d)} \ge (c \cdot \log \delta^{-1})^{2d}, \tag{39}$$

where $c > 0$ is the constant from Corollary 3.7.

Proof of Claim 5.10. If Equation (38) is false, then, by Corollary 3.7, the given polynomial is $(\delta, 2)$-concentrated, and by Lemma 4.2, the probability that its random block restriction is not δ-concentrated is at most δ, and so Equation (37) is satisfied. Next, assume Equation (38), and suppose that Equation (39) is false. Then we get that $a \cdot m^{-1/2} > (\log \delta^{-1})^{-T \cdot d^2}$, for some constant $T > 0$, implying that $a \cdot m^{-1/2} \cdot (\log \delta^{-1})^{B \cdot d^2} > 1$, for $B \ge T + 1$. Hence, the right-hand side of Equation (37) is greater than 1 in this case, and so Equation (37) holds. □

By Claim 5.10, we can assume for the rest of the proof that Equations (38) and (39) both hold.

Claim 5.11. *There is a constant $E > 0$ such that, for any $m \ge 16$ and $0 < \delta \le 1/16$, we have*

$$\mathcal{P}(d, m, \delta, a) \le \frac{1}{2} \cdot (a \cdot m^{-1/2} + \delta) \cdot (\log m)^{E \cdot d \cdot \log d} \cdot (\log \delta^{-1})^{E \cdot d^2} + \mathbf{Exp}_{\aleph}[\mathcal{P}(d, m/\lceil m^{1/(16d)} \rceil, \delta, \aleph)], \tag{40}$$

where $\aleph$ is a non-negative random variable with $\mathbf{Exp}[\aleph] = O\left(d^4 \cdot a \cdot m^{-1/(32d)}\right)$.

Proof. Let $\tau = m^{-1/2}$ and $b = \lceil m^{1/(16d)} \rceil$. By Lemma 5.3, we get that

$$\mathcal{P}(d, m, \delta, a) \le m^{-1/2} \cdot (d \cdot \log m \cdot \log \delta^{-1})^{O(d)} + 2\delta + \mathbf{Exp}_{\aleph_1}[\mathcal{P}_{\mathrm{reg}}(d, m, \delta, \aleph_1, m^{-1/3})], \tag{41}$$

for some non-negative random variable $\aleph_1$ with $\mathbf{Exp}[\aleph_1] \le O(a)$. By Equation (38), we get

$$m^{-1/2} \cdot (d \cdot \log m \cdot \log \delta^{-1})^{O(d)} + 2\delta \le \frac{1}{2} \cdot (a \cdot m^{-1/2} + \delta) \cdot (\log m)^{E \cdot d \cdot \log d} \cdot (\log \delta^{-1})^{E \cdot d^2}, \tag{42}$$

for a sufficiently large constant $E > 0$. Next, by Lemma 5.4, we get

$$\begin{aligned} \mathbf{Exp}_{\aleph_1}[\mathcal{P}_{\mathrm{reg}}(d, m, \delta, \aleph_1, m^{-1/2})] &\le \mathbf{Exp}_{\aleph_1}\left[\mathbf{Exp}_{\aleph_2}[\mathcal{P}(d, m/b, \delta, \aleph_2)]\right] \\ &\le \mathbf{Exp}_{\aleph}[\mathcal{P}(d, m/b, \delta, \aleph)], \end{aligned} \tag{43}$$

for some non-negative random variable $\aleph$ with

$$\begin{aligned} \mathbf{Exp}[\aleph] &= O\left(d^3 \cdot a \cdot b^{-1/2} + d^4 \cdot \tau^{1/(8d)}\right) \\ &\le O\left(d^3 \cdot a \cdot m^{-1/(32d)} + d^4 \cdot m^{-1/(16d)}\right) \\ &\le O\left(d^4 \cdot a \cdot m^{-1/(32d)}\right). && \text{(by Equations (38) and (39))} \end{aligned}$$

Equations (42) and (43) imply Equation (40). □

We now prove Theorem 5.9 by induction on m. We start with the base case $m \le 2^d$. By Equation (38), we only need to consider $a \ge (c \cdot \log \delta^{-1})^{-2d}$. Note that in this case, the bound in Theorem 5.9 is greater 1 when B is sufficiently large. Now suppose Theorem 5.9 holds for all smaller values of m. Let $M = (\log m)^{Bd \log(d)} \cdot (\log \delta^{-1})^{Bd^2}$ for $B > E$ to be determined, where E is the constant in Claim 5.11. By Claim 5.11, we have

$$\mathcal{P}(d, m, \delta, a) \le \frac{1}{2} \cdot (a \cdot m^{-1/2} + \delta) \cdot M + \mathbf{Exp}_{\aleph}[\mathcal{P}(d, m/\lceil m^{1/(16d)} \rceil, \delta, \aleph)],$$

where $\aleph$ is a non-negative random variable with $\mathbf{Exp}[\aleph] = C \cdot d^4 \cdot a \cdot m^{-1/(32d)}$ and C is some constant. Then by the induction hypothesis, we have

$$\begin{aligned} &\mathbf{Exp}_{\aleph}[\mathcal{P}(d, m/\lceil m^{1/(16d)} \rceil, \delta, \aleph)] \\ &\le \mathbf{Exp}_{\aleph}\left[(\aleph \cdot 2 \cdot m^{-1/2+1/(32d)} + \delta) \cdot (\log m^{1-1/(16d)})^{Bd \log(d)} \cdot (\log \delta^{-1})^{Bd^2}\right] \\ &\le 2 \cdot (\mathbf{Exp}_{\aleph}[\aleph] \cdot m^{-1/2+1/(32d)} + \delta) \cdot (1 - 1/(16d))^{Bd \log(d)} \cdot M \\ &\le 2 \cdot (C \cdot d^4 \cdot a \cdot m^{-1/(32d)} \cdot m^{-1/2+1/(32d)} + \delta) \cdot (1 - 1/(16d))^{Bd \log(d)} \cdot M \\ &= 2 \cdot C \cdot d^4 \cdot (1 - 1/(16d))^{Bd \log(d)} \cdot (a \cdot m^{-1/2} + \delta) \cdot M \\ &\le 2 \cdot C \cdot d^4 \cdot e^{-B \log(d)/16} \cdot (a \cdot m^{-1/2} + \delta) \cdot M \\ &\le \frac{1}{2} \cdot (a \cdot m^{-1/2} + \delta) \cdot M, \end{aligned}$$

where the last inequality holds if B is sufficiently large. □

Note that the only reason why we have the factor $(\log \delta^{-1})^{O(d^2)}$ rather than $(\log \delta^{-1})^{O(d)}$ in Equation (37) is to justify the assumption in Equation (39). If we assume this condition explicitly, then we get the following slightly stronger version of the PTF restriction lemma for δ not too small.

Lemma 5.12. *For any degree-d multilinear p, any $m \geq 16$, and any $0 < \delta < 1/16$ such that $m^{1/(64 \cdot d)} \geq (c \cdot \log \delta^{-1})^d$ for the constant $c > 0$ from Corollary 3.7, we have*

$$\mathbf{Pr}_{\rho \sim B_m}[p_\rho \text{ is not } \delta\text{-concentrated}] \leq (m^{-1/2} + \delta) \cdot (\log m)^{O(d \log d)} \cdot (\log \delta^{-1})^{O(d)}.$$

6 Applications

6.1 Lower bounds for depth-2 circuits with PTF gates

Here we generalize the gate and wire complexity lower bound of [KW16] for Andreev's function against depth-2 circuits with LTF gates, to depth-2 circuits with degree-d PTF gates, for any $d \geq 1$. Our lower bounds match those of [KW16] for the case of $d = 1$ (up to polylogarithmic factors), and extend to any degree $d \ll \sqrt{(\log n)/(\log \log n)}$.

The main technical tool used by [KW16] was a restriction lemma saying, roughly, that an n-variate LTF function hit by some "structured" random restriction that leaves $(\log n)$ variables, will become a constant function except with probability $(\log n)/\sqrt{n}$. This restriction lemma is then combined with a careful counting argument to show that Andreev's function requires depth-2 LTF circuits with at least $\Omega(n^{1.5}/(\log^3 n))$ gates, and $\Omega(n^{2.5}/(\log^{7/2} n))$ wires.

The restriction lemma of [KW16] is proved using the Littlewood-Offord lemma from additive combinatorics [LO43]. It is not clear how to prove a similar restriction lemma for higher degree $d > 1$ using the same tools. However, we show that our Block Restriction Lemma yields such a generalization for any $1 \leq d \ll \sqrt{(\log n)/(\log \log n)}$. The reason is that we can make the parameter δ in Lemma 4.1 very small compared to the number of unrestricted variables so that, for the restricted function being δ-close to constant is the same as being constant. We start by proving the following restriction lemma for PTFs.

Lemma 6.1. *Let f be any n-variate degree-d PTF. Let $\mathscr{P}$ be a partition of $[n]$ into parts of equal-sized with $|\mathscr{P}| \leq n/16$, and let $\mathscr{R}_{\mathscr{P}}$ be the distribution on restrictions $\rho : [n] \to \{-1, 1, *\}$ that randomly selects one variables from each part of $\mathscr{P}$ and restricts all other variables uniformly at random. Then*

$$\mathbf{Pr}_{\rho \sim \mathscr{R}_{\mathscr{P}}}[f_\rho \text{ is not a constant}] \leq \frac{1}{\sqrt{n}} \cdot (|\mathscr{P}| \cdot \log n)^{O(d^2)}. \tag{44}$$

Moreover, if f depends on at most w of its inputs, then

$$\mathbf{Pr}_{\rho \sim \mathscr{R}_{\mathscr{P}}}[f_\rho \text{ is not a univariate function}] \leq \frac{1}{n^{3/2}} \cdot w \cdot (|\mathscr{P}| \cdot \log n)^{O(d^2)}. \tag{45}$$

Proof. Consider the following equivalent way of choosing a random restriction $\rho \sim \mathscr{R}_{\mathscr{P}}$.

1. Create $m = \frac{n}{|\mathscr{P}|}$ blocks. For each part in $\mathscr{P}$, randomly assign the $\frac{n}{|\mathscr{P}|}$ variables in the part to the m blocks so that each block takes exactly one of the variables from the part.

2. Apply a random block restriction $\rho' \sim B_m$ based on the partition in the previous step.

By Lemma 5.1 and Lemma 2.16, for any partition into blocks generated in the first step above, the probability over the restrictions in the second step that the restricted PTF is not δ-close to constant is at most

$$\left(\sqrt{\frac{|\mathscr{P}|}{n}}+\delta\right)\cdot\left(\log\frac{n}{|\mathscr{P}|}\cdot\log\delta^{-1}\right)^{O(d^2)}. \tag{46}$$

Now let $\delta=\min\left\{2^{-(|\mathscr{P}|+1)},\sqrt{\frac{|\mathscr{P}|}{n}}\right\}$. In this case, the restricted function, which is on $|\mathscr{P}|$ variables and δ-close to constant, is indeed a constant. Note that for such δ, Equation (46) implies Equation (44).

