Electronic Colloquium on Computational Complexity, Report No. 171 (2013)

Deterministic Approximate Counting for Juntas of Degree-2 Polynomial Threshold Functions

Anindya De*
Institute for Advanced Study

Ilias Diakonikolas†
University of Edinburgh

Rocco A. Servedio‡
Columbia University

Abstract

Let $g : \{-1,1\}^k \to \{-1,1\}$ be any Boolean function and $q_1, \ldots, q_k$ be any degree-2 polynomials over $\{-1,1\}^n$. We give a *deterministic* algorithm which, given as input explicit descriptions of $g, q_1, \ldots, q_k$ and an accuracy parameter $\epsilon > 0$, approximates

$$\mathbf{Pr}_{x\sim\{-1,1\}^n}[g(\mathrm{sign}(q_1(x)), \ldots, \mathrm{sign}(q_k(x))) = 1]$$

to within an additive $\pm\epsilon$. For any constant $\epsilon > 0$ and $k \geq 1$ the running time of our algorithm is a fixed polynomial in n (in fact this is true even for some not-too-small $\epsilon = o_n(1)$ and not-too-large $k = \omega_n(1)$). This is the first fixed polynomial-time algorithm that can deterministically approximately count satisfying assignments of a natural class of depth-3 Boolean circuits.

Our algorithm extends a recent result [DDS13] which gave a deterministic approximate counting algorithm for a single degree-2 polynomial threshold function $\mathrm{sign}(q(x))$, corresponding to the $k = 1$ case of our result. Note that even in the $k = 1$ case it is NP-hard to determine whether $\mathbf{Pr}_{x\sim\{-1,1\}^n}[\mathrm{sign}(q(x)) = 1]$ is nonzero, so any sort of multiplicative approximation is almost certainly impossible even for efficient randomized algorithms.

Our algorithm and analysis requires several novel technical ingredients that go significantly beyond the tools required to handle the $k = 1$ case in [DDS13]. One of these is a new multidimensional central limit theorem for degree-2 polynomials in Gaussian random variables which builds on recent Malliavin-calculus-based results from probability theory. We use this CLT as the basis of a new decomposition technique for k-tuples of degree-2 Gaussian polynomials and thus obtain an efficient deterministic approximate counting algorithm for the Gaussian distribution, i.e., an algorithm for estimating

$$\mathbf{Pr}_{x\sim N(0,1)^n}[g(\mathrm{sign}(q_1(x)), \ldots, \mathrm{sign}(q_k(x))) = 1].$$

Finally, a third new ingredient is a "regularity lemma" for *k-tuples* of degree-d polynomial threshold functions. This generalizes both the regularity lemmas of [DSTW10, HKM09] (which apply to a single degree-d polynomial threshold function) and the regularity lemma of Gopalan et al [GOWZ10] (which applies to a k-tuples of *linear* threshold functions, i.e., the case $d = 1$). Our new regularity lemma lets us extend our deterministic approximate counting results from the Gaussian to the Boolean domain.

*anindya@math.ias.edu. Research supported by Umesh Vazirani's Templeton Foundation Grant 21674.
†ilias.d@ed.ac.uk. Supported in part by a SICSA PECE grant and a Carnegie research grant.
‡rocco@cs.columbia.edu. Supported by NSF grant CCF-1115703.

ISSN 1433-8092

1 Introduction

Unconditional derandomization has been an important research area in computational complexity theory over the past two decades [AW85, Nis91, Nis92, NW94]. A major research goal in this area is to obtain efficient deterministic approximate counting algorithms for "low-level" complexity classes such as constant depth circuits, small space branching programs, polynomial threshold functions, and others [LVW93, LV96, Tre04, GMR13, Vio09, GKM$^+$11, DDS13]. Under the widely-believed hypothesis $\mathbf{P} = \mathbf{BPP}$, there must be a polynomial time deterministic algorithm that can approximate the fraction of satisfying assignments to any polynomial–size circuit. Since finding such an algorithm seems to be out of reach of present day complexity theory [KI02], research efforts have been directed to the aforementioned low-level classes.

A natural class of Boolean functions to consider in this context is the class of polynomial threshold functions (PTFs). Recall that a degree-d PTF, $d \geq 1$, is a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ defined by $f(x) = \mathrm{sign}(p(x))$ where $p : \{-1,1\}^n \to \mathbb{R}$ is a degree-d polynomial over the reals and $\mathrm{sign} : \mathbb{R} \to \{-1,1\}$ is defined as $\mathrm{sign}(z) = 1$ iff $z \geq 0$. In the special case where $d = 1$, degree-d PTFs are often referred to as *linear threshold functions* (LTFs). Understanding the structure of these functions has been a topic of extensive investigation for decades (see e.g., [MK61, MTT61, MP68, Mur71, GHR92, Orp92, Hås94, Pod09] and many other works) due to their importance in fields such as concrete complexity theory [She08, She09, DHK$^+$10, Kan10, Kan12b, Kan12a, KRS12], learning theory [KKMS08, SSSS11, DOSW11, DDFS12], voting theory [APL07, DDS12], and others.

In the context of approximate counting, there is a significant gap in our understanding of low-degree PTFs. An outstanding open problem is to design a deterministic algorithm that approximates the fraction of satisfying assignments to a constant degree PTF over $\{-1,1\}^n$ to an additive $\pm\epsilon$ and runs in time $\mathrm{poly}(n/\epsilon)$. Even for the class of degree-2 PTFs, until recently no deterministic algorithm was known with running time $\mathrm{poly}(n)$ for any sub-constant value of the error ϵ. In previous work [DDS13] we obtained such an algorithm. In the present paper we make further progress on this problem by developing the first efficient deterministic counting algorithm for the class of *juntas of (any constant number of) degree-2 PTFs*.

1.1 Our main result. As our main result, we give a polynomial-time deterministic approximate counting algorithm for any Boolean function of constantly many degree-2 polynomial threshold functions.

Theorem 1. *[Deterministic approximate counting of functions of degree-2 PTFs over $\{-1,1\}^n$] There is an algorithm with the following properties: given an arbitrary function $g : \{-1,1\}^k \to \{-1,1\}$ and k degree-2 polynomials $q_1(x_1,\ldots,x_n),\ldots,q_k(x_1,\ldots,x_n)$ and an accuracy parameter $\epsilon > 0$, the algorithm runs (deterministically) in time $\mathrm{poly}(n) \cdot 2^{(1/\epsilon)^{2^{O(k)}}}$ and outputs a value $v \in [0,1]$ such that*

$$\left|\mathbf{Pr}_{x\in\{-1,1\}^n}[g(\mathrm{sign}(q_1(x)),\ldots,\mathrm{sign}(q_k(x))) = 1] - v\right| \leq \epsilon.$$

Our result may be (somewhat informally) restated in terms of Boolean circuits as a $\mathrm{poly}(n)$-time deterministic approximate counting algorithm for the class NC^0-Thr-AND_2 of depth-3 circuits that have an arbitrary NC^0 gate (i.e., junta) at the top level, arbitrary weighted threshold gates at the middle level, and fanin-2 AND gates at the bottom level. Theorem 1 is a broad generalization of the main result of [DDS13], which establishes the special $k = 1$ case of the current result.

As noted in [DDS13], the problem of determining whether $\mathbf{Pr}_{x\in\{-1,1\}^n}[p(x) \geq 0]$ is nonzero for a degree-2 polynomial p is well known to be NP-hard, and hence no efficient algorithm, even allowing randomness, can give a multiplicative approximation to $\mathbf{Pr}_{x\sim\{-1,1\}^n}[p(x) \geq 0]$ unless NP $\subseteq$ RP. Given this, it is natural to work towards an additive approximation, which is what we achieve.

Previous work. For $k = 1$ and $d = 1$ Gopalan *et al.* in [GKM$^+$11] obtained a multiplicatively $(1 \pm \epsilon)$-accurate deterministic $\mathrm{poly}(n, 1/\epsilon)$ time approximate counting algorithm. For $d \geq 2$, however, as noted above additive approximation is the best one can hope for. For the special case of $k = 1$, in separate

work [DDS13], the authors have given a deterministic approximate counting algorithm that runs in time $\mathrm{poly}(n, 2^{\mathrm{poly}(1/\epsilon)})$. As we explain in detail in the rest of this introduction, more sophisticated ideas and techniques are required to obtain the results of the current paper for general k. These include a new central limit theorem based on Malliavin calculus and Stein's method, and a new decomposition procedure that goes well beyond the decomposition approach employed in [DDS13].

We remark that the only previous deterministic approximate counting algorithm for k-juntas of degree-2 PTFs follows from the *pseudorandom generators* (PRGs) of [DKN10] (which are based on bounded independence). The running time of the resulting algorithm is $n^{\mathrm{poly}(1/\epsilon)}$, even for $k = 1$.

1.2 Techniques. Our high-level approach to establishing Theorem 1 follows a by now standard approach in this area. We first (i) establish the result for general polynomials over Gaussian inputs; then (ii) use a "regularity lemma" to show that every polynomial over Boolean inputs can be decomposed into a "small" number of regular polynomials over Boolean inputs; and finally (iii) use an invariance principle to reduce the problem for "regular" polynomials over Boolean inputs to the problem for regular polynomials over Gaussian inputs. This general approach has been used in a number of previous works, including constructions of unconditional PRGs [DGJ$^+$10, MZ10, GOWZ10, DKN10, Kan11, Kan12b], learning and property testing [MORS10, OS11], and other works. However, we emphasize that significant novel conceptual and technical work is required to make this approach work in our setting. More specifically, to achieve step (i), we require (i.a) a new multidimensional CLT for degree-2 Gaussian polynomials with small eigenvalues and (i.b) a new decomposition procedure that transforms a k-dimensional vector of Gaussian polynomials into a tractable form for the purpose of approximate counting. For step (ii) we establish a novel regularity lemma for k-vectors of low-degree polynomials. Finally, Step (iii) follows by an application of the invariance principle of Mossel [Mos10] combined with appropriate mollification arguments [DKN10]. In the rest of this section we discuss our new approaches to Steps (i) and (ii).

Step (i): The counting problem over Gaussian inputs. The current paper goes significantly beyond the techniques of [DDS13]. To explain our new contributions let us first briefly recall the [DDS13] approach.

The main observation enabling the result in [DDS13] is this: Because of rotational symmetry of the Gaussian distribution, a degree-2 Gaussian polynomial can be "diagonalized" so that there exist no "cross-terms" in its representation. In a little more detail, if $p(x) = \sum_{i,j} a_{ij} x_i x_j$ (we ignore the linear term for simplicity), where $x \sim N(0,1)^n$, then p can be rewritten in the form $p(y) = \sum_i \lambda_i y_i^2$, where $y \sim N(0,1)^n$ and the λ_i's are the eigenvalues of the corresponding matrix. Roughly speaking, once such a representation has been (approximately) constructed, the counting problem can be solved efficiently by dynamic programming. To construct such a decomposition, [DDS13] employs a "critical-index" based analysis on the eigenvalues of the corresponding matrix. For the analysis of the [DDS13] algorithm, [DDS13] proves a CLT for a single degree-2 Gaussian polynomial with small eigenvalues (this CLT is based on a result of Chaterjee [Cha09]). (We note that this informal description suppresses several non-trivial technical issues, see [DDS13] for details.)