Next, for each wire $i\in[w]$, define the following random event E_i: the function f_ρ depends on wire i and on some other wire. Note that if E_i happens, then wire i is assigned $*$ by ρ which happens with probability $|\mathscr{P}|/n$, and that both $(f_\rho)_{w_i=-1}$ and $(f_\rho)_{w_i=1}$ are non-constant functions. It is not hard to see that given wire i is assigned $*$, the probability that $(f_\rho)_{w_i=-1}$ (or $(f_\rho)_{w_i=1}$) is non-constant is

$$\mathbf{Exp}_{\rho_1}[\mathbf{Pr}_{\rho_2}[((f_{\rho_1})_{w=-1})_{\rho_2}\text{ is not a constant}], \tag{47}$$

where ρ_1 is a random partial assignment to the wires (except wire i) in the part that contains wire i, and ρ_2 is the restriction that randomly selects one variables from each of the rest $|\mathscr{P}|-1$ parts and restricts all other variables uniformly at random. Then the inner probability in Equation (47) can be upperbounded by Equation (44). Thus, we get

$$\begin{aligned}
&\mathbf{Pr}_{\rho\sim\mathscr{R}_{\mathscr{P}}}[f_\rho\text{ is not a univariate function}]\\
&=\mathbf{Pr}\left[\vee_{i=1}^{w}E_i\right]\\
&\le\sum_{i=1}^{w}\mathbf{Pr}[E_i]\\
&\le w\cdot\frac{|\mathscr{P}|}{n}\cdot 2\cdot\frac{1}{\sqrt{n}}\cdot((|\mathscr{P}|-1)\cdot\log n)^{O(d^2)},
\end{aligned}$$

implying Equation (45). □

To simplify the presentation, we only argue the worst-case lower bound; a correlation bound as in [KW16] can also be proved in a similar way. We prove a lower bound for the Andreev's function.

Definition 6.2. *Define Andreev's function $A_n\colon\{-1,1\}^{5n}\to\{-1,1\}$ as follows:*

$$A_n(x_1,\ldots,x_{4n},y_1,\ldots,y_n)=x_i,$$

where $i\in[4n]$ is a positive integer uniquely given by the binary string $z\in\{-1,1\}^{\log 4n}$ obtained as follows: partition $[n]$ into $\log 4n$ parts so that each part has $t=\frac{n}{\log 4n}$ variables, and the j-th part P_j is the set $\{y_{(j-1)\cdot t+k}\colon k=1,\ldots,t\}$. Then $z_j=\prod_{y_k\in P_j}y_k$.

For simplicity, we assume here that n is divisible by $\log 4n$. We show the following gate and wire lower bounds for A_n against depth-2 circuits with d-degree PTF gates; for $d=1$, these bounds match those of [KW16], up to polylogarithmic factors.

Theorem 6.3 (Lower bounds for depth-2 degree-d PTF circuits)**.** *Every depth-2 circuit on n inputs and degree-d PTF gates, that computes A_n must have at least $\left(n^{\frac{1}{2}+\frac{1}{d}}\right)\cdot(\log n)^{-O(d^2)}$ gates, and at least $\left(n^{\frac{3}{2}+\frac{1}{d}}\right)\cdot(\log n)^{-O(d^2)}$ wires.*

For the proof of Theorem 6.3, we shall need the following straightforward generalization to degree-d PTFs of the result in [RSO94] about the number of LTFs on s inputs, where each input is some boolean function of n variables; the latter result is in turn a generalization of [Win61, Cho61].

Theorem 6.4 ([RSO94])**.** *For any degree-d PTF g on s variables, and any collection of boolean functions $f_1, \ldots, f_s \colon \{-1,1\}^n \to \{-1,1\}$, the n-variate boolean function*

$$h(\vec{x}) = g(f_1(\vec{x}), \ldots, f_s(\vec{x}))$$

where $\vec{x} = (x_1, \ldots, x_n) \in \{-1,1\}^n$, can be completely specified using $O(s^d \cdot n)$ bits.

As an immediate consequence of Theorem 6.4, we get the following.

Corollary 6.5. *Every depth-2 circuit on n inputs with s degree-d PTF gates can be completely specified using at most $O\left(s^d \cdot n + s \cdot n^{d+1}\right)$ bits.*

Now we are ready to complete the proof of Theorem 6.3.

Proof of Theorem 6.3. For an arbitrary $\vec{a} = (a_1, \ldots, a_{4n}) \in \{-1,1\}^{4n}$, let

$$F(y_1, \ldots, y_n) = A_n(a_1, \ldots, a_{4n}, y_1, \ldots, y_n).$$

Towards contradiction, suppose that A_n, and hence also F, is computable by a depth-2 circuit with degree-d PTF gates of wire or gate complexity less than the bounds claimed in the theorem statement (for sufficiently large constants in the $O(d^2)$ exponents of the polylog factors). Let $\mathscr{P}$ be the partition of $[n]$ into $\log 4n$ parts of equal size as specified in Definition 6.2. We then apply a random restriction $\rho \sim \mathscr{R}_{\mathscr{P}}$ to the function $F(y_1, \ldots, y_n)$. Then F_ρ can be used to reconstruct, in the information-theoretic sense (say, in the sense of Kolmogorov complexity) the string $\vec{a}$ (by the definition of A_n). More precisely, to reconstruct $\vec{a}$, it suffices to know the restriction ρ plus the description of some circuit computing F_ρ. The restriction ρ can be described using at most $2n$ bits (by specifying for each $i \in [n]$ whether it is 1, -1, or unrestricted). Next we bound the size of a circuit computing F_ρ, for some ρ satisfying the above condition.

By Lemma 6.1, the expected number of bottom PTF gates of the depth-2 circuit computing $F_\rho(y_1, \ldots, y_n)$ is at most $s_0 = n^{1/d}/(\log n)^{O(d^2)}$ (if either the number of gates or the number of wires of F is small). By the Markov inequality, the probability over $\rho \sim \mathscr{R}_{\mathscr{P}}$ that the actual number s of gates of the circuit for F_ρ is more than $2 \cdot s_0$ is at most $1/2$.

It follows that with probability at least $1/2$, we get a random restriction $\rho \sim \mathscr{R}_{\mathscr{P}}$ such that F_ρ has at most $2 \cdot s_0$ gates. By Corollary 6.5, the circuit for F_ρ is described with at most n bits.

We conclude that every $\vec{a} \in \{-1,1\}^{4n}$ can be described with at most $2n + n = 3n$ bits. However, by a simple counting argument, we know that almost all $4n$-bit strings $\vec{a}$ require the description size strictly greater than $3n$. A contradiction. □

6.2 Lower bounds for depth-3 circuits with PTF gates

Here we generalize the lower bound of [KW16] against circuits that are majority votes of depth-2 LTF circuits, to majority votes of depth-2 circuits with degree-d PTF gates. In [KW16], it was shown that there exists a polynomial time function that requires circuits of the form mentioned above with $\Omega\left(n^{2.5}/(\log^{7/2} n)\right)$ wires. Here, we show a lower bound against circuits that can have sub-exponential size as long as the total fanin of the bottom layer gates is small.

We first define a generalized Andreev function. Recall that a (ζ, L)-list-decodable code is a function $K\colon \{-1,1\}^k \to \{-1,1\}^n$ that maps k-bits messages to n-bits codewords such that for any codeword $y \in \{-1,1\}^n$, there are at most L codewords in the range of K that have relative hamming distance within ζ from y. We will use the following list-decodable code (see, e.g., [CKK+15] for its construction).

Theorem 6.6. *For any given $0 < \epsilon < 1$, there exists a binary code K mapping $4n$-bit message to a codeword of length 2^{n^ϵ}, such that K is (ζ, L)-list-decodable for $\zeta = 1/2 - O\left(2^{-n^\epsilon/4}\right)$ and $L \leq O\left(2^{n^\epsilon/2}\right)$. Furthermore, there is a polynomial-time algorithm for computing $K(x)$ in position z, for any inputs $x \in \{-1,1\}^{4n}$ and $z \in \{-1,1\}^{n^\epsilon}$.*

Definition 6.7. *Let $0 < \epsilon < 0$. Define the function $B_{n,\epsilon}\colon \{-1,1\}^{5n} \to \{-1,1\}$ as follows:*

$$B_{n,\epsilon}(x_1, \ldots, x_{4n}, y_1, \ldots, y_n) = K(x)_i,$$

where K is the code from Theorem 6.6, and $i \in [2^{n^\epsilon}]$ is a positive integer uniquely given by the binary string $z \in \{-1,1\}^{n^\epsilon}$ obtained as follows: partition $[n]$ into n^{ϵ/d^2} parts so that each part has $t = \frac{n}{n^\epsilon}$ variables, and the j-th part P_j is the set $\left\{y_{(j-1)\cdot t+k}\colon k = 1, \ldots, t\right\}$. Then $z_j = \prod_{y_k \in P_j} y_k$.

Note that the function above is polynomial-time computable since we can compute $K(x)$ in position i in polynomial time.

We are now ready to prove the lower bound.

Theorem 6.8. *For any $\frac{1}{\log n} \ll \epsilon < 1$, let C be a majority vote of depth-2 circuits with degree-d PTF gates such that the top majority gate has fanin at most 2^{n^ϵ} and the total fanin of the gates on the bottom layer at most $w = \left(n^{\frac{3}{2}+\frac{1}{d}}\right) \cdot \left(n^\epsilon \cdot \log n\right)^{-c\cdot d^2}$, where c is a constant. Then C cannot compute $B_{n,\epsilon}$.*

Proof. Let $a \in \{-1,1\}^{4n}$ be a string with Kolmogorov complexity at least $4n$, and let

$$F(y_1, \ldots, y_n) = B_{n,\epsilon}(a_1, \ldots, a_{4n}, y_1, \ldots, y_n).$$

Let D be an arbitrary depth-2 circuit in n^ϵ variables with degree-d PTF gates, of size at most $s_0 = n^{1/d - O\left(\epsilon \cdot d^2\right)}$. Note that by Corollary 6.5, D can be described with at most n bits. Let $\mathscr{P}$ be the partition of $[n]$ into n^ϵ parts of equal size as specified in Definition 6.7. We claim that for any $\rho \sim \mathscr{R}_{\mathscr{P}}$,

$$\mathbf{Corr}(F_\rho, D) < 2^{-n^\epsilon}.$$

Toward a contradiction, suppose D agrees with $F\rho$ on at least $1/2 + 2^{-n^\epsilon}$ of the inputs for some ρ. Then we can recover a as follows. We first use the circuit D and the string ρ to compute the corrupted codeword K' such that K' and $K(a)$ have relative hamming distance at most $1/2 - 2^{-n^\epsilon}$. We then list-decode K' to obtain a list of $L \leq O\left(2^{n^\epsilon/2}\right)$ codewords, which must contain $K(a)$. Finally, we use an index string of length at most $\log(L)$ to get $K(a)$ from the list of codewords and recover a. This shows that we can use fewer than $4n$ bits to describe a, which contradicts the assumption that a has Kolmogorov complexity at least $4n$.

Next, let C_a be the circuit obtained from C by setting the first $4n$ variables to be a, and let $\rho_0 \sim \mathscr{R}_{\mathscr{P}}$ be a restriction such that $(C_a)_{\rho_0}$ has at most s_0 gates on the bottom layer. The existence of such a restriction is guaranteed by Equation (45), when the total fanin of the bottom layer gates

is at most w and c is sufficiently large. By the nature of majority function and a simple averaging argument, we know that $(C_a)_{\rho_0}$ must have correlation at least 2^{-n^ϵ} with one of its sub-circuits, which is a depth-2 circuit in n^ϵ variables with degree-d PTF gates, of size at most s_0. Thus, we conclude that C_a cannot compute F. □

6.3 Lower bounds for constant-depth circuits with PTF gates

Here we extend the wire complexity correlation bounds of [CSS16] for parity and the generalized Andreev's function against constant-depth circuits with LTF gates to constant-depth circuits with degree-d PTF gates, for any $d \geq 1$. We do this by generalizing the structural lemma for LTFs used in [CSS16] to degree-d PTFs.

Lemma 6.9. *For any PTF $f(x) = \mathrm{sgn}(p(x))$ of degree $d \geq 1$, and any $0 < \delta,\ r \leq 1/16$, we have*

$$\mathbf{Pr}_{\rho \sim R_r}\left[f_\rho \text{ is not } \delta\text{-close to constant}\right] \leq (\sqrt{r} + \delta) \cdot (\log r^{-1} \cdot \log \delta^{-1})^{O(d^2)}.$$

Proof. Let r_0 be so that $r_0^{-1} = \lfloor r^{-1} \rfloor$. Then we have

$$\mathbf{Pr}_{\rho \sim R_r}\left[f_\rho \text{ is not } \delta\text{-close to constant}\right] = \mathbf{Pr}_{\rho_1 \sim R_{r_0}, \rho_2 \sim R_{r/r_0}}\left[(f_{\rho_1})_{\rho_2} \text{ is not } \delta\text{-close to constant}\right]. \tag{48}$$

Let $E(\rho_1)$ denote the random event that f_{ρ_1} is δ^2-close to constant. Then Equation (48) can be expressed as

$$\begin{aligned} &\mathbf{Pr}_{\rho_1,\rho_2}[(f_{\rho_1})_{\rho_2} \text{ is not } \delta\text{-close to constant} \mid \neg E(\rho_1)] \cdot \mathbf{Pr}_{\rho_1}[\neg E(\rho_1)] \\ &+ \mathbf{Pr}_{\rho_1,\rho_2}[(f_{\rho_1})_{\rho_2} \text{ is not } \delta\text{-close to constant} \mid E(\rho_1)] \cdot \mathbf{Pr}_{\rho_1}[E(\rho_1)] \end{aligned} \tag{49}$$

By the fact that a function δ^2-close to constant is expected to remain δ^2-close to constant under random restrictions and by Markov's inequality, the second summand in Equation (49) is at most δ. We then upperbound the first summand in Equation (49) by showing the following

$$\mathbf{Pr}_{\rho_1}[\neg E(\rho_1)] \leq (\sqrt{r} + \delta) \cdot (\log r^{-1} \cdot \log \delta^{-1})^{O(d^2)}.$$

Since $r_0 \leq 2r$, it suffices to show

$$\mathbf{Pr}_{\rho_1 \sim R_{r_0}}\left[f_{\rho_1} \text{ is not } \delta^2\text{-close to constant}\right] \leq (\sqrt{r_0} + \delta) \cdot (\log r_0^{-1} \cdot \log \delta^{-1})^{O(d^2)}. \tag{50}$$

Equation (50) follows immediately from Lemma 5.1 and Lemma 2.16 by noting the following equivalent way of choosing a random restriction $\rho_1 \sim R_{r_0}$.

1. Randomly partition the variables of f into $m = 1/r_0$ disjoint blocks, where each variable is assigned to block $i \in [m]$, independently, with probability $1/m$.
2. Apply a random block restriction $\rho' \sim B_m$ based on the partition in the previous step.

□

We now state our correlation bounds against constant-depth circuits with PTF gates. Let Par_n denote the parity function on n variables, and let $A'_n \in \mathsf{P}$ denote the variant of Andreev's function

on $5n$ variables as defined in [CKK+15].[3] For boolean functions $f, g\colon \{-1,1\}^n \to \{-1,1\}$, recall that the correlation between f and g is

$$\mathbf{Corr}(f,g) = \left|\mathbf{Exp}_{x\sim\{-1,1\}^n}\left[f(x)\cdot g(x)\right]\right|.$$

We get the following correlation bounds.

Theorem 6.10. *For any $D \geq 1$ and $1 \leq d \ll \sqrt{\log n/\log\log n}$, let C be any depth-D circuit on n inputs with degree-d PTF gates, of wire complexity at most $n^{1+\varepsilon_D}$, where $\varepsilon_D = B^{-(2D-1)}$, for some constant $B > 0$. Then we have*

$$\mathbf{Corr}(C, \mathrm{Par}_n) \leq O\left(n^{-\varepsilon_D}\right).$$

Theorem 6.11. *For any $D \geq 1$ and $1 \leq d \ll (\log n/\log\log n)^{1/(2D-1)}$, let C be any depth-D circuit on $5n$ inputs with degree-d PTF gates, of wire complexity at most $n^{1+\mu_{D,d}}$, where $\mu_{D,d} = (E\cdot d)^{-(2D-1)}$, for some constant $E > 0$. Then we have*

$$\mathbf{Corr}(C, A'_n) \leq \exp(-n^{\mu_{D,d}/2}).$$

Remark 6.12. *In Theorem 6.10, the exponent ε_D in the correlation bound does not depend on the degree d of the PTF gates in the circuit C, and stays polynomially small even for super-constant degree $d \ll \sqrt{\log n/\log\log n}$. In Theorem 6.11, the correlation bound is exponentially small for a constant degree d, and is super-polynomially small for $d \ll (\log n/\log\log n)^{1/(2D-1)}$.*

The proofs of Theorem 6.10 and Theorem 6.11 are analogous to those in [CSS16] for the case of LTF circuits, with just a couple of changes. The proofs are by induction on the depth D. For the proof of correlation bounds with parity in [CSS16], the base case is the noise sensitivity bound for LTFs due to Peres [Per04]; for the proof of Theorem 6.10, we can use the noise sensitivity bound for degree-d PTFs due to Kane [Kan14]. For the correlation bounds with Andreev's function, the base case in [CSS16] needs an upper bound on the number of distinct LTFs on n variables; we can use the bound for PTFs given by Theorem 6.4. Finally, for the inductive step, [CSS16] use their LTF restriction lemma to show that, under a particular type of random restriction, with high probability, a depth-D circuit with LTF gates will become close to some circuit of depth $D-1$. We can use our PTF Restriction Lemma, Lemma 6.9, with an appropriately small value of δ (for example, $\exp\left(-(r^{-1/(c\cdot d^2)})\right)$, for a sufficiently large constant c).

6.4 Influence bound for PTFs

Here we show that Kane's bound on the total influence (average sensitivity) of degree-d PTFs is a corollary of our Block Restriction Lemma.

Theorem 6.13 ([Kan14])**.** *For any d-degree PTF f on $n > 1$ variables, we have*

$$\mathbf{Inf}[f] \leq \sqrt{n}\cdot(\log n)^{O(d\log d)}\cdot 2^{O\left(d^2\log d\right)}.$$

[3] We have $A'_n(x_1,\ldots,x_{4n},y_1,\ldots,y_n) = \mathrm{Enc}(x_1,\ldots,x_{4n})_{\mathrm{Ext}(y_1,\ldots,y_n)}$, where $\mathrm{Enc}(\cdot)$ denotes the encoding with a certain error-correcting code, and $\mathrm{Ext}(\cdot)$ is a certain extractor; see [CKK+15] or [CSS16] for more details.

Proof. We first partition the variables into n blocks so that each block contains exactly one variable. We then apply a variant of our Block Restriction Lemma, Lemma 5.12, with $\delta = 1/n^2$. For

$$d \leq \sqrt{(\log n)/(c' \cdot \log\log n)}, \tag{51}$$

for some constant $c' > 0$, the assumption of Lemma 5.12 on the largeness of δ is satisfied. Note that since the restricted function is on one variable, being $(1/n^2)$-close to constant is the same as being constant. Therefore, by Lemma 5.12 and Lemma 2.16, we get

$$\mathbf{Pr}_{\rho\sim B_n}\left[f_\rho \text{ is not a constant}\right] \leq n^{-1/2} \cdot (\log n)^{O(d \log d)}. \tag{52}$$

Also, by the definition of random block restriction, we have

$$\mathbf{Pr}_{\rho\sim B_n}\left[f_\rho \text{ is not a constant}\right] = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathbf{Pr}_{A_{\bar{i}}}\left[f_{A_{\bar{i}}} \text{ is not a constant}\right], \tag{53}$$

where $A_{\bar{i}}$ is a random assignment to the variables except the i-th variable. Note that for every fixed i,

$$\mathbf{Pr}_{A_{\bar{i}}}\left[f_{A_{\bar{i}}} \text{ is not a constant}\right] = \mathbf{Pr}_{x\sim\{-1,1\}^n}[f(x) \neq f(x^{\oplus i})] = \mathbf{Inf}_i[f]. \tag{54}$$

Combining Equation (53) and Equation (54), we have

$$\sum_{i=1}^{n} \mathbf{Inf}_i[f] = n \cdot \mathbf{Pr}_{\rho\sim B_n}\left[f_\rho \text{ is not a constant}\right].$$

Together with Equation (52), we get

$$\mathbf{Inf}\left[f\right] \leq \sqrt{n} \cdot (\log n)^{O(d \log d)}. \tag{55}$$

Note that Equation (55) holds for small degrees d satisfying Equation (51). If we multiply the right-hand side of Equation (55) by $2^{O(d^2 \log d)}$, we ensure that the bound on influence holds also for all large d (as then the right-hand side of Equation (55) becomes at least n, which is a trivial upper bound on $\mathbf{Inf}[f]$). □

6.5 Littlewood-Offord type anticoncentration bounds for polynomials

Here we use our Block Restriction Lemma to drive the following anticoncentration bounds for degree-d multilinear polynomials.

Theorem 6.14 ([MNV16])**.** *For any real interval I, and any degree-d multilinear polynomial p such that there exists a set of t disjoint monomials in p, each of which is maximal (i.e., not contained by any other monomials) and has coefficient at least $|I|$ in magnitude, we have*

$$\mathbf{Pr}[p(A) \in I] \leq t^{-1/2} \cdot (\log t)^{O(d \log d)} \cdot 2^{O(d^2 \log d)}.$$

Proof. Our proof is very similar to that of [MNV16], except they used Kane's bound of Theorem 6.13, whereas we use a variant of our Block Restriction Lemma (Lemma 5.12). Without loss of generality, we can assume I is centered at 0; otherwise, the center of I is c and we can bound

the probability that the polynomial $p' = p - c$ takes values within the interval I' centered at 0 with $|I'| = |I|$.

We first partition the variables into t blocks so that p restricted to each block (i.e., the restricted polynomial that only depends on the variables in that block) has at least one monomial with the coefficient at least $|I|$ in magnitude. Consider the following equivalent way of sampling a uniformly random input to p: apply a block restriction based on the partition above and randomly assign 1 or -1 to the variables in the unrestricted block. Then we have that

$$\begin{aligned}\mathbf{Pr}[p(A) \in I] &= \mathbf{Pr}_{\rho\sim B_t, C}[p_\rho(C) \in I] \\ &= \mathbf{Pr}_{\rho,C}[p_\rho(C) \in I \mid p_\rho \text{ is not } \delta\text{-concentrated}] \cdot \mathbf{Pr}_\rho[p_\rho \text{ is not } \delta\text{-concentrated}] \\ &\quad + \mathbf{Pr}_{\rho,C}[p_\rho(C) \in I \mid p_\rho \text{ is } \delta\text{-concentrated}] \cdot \mathbf{Pr}_\rho[p_\rho \text{ is } \delta\text{-concentrated}], \end{aligned} \tag{56}$$

where C is a multidimensional Bernoulli random variable.