At a high level, the approach of the current paper builds on the approach of [DDS13]. To solve the Gaussian counting problem we use a combination of (i.a) a new multidimensional CLT for k-tuples of degree-2 Gaussian polynomials with small eigenvalues, and (i.b) a novel decomposition result for k-tuples of degree-2 Gaussian polynomials. We now elaborate on these steps.

(i.a) As our first contribution, we prove a new multidimensional central limit theorem for k-tuples of degree-2 Gaussian polynomials (Theorem 8). Roughly speaking, our CLT states that if each polynomial in the k-tuple has small eigenvalues, then the joint distribution of the k-tuple is close to a k-dimensional Gaussian random variable with matching mean and covariance matrix. The closeness here is with respect to the k-dimensional Kolmogorov distance over $\mathbb{R}^k$ (a natural generalization of Kolmogorov distance to vector-valued random variables, which we denote d_{K} and which is useful for

analyzing PTFs). To establish our new CLT, we proceed in two steps: In the first (main) step, we make essential use of a recent multidimensional CLT due to Nourdin and Peccati [NP09] (Theorem 11) which is proved using a combination of Malliavin calculus and Stein's method. To use this theorem in our setting, we perform a linear-algebraic analysis which allows us to obtain precise bounds on the Malliavin derivatives of degree-2 Gaussian polynomials with small eigenvalues. An application of [NP09] then gives us a version of our desired CLT with respect to "test functions" with bounded second derivatives (Theorem 12). In the second step, we use tools from mollification [DKN10] to translate this notion of closeness into closeness with respect to k-dimensional Kolmogorov distance, thus obtaining our intended CLT. (As a side note, we believe that this work is the first to use Malliavin-calculus-based tools in the context of derandomization.)

(i.b) As our second contribution, we give an efficient procedure that transforms a k-tuple of degree-2 Gaussian polynomials $p = (p_1, \ldots, p_k)$ into a k-tuple of degree-2 Gaussian polynomials $r = (r_1, \ldots, r_k)$ such that: (1) p and r are d_{K}-close, and (2) the k-tuple r has a "nice structure" that allows for efficient deterministic approximate counting. In particular, there is a "small" set of variables such that for each restriction ρ fixing this set, the restricted k-tuple of polynomials $r|_\rho$ is well-approximated by a k-dimensional Gaussian random variable (with the appropriate mean and covariance matrix). Once such an r has been obtained, deterministic approximate counting is straightforward via an appropriate discretization of the k-dimensional Gaussian distribution (see Section 5).

We now elaborate on Item (1) above. At a high level, the main step of our transformation procedure performs a "change of basis" to convert $p = (p_1(x), \ldots, p_k(x))$ into an essentially equivalent (for the purpose of approximate counting) vector $q = (q_1(y), \ldots, q_k(y))$ of polynomials. The high-level approach to achieve this is reminiscent of (and inspired by) the decomposition procedure for vectors of k linear forms in [GOWZ10]. However, there are significant complications that arise in our setting. In particular, in the [GOWZ10] approach, a vector of k linear forms is simplified by "collecting" variables in a greedy fashion as follows: Each of the k linear forms has a "budget" of at most B, meaning that at most B variables will be collected on its behalf. Thus, the overall number of variables that are collected is at most kB. At each stage some variable is collected which has large influence in the remaining (uncollected) portion of some linear form. The [GOWZ10] analysis shows that after at most B variables have been collected on behalf of each linear form, each of the k linear forms will either be regular or its remaining portion (consisting of the uncollected variables) will have small variance. In our current setting, we are dealing with k degree-2 Gaussian polynomials instead of k linear forms. Recall that every degree-2 polynomial can be expressed as a linear combination of squares of linear forms (i.e., it can be diagonalized). Intuitively, since Gaussians are invariant under change of basis, we can attempt to use an approach where linear forms will play the role that variables had in [GOWZ10]. Mimicking the [GOWZ10] strategy, each quadratic polynomial will have at most B linear forms collected on its behalf, and at most kB linear forms will be collected overall. Unfortunately, this vanilla strategy does not work even for $k = 2$, as it requires a single orthonormal basis in which all the degree-2 polynomials are simultaneously diagonalized.

Instead, we resort to a more refined strategy. Starting with the k quadratic polynomials, we use the following iterative algorithm: If the largest magnitude eigenvalue of each quadratic form is small, we are already in the *regular* case (and we can appeal to our multidimensional CLT). Otherwise, there exists at least one polynomial with a large magnitude eigenvalue. We proceed to collect the corresponding linear form and "reduce" every polynomial by this linear form. (The exact description of this reduction is somewhat involved to describe, but intuitively, it uses the fact that Gaussians are invariant under orthogonal transformations.) This step is repeated iteratively; an argument similar to [GOWZ10] shows that for every quadratic polynomial, we can collect at most B linear forms. At the

end of this procedure, each of the k quadratic polynomials will either be "regular" (have small largest magnitude eigenvalue compared to the variance of the remaining portion), or else the variance of the remaining portion will be small. This completes the informal description of our transformation.

Our main result for the Gaussian setting is the following theorem:

Theorem 2. *[Deterministic approximate counting of functions of degree-2 PTFs over Gaussians] There is an algorithm with the following properties: It takes as input explicit descriptions of n-variable degree-2 polynomials $q_1, \ldots, q_k$, an explicit description of a k-bit Boolean function $g : \{-1,1\}^k \to \{-1,1\}$, and a value $\epsilon > 0$. It runs (deterministically) in time $\mathrm{poly}(n) \cdot 2^{\mathrm{poly}(2^k/\epsilon)}$ and outputs a value $\tilde{v} \in [0,1]$ such that*

$$\left|\mathbf{Pr}_{\mathcal{G}\sim N(0,1)^n}[g(Q_1(\mathcal{G}), \ldots, Q_k(\mathcal{G})) = 1] - \tilde{v}\right| \leq \epsilon, \tag{1}$$

where $Q_i(x) = \mathrm{sign}(q_i(x))$ for $i = 1, \ldots, k$.

We note that in the case $k = 1$, the algorithm of the current work is not the same as the algorithm of [DDS13] (indeed, observe the above algorithm runs in time exponential in $1/\epsilon$ even for $k = 1$, whereas the algorithm of [DDS13] runs in time $\mathrm{poly}(n/\epsilon)$ for a single Gaussian polynomial).

Step (ii): The regularity lemma. Recall that the *influence* of variable i on a multilinear polynomial $p = \sum_{S\subseteq[n]} \widehat{p}(S) \prod_{i\in S} x_i$ over $\{-1,1\}^n$ (under the uniform distribution) is $\mathrm{Inf}_i(p) \stackrel{\text{def}}{=} \sum_{S\ni i} \widehat{p}(S)^2$ and that the *variance* of p is $\mathrm{Var}[p] = \mathbf{E}_{x\in\{-1,1\}^n}[(p(x) - \mathbf{E}[p])^2] = \sum_{\emptyset\neq S} \widehat{p}^2(S)$. For p a degree-d polynomial we have $\mathrm{Var}[p] \leq \sum_{i=1}^n \mathrm{Inf}_i(p) \leq d \cdot \mathrm{Var}[p]$, so for small constant d the variance and the total influence $\sum_{i=1}^n \mathrm{Inf}_i(d)$ are equal up to a small constant factor. A polynomial p is said to be τ-*regular* if for all $i \in [n]$ we have $\mathrm{Inf}_i(p) \leq \tau \cdot \mathrm{Var}[p]$.

As noted earlier, by adapting known invariance principles from the literature [Mos08] it is possible to show that an algorithm for approximately counting satisfying assignments of a junta of degree-2 PTFs over $N(0,1)^n$ will in fact also succeed for approximately counting satisfying assignments of a junta of sufficiently regular degree-2 PTFs over $\{-1,1\}^n$. Since Theorem 2 gives us an algorithm for the Gaussian problem, to complete the chain we need a reduction from the problem of counting satisfying assignments of a junta of *arbitrary* degree-2 PTFs over $\{-1,1\}^n$, to the problem of counting satisfying assignments of a junta of *regular* degree-2 PTFs over $\{-1,1\}^n$.

We accomplish this by giving a novel *regularity lemma* for k-tuples of degree-2 (or more generally, degree-d) polynomials. Informally speaking, this is an efficient deterministic algorithm with the following property: given as input a k-tuple of arbitrary degree-2 polynomials $(p_1, \ldots, p_k)$ over $\{-1,1\}^n$, it constructs a decision tree of restrictions such that for almost every root-to-leaf path (i.e., restriction ρ) in the decision tree, *all* k restricted polynomials $(p_1)_\rho, \ldots, (p_k)_\rho$ are "easy to handle" for deterministic approximate counting, in the following sense: each $(p_i)_\rho$ is either highly regular, or else is highly *skewed*, in the sense that its constant term is so large compared to its variance that the corresponding PTF $\mathrm{sign}((p_i)_\rho)$ is guaranteed to be very close to a constant function. Such leaves are "easy to handle" because we can set the PTFs corresponding to "skewed" polynomials to constants (and incur only small error); then we are left with a junta of regular degree-2 PTFs, which can be handled using the Gaussian algorithm as sketched above.

A range of related "regularity lemmas" have been given in the LTF/PTF literature [DSTW10, HKM09, BELY09, GOWZ10], but none with all the properties that we require. [Ser07] implicitly gave a regularity lemma for a single LTF, and [DSTW10, HKM09, BELY09] each gave (slightly different flavors of) regularity lemmas for a single degree-d PTF. Subsequently [GOWZ10] gave a regularity lemma for k-tuples of LTFs; as noted earlier our decomposition for k-tuples of degree-2 polynomials over Gaussian inputs given in Section 5 uses ideas from their work. However, as we describe in Section 7, their approach does not seem to extend to degrees $d > 1$, so we must use a different approach to prove our regularity lemma.

1.3 Organization. After giving some useful background in Section 2, we prove our new multidimensional CLT in Section 3. We give the transformation procedure that is at the heart of our decomposition approach in Section 4, and present the actual deterministic counting algorithm for the Gaussian case that uses this transformation in Section 5. Section 6 shows how the new regularity lemma for k-tuples of Boolean PTFs gives the main Boolean counting result, and finally the regularity lemma is proved in Section 7.

2 Definitions, Notation and Useful Background

Polynomials and PTFs. Throughout the paper we use lower-case letters p, q, etc. to denote low-degree multivariate polynomials. We use capital letters to denote the corresponding polynomial threshold functions that map to $\{-1, 1\}$, so typically $P(x) = \mathrm{sign}(p(x))$, $Q(x) = \mathrm{sign}(q(x))$, etc.

We consider multivariate polynomials over the domains $\mathbb{R}^n$ (endowed with the standard normal distribution $N(0,1)^n$) and $\{-1,1\}^n$ (endowed with the uniform distribution). Since $x^2 = 1$ for $x \in \{-1, 1\}$, in dealing with polynomials over the domain $\{-1,1\}^n$ we may without loss of generality restrict our attention to multilinear polynomials.

Kolmogorov distance between $\mathbb{R}^k$-valued random variables. It will be convenient for us to use a natural k-dimensional generalization of the Kolmogorov distance between two real-valued random variables which we now describe. Let $X = (X_1, \ldots, X_k)$ and $Y = (Y_1, \ldots, Y_k)$ be two $\mathbb{R}^k$-valued random variables. We define the *k-dimensional Kolmogorov distance* between X and Y to be

$$d_{\mathrm{K}}(X, Y) = \sup_{(\theta_1,\ldots,\theta_k)\in\mathbb{R}^k} \left|\mathbf{Pr}[\forall\, i \in [k]\ X_i \le \theta_i] - \mathbf{Pr}[\forall\, i \in [k]\ Y_i \le \theta_i]\right|.$$

This will be useful to us when we are analyzing k-juntas of degree-2 PTFs over Gaussian random variables; we will typically have $X = (q_1(x), \ldots, q_k(x))$ where $x \sim N(0,1)^n$ and q_i is a degree-2 polynomial, and have $Y = (Y_1, \ldots, Y_k)$ be a k-dimensional Gaussian random variable whose mean and covariance matrix match those of X.

Notation and terminology for degree-2 polynomials. Let $q = (q_1(x), \ldots, q_k(x))$ be a vector of polynomials over $\mathbb{R}^n$. We endow $\mathbb{R}^n$ with the $N(0,1)^n$ distribution, and hence we may view q as a k-dimensional random variable. We sometimes refer to the q_i's as *Gaussian polynomials*.

For A a real $n \times n$ matrix we write $\|A\|_2$ to denote the operator norm $\|A\|_2 = \max_{0\neq x\in\mathbb{R}^n} \frac{\|Ax\|_2}{\|x\|_2}$.

Given a degree-2 polynomial $q : \mathbb{R}^n \to \mathbb{R}$ defined as $q(x) = \sum_{1\le i\le j\le n} a_{ij}x_ix_j + \sum_{1\le i\le n} b_ix_i + C$, we define the (symmetric) matrix A corresponding to its quadratic part as : $A_{ij} = a_{ij}(1/2 + \delta_{ij}/2)$. Note that with this definition we have that $x^T \cdot A \cdot x = \sum_{1\le i\le j\le n} a_{ij}x_ix_j$ for the vector $x = (x_1, \ldots, x_n)$.

Throughout the paper we adopt the convention that the eigenvalues $\lambda_1, \ldots, \lambda_n$ of a real symmetric matrix A satisfy $|\lambda_1| \ge \cdots \ge |\lambda_n|$. We sometimes write $\lambda_{\max}(A)$ to denote λ_1, and we sometimes write $\lambda_i(q)$ to refer to the i-th eigenvalue of the matrix A defined based on q as described above.

Degree-2 polynomials and their heads and tails. The following notation will be useful for us, especially in Section 4. Let $z(y_1, \ldots, y_n) = \sum_{1\le i\le j\le n} a_{ij}y_iy_j + \sum_{1\le i\le n} b_iy_i + c$ be a degree-2 polynomial. For $0 \le t \le n$ we say the *t-head of $z(y)$*, denoted $\mathrm{Head}_t(z(y))$, is the polynomial

$$\mathrm{Head}_t(z(y)) \stackrel{\text{def}}{=} \sum_{1\le i\le t, j\ge i} a_{ij}y_iy_j + \sum_{1\le i\le t} b_iy_i \tag{2}$$

and the *t-tail of $z(y)$*, denoted $\mathrm{Tail}_t(z(y))$, is the polynomial

$$\mathrm{Tail}_t(z(y)) \stackrel{\text{def}}{=} \sum_{t<i\le j\le n} a_{ij}y_iy_j + \sum_{t<i\le n} b_iy_i + c, \tag{3}$$

so clearly we have $z(y) = \mathrm{Head}_t(z(y)) + \mathrm{Tail}_t(z(y))$. (Intuitively, $\mathrm{Tail}_t(z(y))$ is the part of $z(y)$ which does not "touch" any of the first t variables $y_1, \ldots, y_t$ and $\mathrm{Head}_t(z(y))$ is the part which does "touch" those variables.)

Remark 3. *Note that if $\rho = (\rho_1, \ldots, \rho_t) \in \mathbb{R}^t$ is a restriction fixing variables $y_1, \ldots, y_t$, then the restricted polynomial $z|_\rho(y) \stackrel{def}{=} z(\rho_1, \ldots, \rho_t, y_{t+1}, \ldots, y_n)$ is of the form $\mathrm{Tail}_t(z(y)) + L(y_{t+1}, \ldots, y_n)$ where L is an affine form.*

We further define $\mathrm{QuadTail}_t(z(y))$, the "quadratic portion of the t-tail," to be

$$\mathrm{QuadTail}_t(z(y)) \stackrel{\text{def}}{=} \sum_{t<i\le j\le n} a_{ij}y_iy_j. \tag{4}$$

Setting aside heads and tails, it will sometimes be useful for us to consider the sum of squares of all the (non-constant) coefficients of a degree-2 polynomial. Towards that end we have the following definition:

Definition 4. *Given $p : \mathbb{R}^n \to \mathbb{R}$ defined by $p(x) = \sum_{1\le i\le j\le n} a_{ij}x_ix_j + \sum_{1\le i\le n} b_ix_i + C$, define $SS(p)$ as $SS(p) = \sum_{1\le i\le j\le n} a_{ij}^2 + \sum_{1\le i\le n} b_i^2$.*

The following straightforward claim is established in [DDS13]:

Claim 5. *[Claim 20 of [DDS13]] Given $p : \mathbb{R}^n \to \mathbb{R}$, we have that $2\,SS(p) \ge \mathrm{Var}(p) \ge SS(p)$.*

Tail bounds and anti-concentration bounds on low-degree polynomials in Gaussian variables. We will need the following standard concentration bound for low-degree polynomials over independent Gaussians.

Theorem 6 ("degree-d Chernoff bound", [Jan97])**.** *Let $p : \mathbb{R}^n \to \mathbb{R}$ be a degree-d polynomial. For any $t > e^d$, we have*

$$\mathbf{Pr}_{x\sim N(0,1)^n}[|p(x) - \mathbf{E}[p(x)]| > t \cdot \sqrt{\mathrm{Var}(p(x))}] \le de^{-\Omega(t^{2/d})}.$$

We will also use the following anti-concentration bound for degree-d polynomials over Gaussians:

Theorem 7 ([CW01])**.** *Let $p : \mathbb{R}^n \to \mathbb{R}$ be a degree-d polynomial that is not identically 0. Then for all $\epsilon > 0$ and all $\theta \in \mathbb{R}$, we have*

$$\mathbf{Pr}_{x\sim N(0,1)^n}\left[|p(x) - \theta| < \epsilon\sqrt{\mathrm{Var}(p)}\right] \le O(d\epsilon^{1/d}).$$

The model. Throughout this paper, our algorithms will repeatedly be performing basic linear algebraic operations, in particular SVD computation and Gram-Schmidt orthogonalization. In the bit complexity model, it is well-known that these linear algebraic operations can be performed (by deterministic algorithms) up to additive error ϵ in time $\mathrm{poly}(n, 1/\epsilon)$. For example, let $A \in \mathbb{R}^{n\times m}$ have b-bit rational entries. It is known (see [GL96] for details) that in time $\mathrm{poly}(n, m, b, 1/\epsilon)$, it is possible to compute a value $\tilde{\sigma}_1$ and vectors $u_1 \in \mathbb{R}^n$, $v_1 \in \mathbb{R}^m$, such that $\tilde{\sigma}_1 = \frac{u_1^T A v_1}{\|u_1\|\|v_1\|}$ and $|\tilde{\sigma}_1 - \sigma_1| \le \epsilon$, where σ_1 is the largest singular value of A. Likewise, given n linearly independent vectors $v^{(1)}, \ldots, v^{(n)} \in \mathbb{R}^m$ with b-bit rational entries, it is possible to compute vectors $\tilde{u}^{(1)}, \ldots, \tilde{u}^{(n)}$ in time $\mathrm{poly}(n, m, b)$ such that if $u^{(1)}, \ldots, u^{(n)}$ is a Gram-Schmidt orthogonalization of $v^{(1)}, \ldots, v^{(n)}$ then we have $|u^{(i)} \cdot u^{(j)} - \tilde{u}^{(i)} \cdot \tilde{u}^{(j)}| \le 2^{-\mathrm{poly}(b)}$ for all i, j.

In this paper, we work in a unit-cost real number model of computation. This allows us to assume that given a real matrix $A \in \mathbb{R}^{n\times m}$ with b-bit rational entries, we can compute the SVD of A exactly in time $\mathrm{poly}(n, m, b)$. Likewise, given n vectors over $\mathbb{R}^m$, each of whose entries are b-bit rational numbers, we can perform an exact Gram-Schmidt orthogonalization in time $\mathrm{poly}(n, m, b)$. Using high-accuracy approximations of the sort described above throughout our algorithms, it is straightforward to translate our unit-cost real-number algorithms into the bit complexity setting, at the cost of some additional error in the resulting bound.

Using these two observations, it can be shown that by making sufficiently accurate approximations at each stage where a numerical computation is performed by our "idealized" algorithm, the cumulative error resulting from all of the approximations can be absorbed into the final $O(\epsilon)$ error bound. Since inverse polynomial levels of error can be achieved in polynomial time for all of the approximate numerical computations that our algorithm performs, and since only $\mathrm{poly}(n)$ many such approximation steps are performed by $\mathrm{poly}(n)$-time algorithms, the resulting approximate implementations of our algorithms in a bit-complexity model also achieve the guarantee of our main results, at the cost of a fixed $\mathrm{poly}(n)$ overhead in the running time. For the sake of completeness, such a detailed numerical analysis was performed in our previous paper [DDS13]. Since working through the details of such an analysis is tedious and detracts from the clarity of the presentation, we content ourselves with this brief discussion in this work.

3 A multidimensional CLT for degree-2 Gaussian polynomials

In this section we prove a central limit theorem which plays a crucial role in the decomposition result which we establish in the following sections. Let $q = (q_1, \ldots, q_k)$ where each q_i is a degree-2 polynomial in Gaussian random variables $(x_1, \ldots, x_n) \sim N(0, 1)^n$. Our CLT states that under suitable conditions on $q_1, \ldots, q_k$ — all of them have only small–magnitude eigenvalues, no $\mathrm{Var}[q_i]$ is too large and at least one $\mathrm{Var}[q_i]$ is not too small — the distribution of q is close (in k-dimensional Kolmogorov distance) to the distribution of the k-dimensional Gaussian random variable whose mean and covariance matrix match q.