Let $\delta = t^{-1/2}$. By Lemma 5.12, we have

$$\mathbf{Pr}_{\rho\sim B_t}[p_\rho \text{ is not } \delta\text{-concentrated}] \le t^{-1/2} \cdot (\log t)^{O(d \log d)} \cdot 2^{O\left(d^2 \log d\right)}. \tag{57}$$

Note that we multiply by the factor $2^{O\left(d^2 \log d\right)}$ on the right-hand side of Equation (57) so that it holds for all degrees. This bounds the first summand of Equation (56). To bound the second summand of Equation (56), we use the following observation from a preliminary version of [MNV16].

Claim 6.15. *For any real interval I centered at 0, and any δ-concentrated degree-d multilinear polynomial q that has at least one monomial with coefficient greater than $|I|$ in magnitude, we have*

$$\mathbf{Pr}[q(A) \in I] \le \delta.$$

Proof. Let $q = q' + \mu$ where $\mu = \mathbf{Exp}[q(A)]$, and let $\nu = (L \cdot \log \delta^{-1})^d$ where $L > 0$ is the constant from Definition 2.15. Since q is δ-concentrated and has at least one monomial with coefficient greater than $|I|$ in magnitude, we have

$$|\mu|^2 \ge (\nu - 1) \cdot \mathbf{Var}[q] \ge (\nu - 1) \cdot |I|^2 \ge 4 \cdot |I|^2.$$

Now since $|\mu| \ge 2 \cdot |I|$, we note that for all points $x \in \{-1, 1\}^n$ where $q(x) \in I$, it must be the case that $|q'(x)| \ge |\mu| - |I|$. Also, we have

$$|\mu| - |I| \ge \frac{|\mu|}{2} \ge \frac{\sqrt{(\nu - 1) \cdot \mathbf{Var}[q]}}{2} \ge \frac{\sqrt{\nu \cdot \mathbf{Var}[q]}}{4} = \frac{\sqrt{\nu}}{4} \cdot \left\|q'\right\|_2. \tag{58}$$

As a result,

$$\begin{aligned}\mathbf{Pr}[q(A) \in I] &\le \mathbf{Pr}\left[|q'(A)| \ge |\mu| - |I|\right] \\ &\le \mathbf{Pr}\left[|q'(A)| \ge \frac{\sqrt{\nu}}{4} \cdot \left\|q'\right\|_2\right] && \text{(by Equation (58))} \\ &\le \delta. && \text{(by Theorem 2.3)}\end{aligned}$$

□

By Claim 6.15, we get

$$\mathbf{Pr}_{\rho,C}[p_\rho(C) \in I \mid p_\rho \text{ is } \delta\text{-concentrated}] \le \delta,$$

which bounds the second summand of Equation (56). This completes the proof. □

7 Derandomization

7.1 Derandomized Block Restriction Lemma

In this subsection, we show how to derandomize our Block Restriction Lemma, by giving an algorithm for sampling pseudorandom block restrictions (using significantly fewer random bits) so that the probability a given degree-d polynomial is not concentrated under such a pseudorandom block restriction is about the same as that for true random block restrictions. Our pseudorandom block restriction will pick a uniformly random block, and then fix the variables in the remaining blocks in a pseudorandom fashion (using few truly random bits).

Theorem 7.1 (Derandomized Block Restriction Lemma)**.** *For any $0 < \delta \leq 1/16$ and $0 < \zeta < 1$, there is a polynomial-time algorithm for sampling block restrictions $\rho \in B_m$, for any $m \geq 16$, that uses at most $m^{\zeta} \cdot \log n$ random bits, so that the following holds. For any n-variate degree-d multilinear polynomial p whose variables are partitioned into m blocks, we have*

$$\mathbf{Pr}_{\rho}\left[p_{\rho} \text{ is not } \delta\text{-concentrated}\right] \leq \left(m^{-1/2} + \delta\right) \cdot (\log m)^{O(\zeta^{-1} \cdot d \cdot \log d)} \cdot (\log \delta^{-1})^{O(\zeta^{-1} \cdot d^2)}.$$

We first define our pseudorandom block restrictions that yields Theorem 7.1. We start with some notations. Let D be a distribution on $\{-1,1\}^n$. Let S be a set of K coordinates and let ω be an assignment for the coordinates in S. We define D^{ω} to be the distribution on the remaining $n-K$ unfixed coordinates such that for any $a \in \{-1,1\}^{n-K}$,

$$\mathbf{Pr}[D^{\omega} = a] = \mathbf{Pr}[D_{[n]-S} = a \mid D_S = \omega].$$

We will refer to D^{ω} as *the distribution D conditioned on fixing S to ω*.

The main idea of our pseudorandom block restriction is to fix the variables using the output of a pseudorandom generator (PRG) for PTFs. Recall that a function $G : \{-1,1\}^s \to \{-1,1\}^n$ is a *PRG of seed length s that ϵ-fools PTFs of degree d* if, for any degree-d PTF f, we have

$$\left|\mathbf{Pr}_{z \sim \{-1,1\}^s}[f(G(z)) = -1] - \mathbf{Pr}_{x \sim \{-1,1\}^n}[f(x) = -1]\right| \leq \epsilon.$$

Definition 7.2. *Suppose the variables of a polynomial are arbitrarily partitioned into m blocks. We call a random block restriction ρ (m,ϵ)-fooling if it selects a block uniformly at random and fixes all variables outside the selected block using some distribution D that ϵ-fools PTFs of degree $2d$ (in the appropriate number of variables). Moreover, we call such a random block restriction (m,ϵ,K)-fooling if D ϵ-fools PTFs of degree $2d$ even conditioned on fixing at most K coordinates and is $\left(192 \cdot d \cdot \log \delta^{-1} + K\right)$-wise independent.*

We will use the construction of PRGs for PTFs due to Meka and Zuckerman. First recall that a multidimensional distribution on $\{-1,1\}^n$ is called *k-wise independent* if any k coordinates of the distribution are independent. A family of hash functions $\mathcal{H} = \{h \colon [n] \to [\ell]\}$ is called k-wise independent if for any $(x_1, \ldots, x_k) \in [n]^k$, where $x_1, \ldots, x_k$ are distinct, and any $(y_1, \ldots, y_k) \in [\ell]^k$, we have

$$\mathbf{Pr}_{h \sim \mathcal{H}}[h(x_1) = y_1 \wedge \cdots \wedge h(x_k) = y_k] = 1/\ell^k.$$

There exist k-wise independent distributions that can be sampled in $\mathsf{poly}(n,k)$ time using $O\left(k \cdot \log n\right)$ random bits, and there exist 2-wise independent hash families $\mathcal{H}$ such that a random $h \in \mathcal{H}$ can be sampled using $O\left(k \cdot \log(n \cdot \ell)\right)$ bits (see, e.g., [Vad12]).

The generator in the following theorem views its random seed as a tuple of $\ell+1$ disjoint random strings (for a certain parameter $\ell \geq 1$), and uses the first string to sample a hash function h, and the remaining ℓ strings to get ℓ samples from a k-wise independent distribution.

Theorem 7.3 ([MZ13])**.** *For $0 < \epsilon < 1$, let $\ell = 2^{O(d)} \cdot \log^2(\epsilon^{-1}) \cdot \epsilon^{-(4d+1)}$. Let $G : \{-1,1\}^s \to \{-1,1\}^n$ be the following process of generating an assignment for n coordinates using $s = 2^{O(d)} \cdot (\log n) \cdot \epsilon^{-(8d+3)}$ random bits:*

1. *Partition the n coordinates into ℓ buckets using a function $h\colon [n] \to [\ell]$ randomly picked from a 2-wise independent hash family.*
2. *For each bucket, generate a $(\ell + 4d)$-wise independent distribution for the coordinates in that bucket.*

Then G is a PRG that ϵ-fools n-variate PTFs of degree d.

Lemma 7.4. *For $0 < \epsilon < 1$, there exists a (m, ϵ, K)-fooling random block restriction that is samplable using $s = 2^{O(d)} \cdot \left(\epsilon^{-(16d+3)} + \epsilon^{-(8d+2)} \cdot (K + \log \delta^{-1})\right) \cdot \log n$ random bits.*

Proof. Let $\ell = 2^{O(d)} \cdot \log^2(\epsilon^{-1}) \cdot \epsilon^{-(8d+1)}$. Consider the distribution D sampled as follows:

1. Partition the n coordinates into ℓ buckets using a function $h\colon [n] \to [\ell]$ randomly picked from a 2-wise independent hash family.
2. For each bucket, generate a $\left(\ell + 4d + K + 192 \cdot d \cdot \log \delta^{-1}\right)$-wise independent distribution for the coordinates in that bucket.

Note that by Theorem 7.3, D ϵ-fools PTFs of degree $2d$ even conditioned on fixing at most K coordinates. This is because D has sufficient bounded independence for each bucket even conditioned on fixing K coordinates. Also, D is $\left(192 \cdot d \cdot \log \delta^{-1} + K\right)$-wise independent. Note that D is samplable using s random bits. We then define our (m, ϵ, K)-fooling random block restriction as the restriction that randomly selects a block and fixes all variables outside the selected block using D. Finally, note that the number of random bits needed to select a block is at most $\log n$. □

While it is possible to use a (m, ϵ, K)-fooling random block restriction to obtain a derandomized block restriction lemma, it requires large seed length to get small error if we do this in one shot. To deal with this issue, we use a sequence of pseudorandom block restrictions, so that, in each step, we only set the error parameter to match the probability that a random block restriction does not make the polynomial concentrated in the current step. Consider a random block restriction defined as follows. Let $m, \kappa \geq 16$.

1. Partition the m blocks of variables of p into $b = m^{1/\kappa d}$ disjoint super-blocks, where each super-block has m/b blocks.
2. Apply a $(b, \epsilon = b^{-1}, K)$-fooling random block restriction on the b super blocks.
3. Repeat the above two steps for the remaining blocks with m replaced by m/b until there is at most 2^d blocks, in which case we randomly choose a single block and fix the other variables using a $\left(192 \cdot d \cdot \log \delta^{-1} + K\right)$-wise independent distribution.

If at any round the blocks cannot be partitioned evenly into super-blocks, we can divide them so that the sizes of any two super-blocks differ by at most 1. We then select a super-block with probability proportional to its size. This makes sure that a block is selected uniformly at random. To avoid some technicalities that can be overcome easily, we assume here that at each round, the blocks can be partitioned evenly into super-blocks. Note that a random block restriction ρ generated as above can be decomposed into a sequence of sub-restrictions $\rho_1, \ldots, \rho_t, \rho_{t+1}$, where $t = O(\kappa \cdot d \cdot \log\log m)$ so that there are at most 2^d blocks remaining after $\rho_1, \ldots, \rho_t$, and each ρ_i, except the last one, is a (b_i, b_i^{-1}, K)-fooling random block restriction with $b_i = m^{(1-1/\kappa d)^{i-1}/\kappa d}$. Also, the last restriction fixes the variables using some $\left(192 \cdot d \cdot \log \delta^{-1} + K\right)$-wise independent distribution distribution. We call such a random block restriction (m, κ, K)-*good*. By the above, we have

Lemma 7.5. *There exists a (m, κ, K)-good block restriction that is samplable using*

$$2^{O(d)} \cdot \left(m^{(19/\kappa)} + m^{(10/\kappa)} \cdot (K + \log \delta^{-1})\right) \cdot \log n \cdot \kappa \cdot \log\log m$$

random bits.