Theorem 8. *Let $q = (q_1(x), \ldots, q_k(x))$ where each q_i is a degree-2 Gaussian polynomial that satisfies $\mathrm{Var}[q_i] \leq 1$ and $|\lambda_{\max}(q_i)| \leq \epsilon$ for all $i \in [k]$. Suppose that $\max_{i\in[k]} \mathrm{Var}(q_i) \geq \lambda$. Let C denote the covariance matrix of q and let $N = N((\mu_1, \ldots, \mu_k), C)$ be a k-dimensional Gaussian random variable with covariance matrix C and mean $(\mu_1, \ldots, \mu_k)$ where $\mu_i = \mathbf{E}[q_i]$. Then*

$$d_{\mathrm{K}}(q, N) \leq O\left(\frac{k^{2/3}\epsilon^{1/6}}{\lambda^{1/6}}\right).$$

Looking ahead to motivate this result for our ultimate purposes, Theorem 8 is useful for deterministic approximate counting because if $q = (q_1, \ldots, q_k)$ satisfies the conditions of the theorem, then the theorem ensures that $\mathbf{Pr}_{x\sim N(0,1)^n}[\forall \ell \in [k], q_\ell(x) \leq 0]$ is close to $\mathbf{Pr}[\forall \ell \in [k], N_\ell \leq 0]$. Note that the latter quantity can be efficiently estimated by a deterministic algorithm.

A key ingredient in the proof of Theorem 8 is a CLT due to Nourdin and Peccati [NP09] which gives a bound that involves the Malliavin derivative of the functions $q_1, \ldots, q_k$. In Section 3.1we give the necessary background from Malliavin calculus and build on the [NP09] result to prove a result which is similar to Theorem 8 but gives a bound on $\mathbf{E}[h(q)] - \mathbf{E}[h(N)]$ rather than $d_{\mathrm{K}}(q, N)$ for a broad class of "test functions" h (see Theorem 12 below). In Section 3.2we show how Theorem 12can be combined with standard "mollification" techniques to yield Theorem 8.

3.1 Malliavin calculus and test functions with bounded second derivative. We need some notation and conceptual background before we can state the Nourdin-Peccati multi-dimensional CLT from [NP09]. Their CLT is proved using Stein's method; while there is a rich theory underlying their result we give only the

absolute basics that suffice for our purposes. (See e.g. [NP09, Nou12] for detailed treatments of Malliavin calculus and its interaction with Stein's Method.)

We will use $\mathcal{X}$ to denote the space $\mathbb{R}^n$ endowed with the standard $N(0,1)^n$ normal measure and $\mathcal{P}$ to denote the family of all polynomials over $\mathcal{X}$. For integer $d \geq 0$ we let $\mathcal{H}_d$ denote the "d-th Wiener chaos" of $\mathcal{X}$, namely the space of all homogeneous degree-d Hermite polynomials over $\mathcal{X}$. We define the operator $I_d : \mathcal{P} \to \mathcal{H}_d$ as follows : I_d maps $p \in \mathcal{P}$ to the degree-d part of its Hermite expansion, so if p has degree d then $p = I_0(p) + \cdots + I_d(p)$.

We next define the generator of the Ornstein-Uhlenbeck semigroup. This is the operator L which is defined on $\mathcal{P}$ via

$$Lp = \sum_{q=0}^{\infty} -q \cdot I_q(p).$$

It is easy to see that for $p \in \mathcal{P}$ we have the inverse operator

$$L^{-1}p = \sum_{q=1}^{\infty} \frac{-1}{q} I_q(p).$$

Next we introduce the notion of the *Malliavin derivative*. The Malliavin derivative operator D maps a real-valued random variable (defined over $\mathcal{X}$ by a differentiable real-valued function $f : \mathbb{R}^n \to \mathbb{R}$) to an n-dimensional vector of random variables in the following way: for $f : \mathbb{R}^n \to \mathbb{R}$,

$$Df = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right).$$

The following key identity provides the fundamental connection between Malliavin Calculus and Stein's method, which is used to prove Theorem 11 below:

Claim 9 (see e.g. Equation (2.22) of [NP09])**.** *Let $h : \mathbb{R} \to \mathbb{R}$ be a continuous function with a bounded first derivative. Let p and q be polynomials over $\mathcal{X}$ with $\mathbf{E}[q] = 0$. Then $\mathbf{E}[qh(p)] = \mathbf{E}[h'(p) \cdot \langle Dp\ ,\ -DL^{-1}q \rangle]$.*

Specializing to the case $h(x) = x$, we have

Corollary 10. *Let p and q be finite degree polynomials over $\mathcal{X}$ with $\mathbf{E}[q] = 0$. Then, $\mathbf{E}[qp] = \mathbf{E}[\langle Dp\ ,\ -DL^{-1}q \rangle]$.*

We now recall the following CLT due to Nourdin and Peccati:

Theorem 11. *[[NP09], see also [Nou12], Theorem 6.1] Let $p = (p_1, \ldots, p_k)$ where each p_i is a Gaussian polynomial with $\mathbf{E}[p_i] = 0$. Let C be a symmetric PSD matrix in $\mathbb{R}^{k \times k}$ and let N be a mean-0 k-dimensional Gaussian random variable with covariance matrix C. Then for any $h : \mathbb{R}^k \to \mathbb{R}, h \in \mathcal{C}^2$ such that $\|h''\|_\infty < \infty$, we have*

$$|\mathbf{E}[h(p)] - \mathbf{E}[h(N)]| < \frac{1}{2}\|h''\|_\infty \cdot \left(\sum_{i=1}^{k} \sum_{j=1}^{k} \mathbf{E}[|C(i,j) - Y(i,j)|] \right)$$

where $Y(i,j) = \langle Dp_i\ , -DL^{-1}p_j \rangle$.

We now use Theorem 11 to prove our main result of this subsection, which is the following CLT for multidimensional degree-2 Gaussian polynomials with small-magnitude eigenvalues. Our CLT says that such multidimensional random variables must in fact be close to multidimensional Gaussian distributions, where "closeness" here is measured using test functions with bounded second derivative. (In the next subsection we extend this result using mollification techniques to obtain Theorem 8, which uses multidimensional Kolmogorov distance.)

Theorem 12. *Let $q = (q_1, \ldots, q_k)$ where each q_i is a degree-2 mean-0 Gaussian polynomial with $\mathrm{Var}[q_i] \leq 1$ and $|\lambda_{\max}(q_i)| \leq \epsilon$. Let C denote the covariance matrix of q, so $C(i,j) = \mathrm{Cov}(q_i, q_j) = \mathbf{E}[q_i q_j]$. Let N be a mean-zero k-dimensional Gaussian random variable with covariance matrix C. Then for any $h : \mathbb{R}^k \to \mathbb{R}, h \in \mathcal{C}^2$ such that $\|h''\|_\infty < \infty$, we have*

$$|\mathbf{E}[h(q)] - \mathbf{E}[h(N)]| < O(k^2\epsilon) \cdot \|h''\|_\infty.$$

Proof. As in Theorem 11, we write $Y(a,b)$ to denote $\langle Dq_a, -DL^{-1}q_b\rangle$. For any $1 \leq a, b \leq k$, we have

$$C(a,b) = \mathrm{Cov}(q_a, q_b) = \mathbf{E}[q_a q_b] = \mathbf{E}[Y(a,b)], \tag{5}$$

where the second equality is because q_a and q_b have mean 0 and the third equality is by Corollary 10. Since C is a covariance matrix and every covariance matrix is PSD, we may apply Theorem 11, and we get that

$$|\mathbf{E}[h(q)] - \mathbf{E}[h(N)]| < \frac{k^2}{2}\|h''\|_\infty \cdot \max_{1\leq a,b\leq k} \mathbf{E}[|C(a,b) - Y(a,b)|] = \frac{k^2}{2}\|h''\|_\infty \cdot \max_{1\leq a,b\leq k} \mathbf{E}[|Y(a,b) - \mathbf{E}[Y(a,b)]|],$$

where we used (5) for the equality. By Jensen's inequality we have $\mathbf{E}[|Y(a,b) - \mathbf{E}[Y(a,b)]|] \leq \sqrt{\mathrm{Var}[Y(a,b)]}$. Lemma 13 below gives us that $\mathrm{Var}[Y(a,b)] \leq O(\epsilon^2)$, and the theorem is proved. □

It remains to establish the following lemma:

Lemma 13. *For each $1 \leq a, b \leq k$, we have that $\mathrm{Var}[Y(a,b)] = O(\epsilon^2)$.*

Proof. Fix $1 \leq a, b \leq k$, so $q_a(x_1, \ldots, x_n)$ and $q_b(x_1, \ldots, x_n)$ are degree-2 Gaussian polynomials with mean 0. Recalling the spherical symmetry of the $N(0,1)^n$ distribution, by a suitable choice of basis that diagonalizes q_a we may write

$$q_a(x) = \sum_{i=1}^n \lambda_i x_i^2 + \sum_{i=1}^n \beta_i x_i + \gamma \qquad \text{and} \qquad q_b(x) = \sum_{i,j=1}^n \delta_{ij} x_i x_j + \sum_{i=1}^n \kappa_i x_i + \rho,$$

where we take $\delta_{ij} = \delta_{ji}$ for all $1 \leq i, j \leq k$.

Recalling that $Y(a,b) = \langle Dq_a, -DL^{-1}q_b\rangle$, we start by observing that $Dq_a = (2\lambda_\ell x_\ell + \beta_\ell)_{\ell=1,\ldots,n}$. For $-DL^{-1}q_b$, we have that $L^{-1}q_b = -I_1(q_b) - (1/2)I_2(q_b)$. We have $I_1(q_b) = \sum_{i=1}^n \kappa_i x_i$. Recalling that the first two normalized Hermite polynomials are $h_1(x) = x$ and $h_2(x) = (x^2-1)/\sqrt{2}$, it is straightforward to verify that $I_2(q_b)$ (the homogeneous degree-2 part of the Hermite expansion of q_b) is

$$I_2(q_b) = \sum_{1\leq i\neq j\leq k} \delta_{ij} h_1(x_i) h_1(x_j) + \sum_{i=1}^n \sqrt{2} \cdot \delta_{ii} h_2(x_i).$$

Hence

$$L^{-1}q_b = -\sum_{i=1}^n \kappa_i x_i - \frac{1}{2}\sum_{1\leq i\neq j\leq k} \delta_{ij} x_i x_j - \frac{1}{2}\sum_{i=1}^n \delta_{ii}(x_i^2 - 1),$$

so

$$-DL^{-1}q_b = \left(\kappa_\ell + \sum_{i=1}^n \delta_{i\ell} x_i\right)_{\ell=1,\ldots,n}.$$

We thus can write $Y(a,b)$ as a degree-2 polynomial in the variables $x_1, \ldots, x_n$ as

$$\begin{aligned} Y(a,b) &= \sum_{\ell=1}^n (2\lambda_\ell x_\ell + \beta_\ell) \cdot \left(\kappa_\ell + \sum_{i=1}^n \delta_{i\ell} x_i\right) \\ &= \sum_{i=1}^n \sum_{\ell=1}^n 2\lambda_\ell \delta_{i\ell} x_i x_\ell + \sum_{\ell=1}^n 2\kappa_\ell \lambda_\ell x_\ell + \sum_{i=1}^n \left(\sum_{\ell=1}^n \beta_\ell \delta_{i\ell}\right) x_i + \sum_{\ell=1}^n \kappa_\ell \beta_\ell. \end{aligned}$$

By Claim 5, we know that $\mathrm{Var}[Y(a,b)] \le SS(Y(a,b))$. Using the inequality $(r+s)^2 \le 2r^2 + 2s^2$ for the degree-1 coefficients, to prove the lemma it suffices to show that

$$\sum_{i=1}^{n}\sum_{\ell=1}^{n}(\lambda_\ell \delta_{i\ell})^2 + \sum_{\ell=1}^{n}(\kappa_\ell\lambda_\ell)^2 + \sum_{i=1}^{n}\left(\sum_{\ell=1}^{n}\beta_\ell\delta_{i\ell}\right)^2 \le O(\epsilon^2). \tag{6}$$

We bound each term of (6) in turn. For the first, we recall that each λ_ℓ is an eigenvalue of q_a and hence satisfies $\lambda_\ell^2 \le \epsilon^2$; hence we have

$$\sum_{i=1}^{n}\sum_{\ell=1}^{n}(\lambda_\ell\delta_{i\ell})^2 \le \epsilon^2 \sum_{i=1}^{n}\sum_{\ell=1}^{n}(\delta_{i\ell})^2 \le \epsilon^2,$$

where we have used Claim 5 again to get that $\sum_{i,\ell=1}^{n}(\delta_{i\ell})^2 \le SS(q_b) \le \mathrm{Var}[q_b] \le 1$. For the second term, we have

$$\sum_{\ell=1}^{n}(\kappa_\ell\lambda_\ell)^2 \le \epsilon^2 \cdot \sum_{\ell=1}^{n}\kappa_\ell^2 \le \epsilon^2 \cdot SS(q_b) \le \epsilon^2.$$

Finally, for the third term, let us write $M = (\delta_{i\ell})$ for the $k \times k$ matrix corresponding to the quadratic part of q_b and $\bar{\beta}$ for the column vector whose ℓ-th entry is β_ℓ. Then we have that

$$\sum_{i=1}^{n}\left(\sum_{\ell=1}^{n}\beta_\ell\delta_{i\ell}\right)^2 = \|M\bar{\beta}\|_2^2 \le \|\lambda_{\max}(M)\bar{\beta}\|_2^2 \le \epsilon^2\|\bar{\beta}\|_2 \le \epsilon^2,$$

where the second inequality is because each eigenvalue of p_b has magnitude at most 1 and the third is because $\|\bar{\beta}\|_2^2 \le SS(p_a) \le \mathrm{Var}[p_a] \le 1$. This concludes the proof of Lemma 13. □

3.2 From test functions with bounded second derivative to multidimensional Kolmogorov distance. In this subsection we show how "mollification" arguments can be used to extend Theorem 12 to Theorem 8. The main idea is to approximate the (discontinuous) indicator function of an appropriate region by an appropriately "mollified" function (that is continuous with bounded second derivative) so that the corresponding expectations are approximately preserved. There are several different mollification constructions in the literature that could potentially by used for this purpose. We use the following theore from [DKN10].

Theorem 14. *[[DKN10], Theorem 4.8 and Theorem 4.10] Let $I : \mathbb{R}^k \to \{0,1\}$ be the indicator of a region R in $\mathbb{R}^k$ and $c > 0$ be arbitrary. Then there exists a function $\tilde{I}_c : \mathbb{R}^k \to [0,1]$ satisfying:*

- $\|\partial^\beta \tilde{I}_c/\partial x^\beta\|_\infty \le (2c)^{|\beta|}$ *for any $\beta \in \mathbb{N}^k$, and*
- $|I(x) - \tilde{I}_c(x)| \le \min\{1, O((\frac{k}{c\cdot d(x,\partial R)})^2)\}$ *for all $x \in \mathbb{R}^k$,*

where $d(x,\partial R)$ is the Euclidean distance of the point x to the closest point in R.

We use this to prove the following lemma, which says that if a k-dimensional Gaussian X "mimics" the joint distribution Y of a vector of k degree-2 Gaussian polynomials (in the sense of "fooling" all test functions h with bounded second derivative), then X must have small k-dimensional Kolmogorov distance from Y:

Lemma 15. *Let $p_1(x), \ldots, p_k(x) : \mathbb{R}^n \to \mathbb{R}$ be degree-2 polynomials with $\max_{i\in[k]}\mathrm{Var}(p_i) \ge \lambda$, and let Y be their joint distribution when x is drawn from $N(0,1)^n$. Let $X \in \mathbb{R}^k$ be a jointly normal distribution such that $\max_i \mathrm{Var}(X_i) \ge \lambda$. Suppose that for all functions $h : \mathbb{R}^k \to \mathbb{R}, h \in \mathcal{C}^2$, it holds that $|\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]| \le \|h''\|_\infty \cdot \eta$. Then we have*

$$d_{\mathrm{K}}(X,Y) \le O\left(\frac{k^{1/3}\eta^{1/6}}{\lambda^{1/6}}\right).$$

Proof. Fix any $\theta \in \mathbb{R}^n$ and define the function $I : \mathbb{R}^k \to \{0,1\}$ to be the indicator of the region $R \stackrel{\text{def}}{=} \{x \in \mathbb{R}^k : x_i \leq \theta_i\}$. Choose $c > 0$. We have

$$\begin{aligned}
&\left| \mathbf{Pr}[\forall\, i \in [k]\ X_i \leq \theta_i] - \mathbf{Pr}[\forall\, i \in [k]\ Y_i \leq \theta_i] \right| \\
&\quad = \quad |\mathbf{E}[I(X)] - \mathbf{E}[I(Y)]| \\
&\quad \leq \quad \left|\mathbf{E}[\tilde{I}_c(X)] - \mathbf{E}[\tilde{I}_c(Y)]\right| + \left|\mathbf{E}[\tilde{I}_c(Y)] - \mathbf{E}I(Y)]\right| + \left|\mathbf{E}[\tilde{I}_c(X)] - \mathbf{E}[I(X)]\right| \\
&\quad \leq \quad 4c^2\eta + \left|\mathbf{E}[\tilde{I}_c(Y)] - \mathbf{E}I(Y)]\right| + \left|\mathbf{E}[\tilde{I}_c(X)] - \mathbf{E}[I(X)]\right|,
\end{aligned}$$

where we used the first item of Theorem 14 to bound the first term. We proceed to bound the other two terms. For the first one, choose $\delta > 0$ and now note that

$$\begin{aligned}
\left|\mathbf{E}[\tilde{I}_c(Y)] - \mathbf{E}I(Y)]\right| \quad &\leq \quad \mathbf{E}_{y\sim Y}[|\tilde{I}_c(y) - I(y)|] \\
&\leq \quad \mathbf{Pr}_{y\sim Y}[d(y, \partial R) \leq \delta] + O\left(\frac{k^2}{c^2\delta^2}\right) \\
&\leq \quad O\left(\frac{\sqrt{\delta}}{\lambda^{1/4}}\right) + O\left(\frac{k^2}{c^2\delta^2}\right),
\end{aligned}$$

The second inequality above used $0 \leq I, \tilde{I}_c \leq 1$ and the second item of Theorem 14. The final inequality used the Carbery-Wright anti-concentration bound (Theorem 7) together with the observation that in order for $y \sim Y$ to be within distance δ of δR, it must be the case that $|p_i(y) - \theta_i| \leq \delta$ where i is the element of $[k]$ that has $\mathrm{Var}(p_i) \geq \lambda$. Similar reasoning gives that

$$\left|\mathbf{E}[\tilde{I}_c(X)] - \mathbf{E}I(X)]\right| \leq O\left(\frac{\sqrt{\delta}}{\lambda^{1/4}}\right) + O(\frac{k^2}{c^2\delta^2})$$

(in fact here the $\frac{\sqrt{\delta}}{\lambda^{1/4}}$ can be strengthened to $\frac{\delta}{\lambda^{1/2}}$ because now X_i is a degree-1 rather than degree-2 polynomial in $N(0,1)$ Gaussians, but this will not help the overall bound). Optimizing for δ by setting $\delta = k^{4/5}\lambda^{1/10}/c^{4/5}$, we get that

$$\left| \mathbf{Pr}[\forall\, i \in [k]\ X_i \leq \theta_i] - \mathbf{Pr}[\forall\, i \in [k]\ Y_i \leq \theta_i] \right| \leq 4c^2\eta + O\left(\frac{k^{2/5}}{c^{2/5}\lambda^{1/5}}\right).$$

Now optimizing for c by choosing $c = k^{1/6}/(\eta^{5/12}\gamma^{1/12})$, we get that

$$\left| \mathbf{Pr}[\forall\, i \in [k]\ X_i \leq \theta_i] - \mathbf{Pr}[\forall\, i \in [k]\ Y_i \leq \theta_i] \right| \leq O\left(\frac{k^{1/3}\eta^{1/6}}{\lambda^{1/6}}\right),$$

which concludes the proof of Lemma 15. □

With Lemma 15 and Theorem 12 in hand we are ready to prove Theorem 8:

Proof of Theorem 8: For $i \in [k]$ let $\tilde{q}_i(x) = q_i(x) - \mathbf{E}[q_i]$, so $\tilde{q}_i$ has mean zero. Applying Theorem 12 to $\tilde{q} = (\tilde{q}_1, \ldots, \tilde{q}_k)$ we get that any h with $\|h''\|_\infty \leq \infty$ satisfies $|\,\mathbf{E}[h(\tilde{q})] - \mathbf{E}[h(N(0,C))]| \leq O(k^2\epsilon)\cdot\|h''\|_\infty$. Applying Lemma 15, taking X to be $N(0,C)$ and its η parameter to be $O(k^2\epsilon)$, we get that

$$d_{\mathrm{K}}(\tilde{q}, N(0,C)) \ \leq O\left(\frac{k^{2/3}\epsilon^{1/6}}{\lambda^{1/6}}\right),$$

which gives the theorem as claimed. □

4 Transforming a k-tuple of degree-2 Gaussian polynomials

In this section we present a deterministic procedure, called **Transform**, which transforms an arbitrary k-tuple of degree-2 polynomials $(q_1, \ldots, q_k)$ into an "essentially equivalent" (for the purpose of approximately counting PTF satisfying assignments under the Gaussian distribution) k-tuple of degree-2 polynomials $(r_1, \ldots, r_k)$ that have a "nice structure". This structure enables an efficient deterministic decomposition of the joint distribution. In the following section we will give an efficient algorithm to do deterministic approximate counting for vectors of polynomials with this "nice structure."