To prove Theorem 7.1, we will show that the argument in Section 5 still goes through if we replace a truly random block restriction with our pseudorandom block restriction described above. We will need versions of the key lemmas in Section 5 for our pseudorandom block restrictions. In particular, we need a version of Lemma 4.2 for k-wise independent distributions. We first show a version of Theorem 2.3 for k-wise independent distributions. The following can be proved in the same way as Theorem 2.3.

Claim 7.6. *For any degree-d multilinear polynomial $p\colon \mathbb{R}^n \to \mathbb{R}$, any $T \geq 2^d$, and any $\left(\frac{d \cdot T^{2/d}}{2}\right)$-wise independent distribution D on $\{-1,1\}^n$, we have*

$$\mathbf{Pr}\left[|p(D)| \geq T \cdot \|p\|_2\right] \leq \exp\left(-(1/4) \cdot T^{2/d}\right).$$

Proof. Let $W = \frac{T^{2/d}}{2}$. By Markov's inequality, we have

$$\mathbf{Pr}\left[|p(D)| \geq T \cdot \|p\|_2\right] = \mathbf{Pr}\left[|p(D)|^W \geq (T \cdot \|p\|_2)^W\right] \leq \frac{\mathbf{Exp}[|p(D)|^W]}{(T \cdot \|p\|_2)^W}. \tag{59}$$

Since D is a $(d \cdot W)$-wise independent distribution on $\{-1,1\}^n$, we get, using Equation (1), that

$$\mathbf{Exp}[|p(D)|^W] = \|p\|_W^W \leq \left((W-1)^{d/2} \cdot \|p\|_2\right)^W \leq \left(W^{d/2} \cdot \|p\|_2\right)^W. \tag{60}$$

Combining Equations (59) and (60), we get

$$\mathbf{Pr}\left[|p(D)|^W \geq (T \cdot \|p\|_2)^W\right] \leq \left(\frac{W^{d/2}}{T}\right)^W \leq \exp\left(-(1/4) \cdot T^{2/d}\right),$$

as required. □

Claim 7.7. *For any degree-d multilinear polynomial p that is $(\delta, \gamma+1)$-concentrated, let ρ be a random block restriction for p that picks a uniformly random block and assigns the variables outside the block using a $(192 \cdot d \cdot \log \delta^{-1})$-wise independent distribution. Then we have that*

$$\mathbf{Pr}_\rho[p_\rho \text{ is not } (\delta, \gamma)\text{-concentrated}] \leq \delta.$$

Proof. The proof is the same as that of Lemma 4.2, with Claim 7.6 replacing Theorem 2.3. □

Next, we show two recurrence relations that are similar to Lemma 5.3 and Lemma 5.4, but with respect to our pseudorandom block restrictions.

Definition 7.8. *Let $\mathcal{Q}(d, m, \delta, \kappa, K, a)$ be the supremum, over all degree-d polynomials p with $\alpha(p) \leq a$, all possible partitions of the variables into m blocks, and all (m, κ, K)-good random block restrictions ρ, of the probabilities*

$$\mathbf{Pr}_{\rho}\left[p_{\rho} \text{ is not } \delta\text{-concentrated}\right].$$

Let $\mathcal{Q}_{\mathrm{reg}}(d, m, \delta, \kappa, K, a, \tau)$ be the same as $\mathcal{P}$ but only for τ-regular polynomials. For simplicity, we will omit some parameters when they are clear in the context. In particular, we will use $\mathcal{Q}(m, K, a)$ (resp. $\mathcal{Q}_{\mathrm{reg}}(m, K, a, \tau)$) for $\mathcal{Q}(d, m, \delta, \kappa, K, a)$ (resp. $\mathcal{Q}_{\mathrm{reg}}(d, m, \delta, \kappa, K, \tau)$).

Lemma 7.9. *For any $0 < \tau, \delta < 1/4$ and $a > 0$, $m > 4$, $\kappa \geq 16$, $d > 1$, and $K \geq H$, where $H = \tau^{-1} \cdot (d \cdot \log \tau^{-1} \cdot \log \delta^{-1})^{O(d)}$, we have*

$$\mathcal{Q}(m, K, a) \leq \frac{1}{m} \cdot \tau^{-1} \cdot (d \cdot \log \tau^{-1} \cdot \log \delta^{-1})^{O(d)} + 2\delta + \mathbf{Exp}_{\aleph}[\mathcal{Q}_{\mathrm{reg}}(m, K - H, \aleph, \tau)],$$

where $\aleph$ is a non-negative random variable with $\mathbf{Exp}[\aleph] \leq O(a)$.

Proof. The proof is very similar to that of Lemma 5.3, but has a few critical differences. For clarity, we provide details for this proof. Let p be a degree-d multilinear polynomial whose variables are partitioned into m blocks. Let ρ be a (m, κ, K)-good random block restriction. Consider the decision tree given by Theorem 2.11 with $\epsilon = \delta$ and $\gamma = 2$. Note that the depth of this decision tree is H. Since a block is chosen uniformly at random, the probability that ρ is not consistent with any branch of the decision tree is at most H/m.

Next we show the following.

Claim 7.10.

$$\mathbf{Pr}_{\rho}\left[p_{\rho} \text{ is not } \delta\text{-concentrated}\right] \leq \mathbf{Exp}_{\omega}\left[\mathbf{Pr}_{\rho_{\omega}}\left[(p_{\omega})_{\rho_{\omega}} \text{ is not } \delta\text{-concentrated}\right]\right] + 3(H/m),$$

where the expectation is over random leaves ω of the decision tree, p_{ω} is the restriction of p obtained by fixing the variables on the branch leading to ω as specified by the branch, and ρ_{ω} is an $(m, \kappa, K - H)$-good random block restriction.

Proof. We view ρ as $\rho = (\ell, \lambda)$, where ℓ is the selected block and λ is an assignment to the variables outside block ℓ. We can view the distribution of λ as a sequence of distributions, each ϵ-fooling PTFs of degree $2d$ even conditioned on fixing at most K coordinates, and being $\left(192 \cdot d \cdot \log \delta^{-1} + K\right)$-wise independent.

For each leaf ω, let R_{ω} be the set of (m, κ, K)-good random restrictions consistent with the branch leading to ω. As observed above, the probability ξ of choosing a restriction from the complement of $\cup_{\omega} R_{\omega}$ is at most H/m. We get

$$\mathbf{Pr}_{\rho}[p_{\rho} \text{ is not } \delta\text{-concentrated}] \leq (1 - \xi) \cdot \mathbf{Pr}_{\rho \in \cup_{\omega} R_{\omega}}[p_{\rho} \text{ is not } \delta\text{-concentrated}] + \xi. \tag{61}$$

By conditioning on $\rho \in R_\omega$, we get that $\mathbf{Pr}_{\rho \in \cup_\omega R_\omega}[p_\rho \text{ is not } \delta\text{-concentrated}]$ equals to

$$\sum_\omega \mathbf{Pr}_\rho[p_\rho \text{ is not } \delta\text{-concentrated} \mid \rho \in R_\omega] \cdot \mathbf{Pr}[\rho \in R_\omega \mid \rho \in \cup_\omega R_\omega].$$

As ρ is K-wise independent and $K \geq H$, the probability of choosing $\rho \in R_\omega$ conditioned on $\rho \in \cup_\omega R_\omega$ is $2^{-\ell_\omega} \cdot (1-\xi)^{-1}$, where ℓ_ω is the length of the branch leading to ω. Hence, the right-hand side of Equation (61) is at most

$$\mathbf{Exp}_\omega[\mathbf{Pr}_{\rho \in R_\omega}[p_\rho \text{ is not } \delta\text{-concentrated}]] + \xi. \tag{62}$$

Each (m, κ, K)-good restriction $\rho = (\ell, \lambda) \in R_\omega$ can be viewed as a restriction of the variables on the branch leading to ω (as specified by the branch) plus a restriction λ' to the remaining variables outside block ℓ. So we can express p_ρ as $(p_\omega)_{\rho'}$, where $\rho' = (\ell, \lambda')$.

Note that ρ' is a $(m, \kappa, K-H)$-good restriction, which comes from the set of those $(m, \kappa, K-H)$-good restrictions that chose block ℓ outside at most H blocks containing the variables on the branch leading to ω. The set of all such restrictions ρ' has the probability mass at least $1 - H/m$ within the set of all $(m, \kappa, K-H)$-good restrictions ρ_ω (which pick block ℓ uniformly at random from the set of all m blocks). Therefore, we can upperbound the expression in Equation (62) by

$$\begin{aligned}
&\mathbf{Exp}_\omega[\mathbf{Pr}_{\rho'}[(p_\omega)_{\rho'} \text{ is not } \delta\text{-concentrated}]] + \xi \\
&\quad \leq (1 - H/m)^{-1} \cdot \mathbf{Exp}_\omega[\mathbf{Pr}_{\rho_\omega}[(p_\omega)_{\rho_\omega} \text{ is not } \delta\text{-concentrated}]] + \xi \\
&\quad \leq \mathbf{Exp}_\omega[\mathbf{Pr}_{\rho_\omega}[(p_\omega)_{\rho_\omega} \text{ is not } \delta\text{-concentrated}]] + \xi + 2(H/m),
\end{aligned}$$

where the last inequality uses the fact that $(1-x)^{-1} \leq 1 + 2x$ whenever $0 < x \leq 1/2$. The claim follows. □

The proof can now proceed in the same way as that of Lemma 5.3, but instead of using Lemma 4.2 there, we use Claim 7.7 and the fact that ρ_ω fixes the variables $(192 \cdot d \cdot \log \delta^{-1})$-wise independently. □

Lemma 7.11. *For any real $0 < \tau, \delta < 1/4$ and $a > 0$, $m > 4$, $\kappa \geq 16$, $K >\geq 0$ and $d > 1$, we have*

$$\mathcal{Q}_{\text{reg}}(m, K, a, \tau) \leq \mathbf{Exp}_{\aleph}[\mathcal{Q}(m^{1-1/\kappa d}, K, \aleph)],$$

where $\aleph$ is a non-negative random variable with $\mathbf{Exp}[\aleph] = O\left(d^3 a m^{-1/2\kappa d} + d^4 \tau^{1/(8d)} + m^{-1/\kappa d}\right)$.