In more detail, the main theorem of this section, Theorem 16, says the following: Any k-tuple $q = (q_1, \ldots, q_k)$ of degree-2 Gaussian polynomials can be efficiently deterministically transformed into a k-tuple $r = (r_1, \ldots, r_k)$ of degree-2 Gaussian polynomials such that (i) $d_{\mathrm{K}}(r, q) \leq O(\epsilon)$, and (ii) for every restriction fixing the first $t = \mathrm{poly}(k/\epsilon)$ variables, the k-tuple $r|_\rho = (r_1|_\rho, \ldots, r_k|_\rho)$ of restricted polynomials has k-dimensional Kolmogorov distance $O(\epsilon)$ from the k-dimensional Normal distribution with matching mean and covariance matrix. More formally,

Theorem 16. *There is an algorithm* **Transform** *with the following properties: It takes as input a k-tuple $q = (q_1, \ldots, q_k)$ of degree-2 polynomials over $\mathbb{R}^n$ with $\mathrm{Var}_{x \sim N(0,1)^n}[q_i(x)] = 1$ for all $i \in [k]$, and a parameter $\epsilon > 0$. It runs in deterministic time $\mathrm{poly}(n, k, 1/\epsilon)$ and outputs a k-tuple $r = (r_1, \ldots, r_k)$ of degree-2 polynomials over $\mathbb{R}^n$ and a value $0 \leq t \leq O(k \ln(1/\epsilon)/\epsilon^2)$ such that both of the following hold:*

(i) $d_{\mathrm{K}}(q, r) \leq O(\epsilon)$, where q is the random variable $q = (q_1(x), \ldots, q_k(x))$ with $x \sim N(0,1)^n$ and $r = (r_1(y), \ldots, r_k(y))$ with $y \sim N(0,1)^n$; and

(ii) For every restriction $\rho = (\rho_1, \ldots, \rho_t)$, we have

$$d_{\mathrm{K}}(r|_\rho, N(\mu(r|_\rho), \Sigma(r_\rho))) \leq \epsilon.$$

Here "r_ρ" denotes the random variable $(r_1|_\rho(y), \ldots, r_k|_\rho(y))$ where $y \sim N(0,1)^n$ and $r_i|_\rho(y) \stackrel{\text{def}}{=} r_i(\rho_1, \ldots, \rho_t, y_{t+1}, \ldots, y_n)$; $\mu(r|_\rho)$ denotes the vector of means $(\mu_1|_\rho, \ldots, \mu_k|_\rho) \in \mathbb{R}^k$ where $\mu_i|_\rho = \mathbf{E}_{y \sim N(0,1)^n}[r_i|_\rho(y)]$; and $\Sigma(r_\rho)$ denotes the covariance matrix in $\mathbb{R}^{k \times k}$ whose (i,j) entry is $\mathrm{Cov}_{y \sim N(0,1)^n}(r_i|_\rho(y), r_j|_\rho(y))$.

At a high level, the **Transform** procedure first performs a "change of basis" using the procedure **Change-Basis** to convert $q = (q_1(x), \ldots, q_k(x))$ into an "almost equivalent" vector $p = (p_1(y), \ldots, p_k(y))$ of polynomials. (Conceptually the distribution of $(p_1(y), \ldots, p_k(y))$ is identical to the distribution of $(q_1(x), \ldots, q_k(x))$, but in reality some approximations need to be made because we can only approximately compute eigenvalues, etc.; hence the two vector-valued random variables are only "almost equivalent.") Next, the **Transform** procedure runs **Process-Polys** on $(p_1, \ldots, p_k)$; this further changes each p_i slightly, and yields polynomials $r_1, \ldots, r_k$ which are the final output of **Transform**$(q_1, \ldots, q_k)$. A detailed description of the **Transform** procedure follows:

Transform

Input: vector $q = (q_1, \ldots, q_k)$ of degree-2 polynomials $q_\ell(x_1, \ldots, x_n)$ such that $\mathbf{E}_{x \sim N(0,1)^n}[q_\ell(x)^2] = 1$ for all $\ell = 1, \ldots, k$; parameter $\epsilon > 0$

Output: A vector $r = (r_1(y), \ldots, r_k(y))$ of degree-2 polynomials over $\mathbb{R}^n$, and a value $0 \leq t \leq O(k \ln(1/\epsilon)/\epsilon^2)$.

1. Set $\eta = (\epsilon/k)^4/(\log(k/\epsilon))^2$ and $\epsilon' = \epsilon^{12}\eta^2/k^8$.

2. Run **Change-Basis**$((q_1, \ldots, q_k), \epsilon', \eta)$ and let $(p_1, \ldots, p_k), t$ be its output.

3. Run **Process-Polys**$((p_1, \ldots, p_k), t, \eta)$ and let $(r_1, \ldots, r_k), k'$ be its output.

4. Output $(r_1, \ldots, r_k), t$.

Subsection 4.1 below gives a detailed description and analysis of **Change-Basis**, Subsection 4.2 does the same for **Process-Polys**, and Subsection 4.3 proves Theorem 16.

4.1 The Change-Basis procedure.

Intuition. The high-level approach of the **Change-Basis** procedure is similar to the decomposition procedure for vectors of k linear forms that was given in [GOWZ10], but there are significant additional complications that arise in our setting. Briefly, in the [GOWZ10] approach, a vector of k linear forms is simplified by "collecting" variables in a greedy fashion. Each of the k linear forms has a budget of at most B, meaning that at most B variables will be collected on its behalf; thus the overall number of variables that are collected is at most kB. Intuitively, at each stage some variable is collected which has large influence in the remaining (uncollected) portion of some linear form. The [GOWZ10] analysis shows that after at most B variables have been collected on behalf of each linear form, each of the k linear forms will either be regular or its remaining portion (consisting of the uncollected variables) will have small variance. (See Section 7.3 for a more detailed overview of the [GOWZ10] decomposition procedure).

In our current setting, we are dealing with k degree-2 Gaussian polynomials instead of k linear forms, and linear forms will play a role for us which is analogous to the role that single variables played in [GOWZ10]. Thus each quadratic polynomial will have at most B linear forms collected on its behalf and at most kB linear forms will be collected overall. Of course *a priori* there are uncountably many possible linear forms to contend with, so it is not clear how to select a single linear form to collect in each stage. We do this by (approximately) computing the largest eigenvalues of each quadratic form; in each stage we collect some linear form, corresponding to an eigenvector for some quadratic polynomial, whose corresponding eigenvalue is large compared to the variance of the remaining ("uncollected") portion of the quadratic polynomial. An argument similar to that of [GOWZ10] shows that after at most B linear forms have been collected on behalf of each quadratic polynomial, each of the k quadratic polynomials will either be "regular" (have small largest eigenvalue compared to the variance of the remaining portion), or else the variance of the remaining portion will be small.

Remark 17. *In this section we describe an "idealized" version of the algorithm which assumes that we can do certain operations (construct an orthonormal basis, compute eigenvalues and eigenvectors) exactly with no error. In fact these operations can only be carried out approximately, but the errors can in all cases be made extremely small so that running the algorithm with "low-error" implementations of the idealized steps still gives a successful implementation overall. However, keeping track of the errors and approximations is quite cumbersome, so in order to highlight the main ideas we begin by describing the "idealized" version.*

We will try to clearly state all of the idealized assumptions as they come up in the idealized algorithm below. We will state Lemma 26, the main lemma about the **Change-Basis** *algorithm, in versions corresponding both to the "idealized" algorithm and to the "real" algorithm.*

4.1.1 Setup for the Change-Basis procedure. We start with a few definitions. We say that a set $\mathcal{A} = \{L_1(x), \ldots, L_r(x)\}$ of $r \leq n$ linear forms $L_i(x) = v^{(i)} \cdot x$ over $x_1, \ldots, x_n$ is *orthonormal* if $\mathbf{E}_{x \sim N(0,1)^n}[L_i(x) L_j(x)] = \delta_{ij}$ for $1 \leq i, j \leq r$ (equivalently, $v^{(1)}, \ldots, v^{(r)}$ are orthonormal vectors).

Definition 18. *Let $q : \mathbb{R}^n \to \mathbb{R}$ be a degree-2 polynomial*

$$q(x) = \sum_{1 \le i \le j \le n} a_{ij} x_i x_j + \sum_{1 \le i \le n} b_i x_i + c, \tag{7}$$

and let $\{L_i(x) = v^{(i)} \cdot x\}_{i=1,\ldots,n}$ be a full orthonormal set of linear forms. Let $\mathcal{A} = \{L_1, \ldots, L_r\}$ and $\mathcal{B} = \{L_{r+1}, \ldots, L_n\}$ for some $0 \le r \le n$. We define $\mathrm{Proj}(q, \mathcal{A}, \mathcal{B})$, the projection of q onto $\mathcal{A}$, *and $\mathrm{Res}(q, \mathcal{A}, \mathcal{B})$, the* residue of q w.r.t. $\mathcal{A}$, *as follows. Rewrite q using the linear forms $L_i(x)$, i.e.*

$$q = \sum_{1 \le i \le j \le n} \alpha_{ij} L_i(x) L_j(x) + \sum_{1 \le i \le n} \beta_i L_i(x) + c. \tag{8}$$

Define

$$\mathrm{Res}(q, \mathcal{A}, \mathcal{B}) \stackrel{def}{=} \sum_{r < i \le j \le n} \alpha_{ij} L_i(x) L_j(x) + \sum_{r < i \le n} \beta_i L_i(x) + c \tag{9}$$

and

$$\mathrm{Proj}(q, \mathcal{A}, \mathcal{B}) \stackrel{def}{=} q - \mathrm{Res}(q, \mathcal{A}, \mathcal{B}).$$

Note that the residue (resp. projection) of q corresponds to the tail (resp. head) of q in the basis of the linear forms L_i.

Idealized Assumption #1: There is a $\mathrm{poly}(n)$ time deterministic procedure **Complete-Basis** which, given a set $\mathcal{A} = \{L_i(x)\}_{i=1,\ldots,r}$ of orthonormal linear forms, outputs a set $\mathcal{B} = \{L_j(x)\}_{j=r+1,\ldots,n}$ such that $\mathcal{A} \cup \mathcal{B}$ is a full orthonormal set of linear forms.