Proof. For a (m, κ, K)-good random block restriction ρ, we can decompose it into two restrictions ρ_1 and ρ', where ρ_1 is a $(b = m^{1/\kappa d}, \epsilon = m^{-1/\kappa d}, K)$-fooling random block restriction and ρ' is a $(m^{1-1/\kappa d}, \kappa, K)$-good random block restriction. Then

$$\begin{aligned}
\mathbf{Pr}_\rho[p_\rho \text{ is not } \delta\text{-concentrated}] &= \mathbf{Exp}_{\rho_1}\left[\mathbf{Pr}_{\rho'}[(p_{\rho_1})_{\rho'} \text{ is not } \delta\text{-concentrated}]\right] \\
&\leq \mathbf{Exp}_{\rho_1}\left[\mathcal{Q}(m^{1-1/\kappa d}, K, \alpha(p_{\rho_1}))\right].
\end{aligned}$$

Therefore, we need to show

$$\mathbf{Exp}_{\rho_1}[\alpha(p_{\rho_1})] = O\left(d^3 a b^{-1/2} + d^4 \tau^{1/(8d)} + \epsilon\right).$$

From Equation (36), for a truly random block restriction $\sigma \sim B_b$, we have

$$\mathbf{Exp}_\sigma[\alpha(p_\sigma)] = O\left(d^3 a b^{-1/2} + d^4 \tau^{1/(8d)}\right).$$

Thus, it suffices to show that

$$|\ \mathbf{Exp}_\sigma[\alpha(p_\sigma)] - \mathbf{Exp}_{\rho_1}[\alpha(p_{\rho_1})]\ | \leq \epsilon.$$

Consider a degree-d multilinear polynomial p whose variables are partitioned into m blocks. Fixed a block ℓ. Let A_ℓ be an assignment to the variables in block ℓ, B be a vector of dimension the same as the number of variables in block ℓ, and t be an arbitrary number. Define $T_{\ell,A_\ell,B,t}$ to be the boolean function on input D such that $T_{\ell,A_\ell,B,t}(D) = -1$ if and only if

$$|p_D(A_\ell)|^2 \leq t \cdot |\mathsf{D}_B\, p_D(A_\ell)|^2.$$

It is easy to see that $T_{\ell,A_\ell,B,t}$ is a PTF of degree at most $2d$.

Now for a block, ℓ, we let A_ℓ denote the random assignment to the variables in ℓ and let $A_{\bar{\ell}}$ denote the random assignment to the variables that are not in ℓ. Let D be the distribution by which ρ_1 fixes the variables. Note that D ϵ-fools PTF of degree $2d$. We have

$$\begin{aligned}
\mathbf{Exp}_\sigma[\alpha(p_\sigma)] &= \frac{1}{b} \cdot \sum_{\ell \in [b]} \mathbf{Exp}_{A_{\bar{\ell}}}[\alpha(p_{A_{\bar{\ell}}})] \\
&= \frac{1}{b} \cdot \sum_{\ell} \mathbf{Exp}\left[\min\left\{1, \frac{|\mathsf{D}_B\, p_{A_{\bar{\ell}}}(A_\ell)|^2}{|p_{A_{\bar{\ell}}}(A_\ell)|^2}\right\}\right] \\
&= \frac{1}{b} \cdot \sum_{\ell} \int_0^1 \mathbf{Pr}\left[|p_D(A_\ell)|^2 \leq t \cdot |\mathsf{D}_B\, p_D(A_\ell)|^2\right] dt \\
&= \frac{1}{b} \cdot \sum_{\ell} \int_0^1 \mathbf{Pr}\left[T_{\ell,A_\ell,B,t}(A_{\bar{\ell}})\right] dt \\
&\leq \frac{1}{b} \cdot \sum_{\ell} \int_0^1 \left(\mathbf{Pr}\left[T_{\ell,A_\ell,B,t}(D)\right] + \epsilon\right) dt \\
&\leq \left(\frac{1}{b} \cdot \sum_{\ell} \int_0^1 \mathbf{Pr}\left[T_{\ell,A_\ell,B,t}(D)\right] dt\right) + \epsilon \\
&= \mathbf{Exp}_{\rho_1}[\alpha(p_{\rho_1})] + \epsilon.
\end{aligned}$$

The other direction can be shown similarly. □

We are now ready to prove the following result, which, together with our construction of (m, κ, K)-good random block restrictions in Lemma 7.5, will imply Theorem 7.1

Theorem 7.12. *There exist constants $B, C > 0$ such that, for any $d > 0$, $m, \kappa \geq 16$, $0 < \delta \leq 1/16$, $0 < a \leq 1$, and $K = m^{8/\kappa} \cdot (d \cdot \log m \cdot \log \delta^{-1})^{C \cdot d}$, we have*

$$\mathcal{Q}(d, m, \delta, \kappa, K, a) \leq (a \cdot m^{-1/2} + \delta) \cdot (\log m)^{B\kappa \cdot d \cdot \log d} \cdot (\log \delta^{-1})^{B\kappa \cdot d^2}. \tag{63}$$

Proof. The proof of is similar to that of Theorem 5.9. As in the proof of Theorem 5.9, we can assume, for a sufficiently large constant $B > 0$, that both

$$a \geq (c \cdot \log \delta^{-1})^{-2d}, \tag{64}$$

and

$$m^{1/(2\kappa \cdot d)} \geq (c \cdot \log \delta^{-1})^{2d}, \tag{65}$$

where $c > 0$ is the constant from Corollary 3.7. Note that if Equation (64) if false. Then by corollary 3.7, the polynomial is $(\delta, 2)$-concentraed and by Claim 7.7 such a polynomial will remain δ-concentrated under restrictions that fix variables $(192 \cdot d \cdot \log \delta^{-1})$-wise independently except with probability at most δ.

Now let $\tau = m^{-8/\kappa}$ and $H = \tau^{-1} \cdot (d \cdot \log \tau^{-1} \cdot \log \delta^{-1})^{O(d)}$. Note that $K \geq H$ when C is sufficiently large. Then combining Lemma 7.9 and Lemma 7.11, and proceeding as in the proof of Claim 5.11, we have, for a sufficiently large constant E,

$$\begin{aligned} \mathcal{Q}(m, K, a) \leq (a \cdot m^{-1/2} + \delta) \cdot (\log m)^{E \cdot \kappa \cdot d \cdot \log d} \cdot (\log \delta^{-1})^{E \cdot \kappa \cdot d^2} \\ + \mathbf{Exp}_{\aleph}[\mathcal{Q}(m^{1-1/(\kappa d)}, K - H, \aleph)], \end{aligned} \tag{66}$$

where $\aleph$ is a non-negative random variable with $\mathbf{Exp}[\aleph] = O\left(d^4 \cdot a \cdot m^{-1/(2\kappa d)}\right)$.

To solve the recurrence relation given by Equation (66), we again use induction on m. The base case is $m \leq 2^d$. As $a \geq (c \cdot \log \delta^{-1})^{-2d}$, in this case the right hand side of Equation (63) is greater than 1 for when B is sufficiently large. Now suppose Theorem 5.9 holds for all smaller values of m. Let $m_1 = m^{1-1/(\kappa d)}$. Then when C is sufficiently large, we have

$$K - H \geq m_1^{8/\kappa} \cdot (d \cdot \log m_1 \cdot \log \delta^{-1})^{C \cdot d}.$$

Therefore, we can apply the induction hypothesis on

$$\mathcal{Q}(m^{1-1/(\kappa d)}, K - H, \aleph)$$

in Equation (66). After applying the induction hypothesis and proceeding as in the proof of Theorem 5.9, we can complete the induction step and hence the proof. □

By Lemma 7.5 and Theorem 7.12, there exist constants $B, C > 0$ and a pseduorandom block restriction samplable using

$$m^{(19/\kappa)} \cdot \log n \cdot \kappa \cdot (d \cdot \log m \cdot \log \delta^{-1})^{C \cdot d} \tag{67}$$

random bits such that for any degree-d multilinar polynomial p,

$$\mathbf{Pr}_{\rho}\left[p_{\rho} \text{ is not } \delta\text{-concentrated}\right] \leq \left(m^{-1/2} + \delta\right) \cdot (\log m)^{B \cdot \kappa \cdot d \cdot \log d} \cdot (\log \delta^{-1})^{B \cdot \kappa \cdot d^2}. \tag{68}$$

Note that we can assume without loss of generality that both

$$\kappa \leq \frac{\log m}{d^2 \cdot \log \log m}$$

and

$$(d \cdot \log m \cdot \log \delta^{-1})^{C \cdot d} \leq m^{1/\kappa}.$$

Otherwise, the right hand side of the Equation (68) is greater than 1 when B is sufficiently large. Then Equation (67) is at most

$$m^{O(1/\kappa)} \cdot \log n.$$

By changing the parameter κ, we obtain Theorem 7.1.

7.2 Derandomized Littlewood-Offord type anticoncentration bounds

Here we show two versions of derandomized anticoncentration bounds for degree-d multilinear polynomials. We first show the following.

Theorem 7.13. *For any positive integers t and d, and $0 < \zeta < 1$, there exists a distribution D on $\{-1,1\}^n$, samplable in $\mathsf{poly}(n)$ time using $t^{\zeta/d} \cdot \log n$ random bits, such that the following holds. For any real interval I, and any n-variate degree-d multilinear polynomial p that has at least t disjoint degree-d monomials with coefficient at least $|I|$ in magnitude, we have*

$$\mathbf{Pr}\left[p(D) \in I\right] \leq t^{-\frac{1}{2d}} \cdot (\log t)^{O\left(\zeta^{-1} \cdot d^2\right)}.$$

Let us call a degree-d monomial *good* if its coefficient is at least $|I|$ in magnitude, and say that a set of variables *contains* a monomial if every variable in the monomial is in the set. Also, let us call a partition of variables into blocks *good* if every block contains at least one good monomial. From the analysis in Theorem 6.14, it is easy to see that if we are explicitly given a degree-d polynomial with at least t disjoint good monomials, then we can obtain a good partition with t blocks and use our derandomized Block Restriction Lemma to generate the inputs so that the polynomial will take value inside the interval with probability at most about $t^{-1/2}$.

However, we want to derandomize obliviously, without knowing the structure of the polynomial. The idea is to partition the variables randomly, using bounded-independent hashing, so that we get a good partition with high probability. To show that bounded-independent hashing will produce a good partition with high probability, we need the following version of Chernoff bounds for bounded-independent random variables.

Theorem 7.14 ([SSS95])**.** *Let $\epsilon \leq 1$. If X is the sum of k-wise independent random variables taking values in $[0,1]$, and $\mu = \mathbf{Exp}[X]$ such that $k \leq \lfloor \epsilon^2 \mu e^{-1/2} \rfloor$, then*

$$\mathbf{Pr}[|X - \mu| > \epsilon\mu] < \exp(-\lfloor k/2 \rfloor).$$

We now show the following.

Lemma 7.15. *Let p be a n-variate degree-d multilinear polynomial with at least t disjoint good monomials. If the variables are partitioned into $m = t^{1/d}/\log^{2/d}(t)$ blocks using a random hash function from a $(Cd \log t)$-wise independent hash family, where $C > 0$ is some constant, then the probability that the partition is not good is at most $1/t$.*

Proof. Fix a block ℓ. Let $m_1, \ldots, m_t$ be the t disjoint good monomials. For $i = 1, \ldots, t$, let X_i be the indicator random variable for the event that ℓ contains m_i, using a $(Cd \log t)$-wise independent hashing (i.e., X_i is 1 if every variable in m_i is hashed to ℓ, and 0 otherwise). Note that $\mathbf{Pr}[X_i = 1] = 1/m^d$ for every i, and $X_1, \ldots, X_t$ are $(C \log t)$-wise independent. Let $X = X_1 + \cdots + X_t$ and $\mu = \mathbf{Exp}[X] = t/m^d = \log^2 t$. By Theorem 7.14, we have

$$\mathbf{Pr}[X = 0] \leq \mathbf{Pr}[|X - \mu| > (1/2)\mu] \leq \exp(-\lfloor (C \log t)/2 \rfloor) \leq 1/t^2,$$

where the last inequality holds if C is sufficiently large. Taking the union bound over the m blocks, we conclude that probability that there exists one block that does not contain any good monomial is at most $1/t$. □

We will also need the following version of Claim 6.15 for bounded-independent distributions, whose proof is the same as Claim 6.15, with Claim 7.6 replacing Theorem 2.3.

Claim 7.16. *For any real interval I centered at 0, any δ-concentrated degree-d multilinear polynomial q that has at least one monomial with coefficient greater than $|I|$ in magnitude, and any $\left(192 \cdot d \cdot \log \delta^{-1}\right)$-wise independent distribution D on $\{-1,1\}^n$, we have*

$$\mathbf{Pr}[q(D) \in I] \leq \delta.$$

Proof of Theorem 7.13. Consider the following process of sampling from D.

1. Partition the variables of p into $m = t^{1/d}/\log^{2/d}(t)$ blocks using $(Cd \log t)$-wise independent hashing, where C is the constant from Lemma 7.15.
2. Apply the derandomized Block Restriction Lemma (Theorem 7.1) based on the partition in the previous step with $\delta = m^{-1/2}$.
3. Fix the variables in the last block (i.e., the unrestricted block after applying random block restriction in the previous step) using a $(192 \cdot d \cdot \log \delta^{-1})$-wise independent distribution.

The amount of random bits used in the first step is $O\left(d \cdot \log t \cdot \log n\right)$, and the amount of random bits needed in the second step is at most

$$m^{\zeta} \cdot \log n \leq t^{\zeta/d} \cdot \log n.$$

The last step only needs $O\left(d \cdot \log \delta^{-1} \cdot \log n\right)$ random bits. Therefore, the total amount of random bits needed for the above process is at most $O\left(t^{\zeta/d} \cdot \log n\right)$.

We now show the correctness. By Lemma 7.15, the probability that the partition obtained in the first step is not good is at most $1/t$. Given that the partition in the first step is good, any restricted polynomial after the second step will have at least one good monomial, and by Theorem 7.1 the probability that the restricted polynomial is not δ-concentrated is at most

$$\left(m^{-1/2} + \delta\right) \cdot \left(\log m \cdot \log \delta^{-1}\right)^{O\left(\zeta^{-1} \cdot d^2\right)}.$$

Finally, given that the restricted polynomial in the second step has at least one good monomial and is δ-concentrated, the probability that it falls inside the interval I after taking an input from a $(192 \cdot d \cdot \log \delta^{-1})$-wise independent distribution is at most δ by Claim 7.16. Therefore, by noting $\delta = m^{-1/2}$, the probability that an input obtained in the above process makes p fall inside I is at most

$$1/t + \left(m^{-1/2}\right) \cdot (\log m)^{O\left(\zeta^{-1} \cdot d^2\right)} + m^{-1/2} \leq t^{-1/(2d)} \cdot (\log t)^{O\left(\zeta^{-1} \cdot d^2\right)}.$$

This completes the proof. □

Next, we show another derandomized anticoncentration bound that is quantitatively better when the polynomials are dense (i.e., have many good monomials).

Theorem 7.17. *For any positive integers t and d, and $0 < \zeta < 1$, there exists a distribution D on $\{-1,1\}^n$, samplable in $\mathsf{poly}(n)$ time using $t^{\zeta} \cdot \log n$ random bits, such that the following holds. For any real interval I, and any n-variate degree-d multilinear polynomial p with at least $t \cdot n^{d-1}$ degree-d monomials whose coefficients are at least $|I|$ in magnitude, we have*

$$\mathbf{Pr}\left[p(D) \in I\right] \leq t^{-\frac{1}{2}} \cdot (\log t)^{O\left(\zeta^{-1} \cdot d^2\right)}.$$

Remark 7.18. *For dense polynomials with $t = n^{1-o(1)}$ and any $\epsilon > (C \cdot d^2 \cdot \log\log n)/\log n$, where C is some constant, setting*

$$\zeta = (C \cdot d^2 \cdot \log\log n)/(\epsilon \cdot \log n),$$

we get that the bound in Theorem 7.17 is at most $n^{-1/2+o(1)+\epsilon}$, matching the bound in Theorem 6.14 up to the $n^{o(1)+\epsilon}$ factor, and that the seed length is at most $(\log n)^{O\left(\epsilon^{-1} \cdot d^2\right)}$. Such a short seed is beyond reach of the naive derandomization using the PRG from Theorem 7.3, when the error is inverse-polynomially small.

To show Theorem 7.17, we a gain use bounded-independent hashing to partition the variables.

Lemma 7.19. *Let p be a n-variate degree-d multilinear polynomial with at least $t \cdot n^{d-1}$ good monomials. If the variables are partitioned into $m = t/(C \log t)$ blocks using a random hash function from a $(Cd \log t)$-wise independent hash family, where C is a constant, then the probability that the partition is not good is at most $1/t$.*

Proof. We first consider using full randomness to partition the variables. It will be convenient to view a set of variables as an n-bit characteristic string, where a coordinate i is 1 if the ith variable is in the set, and 0 otherwise. For a set S, we will also use S to denote its characteristics string. Let $K = (C/2) \log t$, and let p be so that $1 - (1-p)^K = 1/m$.

Consider a random set U that picks each variable independently with probability p. Note that U can be viewed as a random n-bit string such that each coordinate is 1 with probability p. Also, given U, we can compute the number of good monomials contained in U, using a degree-d polynomial, which is simply the sum of the $t \cdot n^{d-1}$ monomials of p. Let q denote this polynomial and let $\mu = \mathbf{Exp}[q(U)]$. Now define $r(U) = (q(U) - \mu)^2$. Note that r is a polynomial of degree at most $2d$ and, given our value of p, we have

$$\mu^2 > 2 \cdot \mathbf{Exp}[r(U)]. \tag{69}$$

Also, if U does not contain any good monomial, then $r(U) = \mu^2$.

Let $U_1, \ldots, U_K$ be K independent random sets, where each U_i picks each variable independently with probability p. Let T be a random set that picks each variable independently with probability $1/m = 1 - (1-p)^K$. Note that $U_1 \cup \cdots \cup U_K$ and T have the same distribution. Given a set T, consider the following way of picking a tuple of K random subsets $V^T = V_1^T, \ldots, V_K^T$ of T: pick V^T from the distribution of $U_1, \ldots, U_K$, conditioned on $U_1 \cup \cdots \cup U_K = T$. Now define f as

$$f(T) = \mathbf{Exp}_{V^T}\left[\prod_{i=1}^{K} r(V_i^T)\right].$$

Note that f can be written as a polynomial of degree at most $2dK$. To see this, consider the following equivalent way of picking V^{T_0} for some T_0: for each variable that appears in T_0, we assign it to each of the K subsets with probability p, conditioned on at least one subset containing the variable. Now consider picking a tuple of K random sets $W = W_1, \ldots, W_K$ in the above way, with $T_0 = \{1, \ldots, n\}$. Then it is easy to see that, for any given set T, $W \cap T = W_1 \cap T, \ldots, W_1 \cap T$ (i.e., after we pick W we remove all the variables that are not in T) and V^T have the same distribution. Therefore, we have

$$f(T) = \mathbf{Exp}_{W}\left[\prod_{i=1}^{K} r(W_i \cap T)\right],$$

which is clearly a polynomial of degree at most $2dK$ since r is of degree at most $2d$. Note that since each V_i^T is a subset of T, if T does not contain any good monomial, then

$$f(T) = \mu^{2K}. \tag{70}$$

Also, by the definition of the distribution for V^T, we have

$$\begin{aligned}
\mathbf{Exp}_T[f(T)] &= \mathbf{Exp}_T\left[\mathbf{Exp}_{V^T}\left[\prod_{i=1}^{K} r(V_i^T)\right]\right] \\
&= \mathbf{Exp}_{U_1,\ldots,U_K}\left[\prod_{i=1}^{K} r(U_i)\right] \\
&= \prod_{i=1}^{K} \mathbf{Exp}\left[r(U_i)\right] \\
&< \mu^{2K}/2^K,
\end{aligned} \tag{71}$$

where the last inequality is by Equation (69).

Now consider partitioning the n variables into m blocks, using a $(Cd\log t)$-wise independent hash family $\mathcal{H}$. Let D be a random n-bits string such that the coordinates are $(2dK)$-wise independent and each coordinate is 1 with probability $1/m$. Let T be a random set that takes each variable independently with probability $1/m$. Then, for a block $\ell \in [m]$, we have

$$\begin{aligned}
&\mathbf{Pr}_{h\sim\mathcal{H}}[\ell \text{ does not contain any good monomial under } h] \\
&\leq \mathbf{Pr}_D[f(D) = \mu^{2K}] && \text{(by Equation (70))} \\
&\leq \mathbf{Pr}_D\left[f(D) > 2^K \cdot \mathbf{Exp}_T[f(T)]\right] && \text{(by Equation (71))} \\
&= \mathbf{Pr}_D\left[f(D) > 2^K \cdot \mathbf{Exp}_D[f(D)]\right] \\
&\leq 2^{-K} \\
&\leq t^{-C/2},
\end{aligned}$$

where the forth line above is by the fact that f is a polynomial of degree at most $2dK = Cd\log t$ and that D is $(Cd\log t)$-wise independent, and the second last line is by Markov's inequality. Finally, by the union bound over the m blocks, and for C sufficiently large, we get that the probability that there exists one block that does not contain any good monomial is at most $1/t$. □

Given Lemma 7.19, Theorem 7.17 is now proved in the same way as Theorem 7.13.

8 Open problems

We proved a restriction lemma for PTFs of degree $d \geq 1$, and used it to derive new lower bounds against constant-depth circuits with PTF gates. What are other applications of the (derandomized) PTF Restriction Lemma? For example, can it be used to get a PRG for constant-depth PTF circuits? Can we get a nontrivial (better than brute force) SAT algorithm for PTFs (constant-depth PTF circuits)? Finally, what are the applications of derandomized Littlewood-Offord type anticoncentration bounds?

References

[ACW16] Josh Alman, Timothy M. Chan, and Ryan Williams. Polynomial representations of threshold functions and algorithmic applications. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS'16)*, 2016. 4

[Ajt83] Miklós Ajtai. Σ^1_1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24(1):1 – 48, 1983. 1, 3, 7

[AK10] Eric Allender and Michal Koucký. Amplifying lower bounds by means of self-reducibility. *J. ACM*, 57(3), 2010. 8

[All89] Eric Allender. A note on the power of threshold circuits. In *Proceedings of the Thirtieth Annual Symposium on Foundations of Computer Science (FOCS), Research Triangle Park, North Carolina, USA, 30 October - 1 November 1989*, pages 580–584, 1989. 1, 3

[And87] Alexander E. Andreev. On a method of obtaining more than quadratic effective lower bounds for the complexity of π-schemes. *Vestnik Moskovskogo Universiteta. Matematika*, 42(1):70–73, 1987. English translation in *Moscow University Mathematics Bulletin*. 1, 4, 7

[Ant01] Martin Anthony. *Discrete Mathematics of Neural Networks: Selected Topics*. SIAM monographs on discrete mathematics and applications. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2001. 3

[Bon70] Aline Bonami. Étude des coefficients Fourier des fonctions de $L^p(G)$. *Annales de l'Institut Fourier*, 20(2):335–402, 1970. 11

[Bru90] Jehoshua Bruck. Harmonic analysis of polynomial threshold functions. *SIAM Journal on Discrete Mathematics*, 3(2):168–177, 1990. 4

[BS92] Jehoshua Bruck and Roman Smolensky. Polynomial threshold functions, AC^0 functions, and spectral norms. *SIAM J. Comput.*, 21(1):33–42, 1992. 4

[Cho61] Chao-Kong Chow. On the characterization of threshold functions. In *Proceedings of the 2nd Annual Symposium on Switching Circuit Theory and Logical Design (FOCS)*, pages 34–38, 1961. 33