Claim 19. *There is an efficient algorithm* **Rewrite** *which, given as input q (in the form (7)and sets $\mathcal{A} = \{L_i(x)\}_{i=1,\ldots,r}$, $\mathcal{B} = \{L_{r+1}(x), \ldots, L_n(x)\}$ such that $\mathcal{A} \cup \mathcal{B}$ is a full orthonormal basis, outputs coefficients α_{ij}, β_i, c such that (8) holds.*

Proof sketch: Given $\mathcal{A}$ and $\mathcal{B}$ by performing a matrix inversion it is possible to efficiently compute coefficients u_{ij} such that for $i \in [n]$ we have $x_i = \sum_{j=1}^n u_{ij} L_j(x)$. Substituting $\sum_{j=1}^n u_{ij} L_j(x)$ for each occurrence of x_i in (8) we may rewrite q in the form (8) and obtain the desired coefficients. □

Next we observe that the largest eigenvalue can never increase as we consider the residue of q with respect to larger and larger orthonormal sets of linear forms:

Lemma 20. *Fix any degree-2 polynomial q and any full orthonormal set $\{L_i(x) = v^{(i)} \cdot x\}_{i=1,\ldots,n}$ of linear forms. Let $\mathcal{A} = \{L_i(x) = v^{(i)} \cdot x\}_{i=1,\ldots,r}$ and $\mathcal{B} = \{L_i(x) = v^{(i)} \cdot x\}_{i=r+1,\ldots,n}$. Then we have that $|\lambda_{\max}(\mathrm{Res}(q, \mathcal{A}, \mathcal{B}))| \le |\lambda_{\max}(q)|$.*

Proof. Let M be the $n \times n$ symmetric matrix corresponding to the quadratic part of q, and let M' be the $n \times n$ symmetric matrix corresponding to the quadratic part of $\mathrm{Res}(q, \mathcal{A}, \mathcal{B})$. Let $\tilde{M}$ be the symmetric matrix obtained from M by a change of basis to the new coordinate system defined by the n orthonormal linear forms $L_1, \ldots, L_n$, and likewise let $\tilde{M}'$ be the matrix obtained from M' by the same change of basis. Note that $\tilde{M}'$ is obtained from $\tilde{M}$ by zeroing out all entries $\tilde{M}_{ij}$ that have either $i \in \mathcal{A}$ or $j \in \mathcal{A}$, i.e. $\tilde{M}'$ corresponds to the principal minor $\tilde{M}_{\mathcal{B},\mathcal{B}}$ of $\tilde{M}$. Since eigenvalues are unaffected by a change of basis, it suffices to show that $|\lambda_{\max}(\tilde{M})| \ge |\lambda_{\max}(\tilde{M}')|$.

We may suppose without loss of generality that $\lambda_{\max}(\tilde{M}')$ is positive. By the variational characterization of eigenvalues we have that $\lambda_{\max}(\tilde{M}') = \max_{\|x\|=1} x^T \tilde{M}' x$. Since $\tilde{M}'$ corresponds to the principal minor $\tilde{M}_{\mathcal{B},\mathcal{B}}$ of $\tilde{M}$, a vector x' that achieves the maximum must have nonzero coordinates only in $\mathcal{B}$, and thus

$$\lambda_{\max}(\tilde{M}') = (x')^T \tilde{M}' x' = (x')^T \tilde{M} x' \le \max_{\|x\|=1} x^T \tilde{M}' x \le |\lambda_{\max}(\tilde{M})|.$$

□

4.1.2 The Change-Basis procedure. We now describe the **Change-Basis** procedure. This procedure takes as input a vector $q = (q_1, \ldots, q_k)$ of k degree-2 polynomials, where each q_i is specified explicitly by its coefficientsas in (7), and two parameters $\epsilon', \eta > 0$. It outputs a vector of polynomials $p = (p_1(y), \ldots, p_k(y))$ where each $p_\ell(y_1, \ldots, y_n)$ is also specified explicitly by coefficients $\alpha_{ij}^{(\ell)}$, $\beta_i^{(\ell)}$, $c^{(\ell)}$ that define $p_\ell(y)$ as

$$p_\ell(y) = \sum_{1 \le i \le j \le n} \alpha_{ij}^{(\ell)} y_i y_j + \sum_{1 \le i \le n} \beta_i^{(\ell)} y_i + c^{(\ell)}, \tag{10}$$

and an integer $0 \le t \le k \ln(1/\eta)/\epsilon'^2$. As its name suggests, the **Change-Basis** procedure essentially performs a change of basis on $\mathbb{R}^n$ and rewrites the polynomials $q_\ell(x)$ in the new basis as $p_\ell(y)$. It is helpful to think of y_i as playing the role of $L_i(x)$ where $\{L_i(x)\}_{i=1,\ldots,n}$ is a set of orthonormal linear forms computed by the algorithm, and to think of the coefficients $\alpha_{ij}^{(\ell)}$, $\beta_i^{(\ell)}$, $c^{(\ell)}$ defining $p_\ell(y)$ as being obtained from $q_\ell(x)$ by rewriting $q_\ell(x)$ using the linear forms $L_i(x)$ as in (8).

The **Change-Basis** procedure has two key properties. The first is that the two vector-valued random variables $(q_1(x), \ldots, q_k(x))$ (where $x \sim N(0,1)^n$) and $(p_1(y), \ldots, p_k(y))$ (where $y \sim N(0,1)^n$) are very close in Kolmogorov distance. (In the "idealized" version they are identically distributed, and in the "real" version they are close in k-dimensional Kolmogorov distance.) The second is that each of the p_ℓ polynomials is "nice" in a sense which we make precise in Lemma 26 below. (Roughly speaking, p_ℓ either almost entirely depends only on a few variables, or else has a small-magnitude max eigenvalue.)

Change-Basis

Input: vector $q = (q_1, \ldots, q_k)$ of degree-2 polynomials $q_\ell(x_1, \ldots, x_n)$ such that $\mathbf{E}_{x \sim N(0,1)^n}[q_\ell(x)^2] = 1$ for all $\ell = 1, \ldots, k$; parameters $\epsilon', \eta > 0$

Output: A vector $p = (p_1(y), \ldots, p_k(y))$ of degree-2 polynomials (described explicitly via their coefficients as in (10)) satisfying the guarantees of Lemma 26, and an integer $t \ge 0$.

1. Initialize the set of linear forms $\mathcal{A}$ to be $\emptyset$. Let $\tilde{q}_\ell(x) = q_\ell(x)$ for all $\ell = 1, \ldots, k$.
2. If each $\ell = 1, \ldots, k$ is such that $\tilde{q}_\ell$ satisfies either

   (a) $\mathrm{Var}[\tilde{q}_\ell] \le \eta$, or (b) $\frac{(\lambda_{\max}(\tilde{q}_\ell))^2}{\mathrm{Var}[\tilde{q}_\ell]} \le \epsilon'$,

   then use **Complete-Basis** to compute a set $\mathcal{B}$ of linear forms $\mathcal{B} = \{L_{|\mathcal{A}|+1}(x), \ldots, L_n(x)\}$ such that $\mathcal{A} \cup \mathcal{B}$ is a full orthonormal basis, and go to Step 5. Otherwise, proceed to Step 3.
3. Let $\ell' \in [k]$ be such that $\tilde{q}_{\ell'}$ does not satisfy either (a) or (b) above. Let $v \in \mathbb{R}^n$ be a unit eigenvector corresponding to the maximum magnitude eigenvalue $\lambda_{\max}(\tilde{q}_{\ell'})$. Let $L(x) = v \cdot x$. Add $L(x)$ to $\mathcal{A}$.
4. Use **Complete-Basis**$(\mathcal{A})$ to compute a set of linear forms $\mathcal{B} = \{L_{|\mathcal{A}|+1}(x), \ldots, L_n(x)\}$ such that $\mathcal{A} \cup \mathcal{B}$ is a full orthonormal basis. For all $\ell = 1, \ldots, k$ use **Rewrite**$(q_\ell, \mathcal{A}, \mathcal{B})$ to compute coefficients $\alpha_{ij}^{(\ell)}, \beta_i^{(\ell)}$, $c^{(\ell)}$ as in (8)). Set $\tilde{q}_\ell(x) = \mathrm{Res}(q_\ell, \mathcal{A}, \mathcal{B})$ and $\mathrm{Proj}(q_\ell, \mathcal{A}, \mathcal{B}) = q_\ell(x) - \tilde{q}_\ell(x)$. Go to Step 2.
5. We have $\mathcal{A} = \{L_1(x), \ldots, L_{|\mathcal{A}|}(x)\}$ and $\mathcal{B} = \{L_{|\mathcal{A}|+1}(x), \ldots, L_n(x)\}$. For each $\ell \in [k]$ use **Rewrite** on q_ℓ to compute coefficients $\alpha_{ij}^{(\ell)}$, $\beta_i^{(\ell)}$, $c^{(\ell)}$ such that

$$q_\ell(x) = \sum_{1 \le i \le j \le n} \alpha_{ij}^{(\ell)} L_i(x) L_j(x) + \sum_{1 \le i \le n} \beta_i^{(\ell)} L_i(x) + c^{(\ell)}.$$

Output the polynomials $p_1(y), \ldots, p_k(y)$ defined by these coefficients as in (10), and the value $t = |\mathcal{A}|$.

Idealized assumption #2: There is a $\mathrm{poly}(n)$ time deterministic procedure which, given $\tilde{q}_\ell$ as input,

- exactly computes the maximum eigenvalue $\lambda_{\max}(\tilde{q}_\ell)$, and
- exactly computes a unit eigenvector corresponding to $\lambda_{\max}(\tilde{q}_\ell)$.

Before we proceed with the proof, we recall some basic facts:

Definition 21 (Rotational invariance of polynomials)**.** *Given two polynomials $p(x) = \sum_{1 \le i \le j \le n} a_{ij} x_i x_j + \sum_{1 \le i \le n} b_i x_i + C$ and $q(x) = \sum_{1 \le i \le j \le n} a'_{ij} x_i x_j + \sum_{1 \le i \le n} b'_i x_i + C$ with the same constant term, we say that they are* rotationally equivalent *if there is an orthogonal matrix Q such that $Q^T \cdot A \cdot Q = A'$ and $Q^T \cdot b = b'$. If the matrix A' is diagonal then the polynomial q is said to be the* decoupled equivalent of p*. In this case, the eigenvalues of A (or equivalently A') are said to be the eigenvalues of the quadratic form p.*

Claim 22. *For any degree-2 polynomials $p(x)$ and $q(x)$ which are rotationally equivalent, the distributions of $p(x)$ and $q(x)$ are identical when $(x_1, \ldots, x_n) \sim N(0,1)^n$.*

For $x \sim N(0,1)^n$), since $L_1, \ldots, L_n$ is an orthonormal basis, we have that $(L_1(x), \ldots, L_n(x))$ is distributed identically to $(y_1, \ldots, y_n) \sim N(0,1)^n$. By construction, we have that the matrix corresponding to p_ℓ is an orthogonal transformation of the matrix corresponding to q_ℓ. That is, p_ℓ and q_ℓ are rotationally equivalent.

Recalling the $\mathrm{Tail}_t(\cdot)$ and $\mathrm{Head}_t(\cdot)$ notation from Section 2, we see that the polynomial $\mathrm{Tail}_t(p_\ell(y))$ corresponds precisely to the polynomial $\mathrm{Res}(q_\ell, \mathcal{A}, \mathcal{B})$ and that $\mathrm{Head}_t(p_\ell(y))$ corresponds precisely to $\mathrm{Proj}(q_\ell, \mathcal{A}, \mathcal{B})$. As a consequence, the eigenvalues of $\mathrm{Tail}_t(p_\ell)$ are identical to the eigenvalues of $\tilde{q}_\ell$.