[CKK+15] Ruiwen Chen, Valentine Kabanets, Antonina Kolokolova, Ronen Shaltiel, and David Zuckerman. Mining circuit lower bound proofs for meta-algorithms. *Computational Complexity*, 24(2):333–392, 2015. 34, 36

[CSS16] Ruiwen Chen, Rahul Santhanam, and Srikanth Srinivasan. Average-case lower bounds and satisfiability algorithms for small threshold circuits. In Ran Raz, editor, *31st Conference on Computational Complexity (CCC 2016)*, volume 50 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 1:1–1:35, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. 1, 4, 5, 7, 8, 35, 36

[CW01] Anthony Carbery and James Wright. Distributional and L^q norm inequalities for polynomials over convex bodies in R^n. *Mathematical Research Letters*, 8(3):233–248, 5 2001. 9, 11

[DDS14] Anindya De, Ilias Diakonikolas, and Rocco A. Servedio. Deterministic approximate counting for juntas of degree-2 polynomial threshold functions. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 229–240, 2014. 9

[DGJ+10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. Bounded independence fools halfspaces. *SIAM J. Comput.*, 39(8):3441–3462, 2010. 9

[DKN10] Ilias Diakonikolas, Daniel M. Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 11–20, 2010. 9

[DOSW11] Ilias Diakonikolas, Ryan O'Donnell, Rocco A. Servedio, and Yi Wu. Hardness results for agnostically learning low-degree polynomial threshold functions. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, San Francisco, California, USA, January 23-25, 2011*, pages 1590–1606, 2011. 9

[DRST14] Ilias Diakonikolas, Prasad Raghavendra, Rocco A. Servedio, and Li-Yang Tan. Average sensitivity and noise sensitivity of polynomial threshold functions. *SIAM J. Comput.*, 43(1):231–253, 2014. 9

[DS14] Anindya De and Rocco A. Servedio. Efficient deterministic approximate counting for low-degree polynomial threshold functions. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 832–841, 2014. 9

[DSTW10] Ilias Diakonikolas, Rocco A. Servedio, Li-Yang Tan, and Andrew Wan. A regularity lemma, and low-weight approximators, for low-degree polynomial threshold functions. In *IEEE Conference on Computational Complexity*, 2010. 22

[DSTW14] Ilias Diakonikolas, Rocco A. Servedio, Li-Yang Tan, and Andrew Wan. A regularity lemma and low-weight approximators for low-degree polynomial threshold functions. *Theory of Computing*, 10:27–53, 2014. 9

[Erd45] Paul Erdős. On a lemma of littlewood and offord. *Bull. Amer. Math. Soc.*, 51:898–902, 1945. 6, 8

[FSS84] Merrick L. Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, 1984. 1, 3, 7

[GHR92] Mikael Goldmann, Johan Håstad, and Alexander A. Razborov. Majority gates vs. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992. 3, 8

[GL94] Craig Gotsman and Nathan Linial. Spectral properties of threshold functions. *Combinatorica*, 14(1):35–50, 1994. 6, 9

[Hås89] Johan Håstad. Almost optimal lower bounds for small depth circuits. In S. Micali, editor, *Randomness and Computation*, pages 143–170, Greenwich, Connecticut, 1989. Advances in Computing Research, vol. 5, JAI Press. 1, 3, 7

[Hås98] Johan Håstad. The shrinkage exponent of de Morgan formulas is 2. *SIAM J. Comput.*, 27(1):48–64, 1998. 7

[Hås16] Johan Håstad. An average-case depth hierarchy theorem for higher depth. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:41, 2016. 7

[HMP+93] András Hajnal, Wolfgang Maass, Pavel Pudlák, Mario Szegedy, and György Turán. Threshold circuits of bounded depth. *J. Comput. Syst. Sci.*, 46(2):129–154, 1993. 8

[IN93] Russell Impagliazzo and Noam Nisan. The effect of random restrictions on formula size. *Random Struct. Algorithms*, 4(2):121–134, 1993. 7

[IPS97] Russell Impagliazzo, Ramamohan Paturi, and Michael E. Saks. Size-depth tradeoffs for threshold circuits. *SIAM J. Comput.*, 26(3):693–707, 1997. 1, 4, 7, 8

[Kan11] Daniel M. Kane. A small PRG for polynomial threshold functions of Gaussians. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 257–266, 2011. 9

[Kan12] Daniel M. Kane. A structure theorem for poorly anticoncentrated Gaussian chaoses and applications to the study of polynomial threshold functions. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 91–100, 2012. 9

[Kan14] Daniel M. Kane. The correct exponent for the Gotsman-Linial conjecture. *Computational Complexity*, 23(2):151–175, 2014. 1, 6, 8, 9, 12, 14, 28, 36

[Kan15] Daniel M. Kane. A polylogarithmic PRG for degree 2 threshold functions in the Gaussian setting. In *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA*, pages 567–581, 2015. 9

[Khr71] V.M. Khrapchenko. A method of determining lower bounds for the complexity of π-schemes. *Matematicheskie Zametki*, 10(1):83–92, 1971. English translation in *Mathematical Notes of the Academy of Sciences of the USSR*. 7

[KW16] Daniel M. Kane and Ryan Williams. Super-linear gate and super-quadratic wire lower bounds for depth-two and depth-three threshold circuits. In *Proceedings of the Forty-Eighth ACM Symposium on Theory of Computing (STOC'16)*, 2016. 1, 4, 7, 8, 31, 32, 33

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *J. ACM*, 40(3):607–620, 1993. 1

[LO43] John E. Littlewood and A. Cyril Offord. On the number of real roots of a random algebraic equation (III). *Rec. Math. (Mat. Sbornik) N.S.*, 12 (54)(3):277–286, 1943. 6, 8, 31

[MK61] J. Myhill and W. H. Kautz. On the size of weights required for linear-input switching functions. *IRE Transactions on Electronic Computers*, EC-10(2):288–290, June 1961. 3

[MNV16] Raghu Meka, Oanh Nguyen, and Van Vu. Anti-concentration for polynomials of independent random variables. *Theory of Computing*, 12(11):1–17, 2016. 2, 6, 37, 38

[MOO10] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Annals of Mathematics*, 171(1):295–341, 2010. 9, 14

[MP43] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5(4):115–133, 1943. 3

[MP69] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, Mass., 1969. (3rd Edition published in 1988). 3

[MZ13] Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. *SIAM J. Comput.*, 42(3):1275–1301, 2013. 9, 40

[Nis94] Noam Nisan. The communication complexity of threshold gates. In *In Proceedings of Combinatorics, Paul Erdos is Eighty*, pages 301–315, 1994. 4, 8

[NR04] Moni Naor and Omer Reingold. Number-theoretic constructions of efficient pseudo-random functions. *J. ACM*, 51(2):231–262, 2004. 3

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. 10

[OS08] Ryan O'Donnell and Rocco A. Servedio. Extremal properties of polynomial threshold functions. *J. Comput. Syst. Sci.*, 74(3):298–312, 2008. 9

[Per04] Yuval Peres. Noise Stability of Weighted Majority. *arXiv.math/0412377*, 2004. 36

[PS94] Ramamohan Paturi and Michael E. Saks. Approximating threshold circuits by rational functions. *Inf. Comput.*, 112(2):257–272, 1994. 8

[PZ93] Mike Paterson and Uri Zwick. Shrinkage of de Morgan formulae under restriction. *Random Struct. Algorithms*, 4(2):135–150, 1993. 7

[Raz87] Alexander A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987. 3

[Raz92] Alexander A. Razborov. On small depth threshold circuits. In Otto Nurmi and Esko Ukkonen, editors, *Algorithm Theory — SWAT '92: Third Scandinavian Workshop on Algorithm Theory Helsinki, Finland, July 8–10, 1992 Proceedings*, pages 42–52, Berlin, Heidelberg, 1992. Springer Berlin Heidelberg. 3

[RR97] Alexander A. Razborov and Steven Rudich. Natural proofs. *J. Comput. Syst. Sci.*, 55(1):24–35, 1997. 3

[RSO94] Vwani Roychowdhury, Kai-Yeung Siu, and Alon Orlitsky. *Theoretical Advances in Neural Computation and Learning*, chapter Neural Models and Spectral Methods, pages 3–36. Springer US, Boston, MA, 1994. 33

[RST15] Benjamin Rossman, Rocco A. Servedio, and Li-Yang Tan. An average-case depth hierarchy theorem for boolean circuits. In Venkatesan Guruswami, editor, *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 1030–1048. IEEE Computer Society, 2015. 7

[Sak93] Michael E. Saks. Slicing the hypercube. In K. Walker, editor, *Surveys in Combinatorics, 1993*, pages 211–256. Cambridge University Press, 1993. 9

[SB91] Kai-Yeung Siu and Jehoshua Bruck. On the power of threshold circuits with small weights. *SIAM J. Discrete Math.*, 4(3):423–435, 1991. 8

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 77–82. ACM, 1987. 3

[SRK94] Kai-Yeung Siu, Vwani P. Roychowdhury, and Thomas Kailath. Rational approximation techniques for analysis of neural networks. *IEEE Trans. Information Theory*, 40(2):455–466, 1994. 8

[SSS95] Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff-hoeffding bounds for applications with limited independence. *SIAM J. Discrete Math.*, 8(2):223–250, 1995. 46

[STT12] Rocco A. Servedio, Li-Yang Tan, and Justin Thaler. Attribute-efficient learning and weight-degree tradeoffs for polynomial threshold functions. In *COLT 2012 - The 25th Annual Conference on Learning Theory, June 25-27, 2012, Edinburgh, Scotland*, pages 14.1–14.19, 2012. 9

[Sub61] Bella A. Subbotovskaya. Realizations of linear function by formulas using $\vee$, $\&$, $^-$. *Doklady Akademii Nauk SSSR*, 136(3):553–555, 1961. English translation in *Soviet Mathematics Doklady*. 1, 7

[Tal14] Avishay Tal. Shrinkage of de Morgan formulae by spectral techniques. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 551–560, 2014. 7

[Tam16] Suguru Tamaki. A satisfiability algorithm for depth two circuits with a sub-quadratic number of symmetric and threshold gates. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:100, 2016. 4

[Vad12] Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012. 39

[Wil13] Ryan Williams. Improving exhaustive search implies superpolynomial lower bounds. *SIAM J. Comput.*, 42(3):1218–1244, 2013. 4

[Wil14] Ryan Williams. Nonuniform ACC circuit lower bounds. *J. ACM*, 61(1):2:1–2:32, 2014. 3, 4

[Win61] Robert O. Winder. Single stage threshold logic. In *Proceedings of the Second Annual Symposium on Switching Circuit Theory and Logical Design (FOCS)*, pages 321–332, Oct 1961. 33

[Yao85] Andrew Chi-Chih Yao. Separating the polynomial-time hierarchy by oracles (preliminary version). In *Proceedings of the Twenty-Sixth Annual Symposium on Foundations of Computer Science (FOCS), Portland, Oregon, USA, 21-23 October 1985*, pages 1–10, 1985. 1, 3, 7

[Yao90] Andrew Chi-Chih Yao. On ACC and threshold circuits. In *Proceedings of the Thirty-First Annual Symposium on Foundations of Computer Science (FOCS), St. Louis, Missouri, USA, October 22-24, 1990, Volume II*, pages 619–627, 1990. 1, 3