Claim 23. *Let $q(x)$ be a degree-2 Gaussian polynomial and $\mathcal{A} = \{L_1(x), \ldots, L_r(x)\}$ be an orthonormal set of linear forms. Let $\tilde{q}_\ell(x) = \mathrm{Res}(q_\ell, \mathcal{A}, \mathcal{B})$ and let v be a unit eigenvector of (the symmetric matrix corresponding to) $\tilde{q}_\ell$. Then the linear form $L'(x) \stackrel{def}{=} v \cdot x$ is orthogonal to all of $L_1, \ldots, L_r$, i.e., $\mathbf{E}[L'(x) \cdot L_i(x)] = 0$ for all $\ell = 1, \ldots, r$.*

Proof. The claim follows from the fact that v lies in the span of $\mathcal{B}$ (as follows by the definition of the residue) and that the sets of vectors $\mathcal{A}$ and $\mathcal{B}$ are orthonormal. □

The above claim immediately implies that throughout the execution of **Change-Basis**, $\mathcal{A}$ is always an orthonormal set of linear forms:

Corollary 24. *At every stage in the execution of* **Change-Basis***, the set $\mathcal{A}$ is orthonormal.*

(As a side note we observe that since $\mathcal{A} \cup \mathcal{B}$ is a full orthonormal set, it is indeed straightforward to compute $\mathrm{Var}[\tilde{q}_\ell]$ in Step 2; the first time Step 2 is reached this is simply the same as $\mathrm{Var}[q_\ell]$, and in subsequent iterations we can do this in a straightforward way since we have computed the coefficients $\alpha_{ij}^{(\ell)}, \beta_i^{(\ell)}$ in Step 4 immediately before reaching Step 2.)

Next we bound the value of t that the algorithm outputs:

Claim 25. *The number of times that* **Change-Basis** *visits Step 2 is at most $k \ln(1/\eta)/\epsilon'^2$. Hence the value t that the algorithm returns is at most $k \ln(1/\eta)/\epsilon'^2$.*

Proof. It is easy to see that after the end of any iteration, for any fixed $\ell \in [k]$, the variance of $\tilde{q}_\ell$ does not increase. This follows by the definition of the residue and the expression of the variance as a function of the coefficients. At the start of the algorithm each $\tilde{q}_\ell$ has $\mathrm{Var}(\tilde{q}_\ell) = 1$. We claim that each time Step 3 is reached, the polynomial $\tilde{q}_{i'}$, $i' \in [k]$, that is identified in that step has its variance $\mathrm{Var}(\tilde{q}_{i'})$ multiplied by a value which is at most $(1 - \Omega(\epsilon'^2))$ in the corresponding iteration. The claim follows from the fact that the maximum magnitude eigenvalue of $\tilde{q}_{i'}$ is at least $\epsilon' \cdot \sqrt{\mathrm{Var}[\tilde{q}_{i'}]}$ and the definition of the residue. Thus each specific $j \in [k]$ can be chosen as the i' in Step 3 at most $O(\ln(1/\eta)/\epsilon'^2)$ times (after this many iterations it will be the case that $\mathrm{Var}[\tilde{q}_j] \leq \eta$). This proves the claim. □

Thus we have proved the following:

Lemma 26. *(Idealized lemma about* **Change-Basis***:) Given as input a vector $q = (q_1, \ldots, q_k)$ of degree-2 polynomials such that $\mathbf{E}_{x \sim N(0,1)^n}[q_i(x)^2] = 1$ and parameters $\epsilon', \eta > 0$, the algorithm* **Change-Basis**$((q_1, \ldots, q_k), \epsilon', \eta))$ *runs in time $\mathrm{poly}(n, t, 1/\epsilon')$ and outputs polynomials $p_1(y), \ldots, p_k(y)$ (described via their coefficients as in (10)) and a value $0 \leq t \leq k \ln(1/\eta)/\epsilon'^2$ such that items (1) and (2) below both hold.*

1. *The vector-valued random variables $q = (q_1(x), \ldots, q_k(x))$ (where $x \sim N(0,1)^n$) and $p = (p_1(y), \ldots, p_k(y))$ (where $y \sim N(0,1)^n$) are identically distributed.*
2. *For each $\ell \in [k]$, at least one of the following holds:*

   (a) $\mathrm{Var}_{y \sim N(0,1)^n}[\mathrm{Tail}_t(p_\ell(y))] \leq \eta$, *or* *(b)* $\frac{(\lambda_{\max}(\mathrm{Tail}_t(p_\ell))^2}{\mathrm{Var}[\mathrm{Tail}_t(p_\ell)]} \leq \epsilon'$.

(Non-idealized lemma about **Change-Basis***:) This is the same as the idealized lemma except that (1) above is replaced by*

$$d_{\mathrm{K}}(p, q) \leq O(\epsilon'). \tag{11}$$

4.2 The Process-Polys procedure. In this subsection we describe and analyze the **Process-Polys** procedure. Our main result about this procedure is the following:

Lemma 27. *There is a deterministic procedure* **Process-Polys** *which runs in time $\mathrm{poly}(n, k, t, 1/\epsilon', 1/\eta)$ and has the following performance guarantee: Given as input degree-2 polynomials $p_1(y), \ldots, p_k(y)$ satisfying item (2) of Lemma 26, an integer $0 \leq t \leq n$, and a parameter η,* **Process-Polys** *outputs a vector $r = (r_1, \ldots, r_k)$ of degree-2 polynomials over $\mathbb{R}^n$, and a value $0 \leq k' \leq k$, such that r, t, k' satisfy the following properties:*

1. *(r is as good as p for the purpose of approximate counting:)*

$$\left|\mathbf{Pr}_{y \sim N(0,1)^n}[\forall \ell \in [k], r_\ell(y) \leq 0] - \mathbf{Pr}_{x \sim N(0,1)^n}[\forall \ell \in [k], p_\ell(y) \leq 0]\right| \leq O(\epsilon);$$

2. *For any restriction $\rho = (\rho_1, \ldots, \rho_t) \in \mathbb{R}^n$ and all $1 \leq \ell \leq k'$, the polynomial $r_\ell|_\rho$ has degree at most 1;*
3. *For all $k' < \ell \leq k$, the polynomial $r_\ell(y)$ has $\frac{\lambda_{\max}(\mathrm{Tail}_t(r_\ell))^2}{\mathrm{Var}[\mathrm{Tail}_t(r_\ell)]} \leq \epsilon'$;*
4. *For all $k' < \ell \leq k$, the polynomial $r_\ell(y)$ has $\mathrm{Var}[\mathrm{QuadTail}_t(r_\ell)] \geq \eta/2$.*

(Looking ahead, in Section 4.3 Items (2)–(4) of Lemma 27 will be used to show that for most restrictions $\rho = (\rho_1, \ldots, \rho_t)$, the distribution of $(r_1|\rho, \ldots, r_k|\rho)$ is close to the distribution of a multivariate Gaussian with the right mean and covariance. Item (2) handles polynomials $r_1, \ldots, r_{k'}$ and Items (3) and (4) will let us use Theorem 8 for the remaining polynomials.)

Process-Polys

Input: k-tuple $p = (p_1, \ldots, p_k)$ of degree-2 polynomials $p_\ell(y_1, \ldots, y_n)$ such that $\mathrm{Var}_{y \sim N(0,1)^n}[p_\ell(y)] = 1$; integer $t \geq 0$; parameter $\eta > 0$.
Output: k-tuple $r = (r_1, \ldots, r_k)$ of degree-2 polynomials $r_\ell(y_1, \ldots, y_n)$ and integer $0 \leq k' \leq k$.

1. Reorder the polynomials $p_1(y), \ldots, p_k(y)$ so that $p_1, \ldots, p_{k_1}$ are the ones that have $\mathrm{Var}[\mathrm{Tail}_t(p_\ell)] \leq \eta$. For each $\ell \in [k_1]$, define $r_\ell(y) = \mathrm{Head}_t(p_\ell(y)) + \mathbf{E}[\mathrm{Tail}_t(p_\ell)]$.
2. Reorder the polynomials $p_{k_1+1}(y), \ldots, p_k(y)$ so that $p_{k_1+1}(y), \ldots, p_{k_2}(y)$ are the ones that have $\mathrm{Var}[\mathrm{QuadTail}_t(p_\ell(y))] \leq \eta/2$. For each $\ell \in [k_1+1, \ldots, k_2]$, define $r_\ell(y) = p_\ell(y) - \mathrm{QuadTail}_t(p_\ell(y))$.
3. For each $\ell \in [k_2+1, \ldots, k]$ define $r_\ell(y) = p_\ell(y)$. Set $k' = k_2$ and output $(r_1, \ldots, r_k), k'$.

Recall that for $1 \leq \ell \leq k$ each polynomial p_ℓ is of the form

$$p_\ell(y) = \sum_{1 \leq i \leq j \leq n} \alpha_{ij}^{(\ell)} y_i y_j + \sum_{1 \leq i \leq n} \beta_i^{(\ell)} y_i + c^{(\ell)}.$$

Because of Step 2 of **Process-Polys**, for $1 \leq \ell \leq k_1$ we have that each polynomial r_ℓ is of the form

$$r_\ell(y) = \sum_{1 \leq i \leq t, j \geq i} \alpha_{ij}^{(\ell)} y_i y_j + \sum_{1 \leq i \leq t} \beta_i^{(\ell)} y_i + c^{(\ell)},$$

which gives part (2) of the lemma for $1 \leq \ell \leq k_1$. Because of Step 3, for $k_1 + 1 \leq \ell \leq k_2$ we have that each polynomial r_ℓ is of the form

$$r_\ell(y) = \sum_{1 \leq i \leq t, j \geq i} \alpha_{ij}^{(\ell)} y_i y_j + \sum_{1 \leq i \leq n} \beta_i^{(\ell)} y_i + c^{(\ell)},$$

which gives part (2) of the lemma for $k_1 + 1 \leq \ell \leq k_2 = k'$. For $k_2 + 1 \leq \ell \leq k$ each polynomial $r_\ell(y)$ is of the form

$$r_\ell(y) = \sum_{1 \leq i \leq j \leq n} \alpha_{ij}^{(\ell)} y_i y_j + \sum_{1 \leq i \leq n} \beta_i^{(\ell)} y_i + c^{(\ell)} \qquad \text{with } \mathrm{Var}[\mathrm{QuadTail}_t(p_\ell(y))] > \eta/2,$$

which gives part (4) of the lemma.

Part (3) of the lemma follows immediately from item (2) of Lemma 26. Thus the only part which remains to be shown is part (1).

We first deal with the polynomials $r_1, \ldots, r_{k_1}$ using the following simple claim:

Claim 28. *For each $\ell \in [k_1]$ we have that the $r_\ell(x)$ defined in Step 1 of* **Process-Polys** *satisfies*

$$\mathbf{Pr}_{x \sim N(0,1)^n}[\mathrm{sign}(r_\ell(x) \neq \mathrm{sign}(p_\ell(x))] \leq O(\sqrt{\log(1/\eta)} \cdot \eta^{1/4}).$$

Proof. Recall that for $\ell \in [k_1]$ we have $r_\ell = \mathrm{Head}_t(p_\ell) + \mathbf{E}[\mathrm{Tail}_t(p_\ell)]$ while $p_\ell = \mathrm{Head}_t(p_\ell) + \mathrm{Tail}_t(p_\ell)$. Hence $\mathrm{sign}(r_\ell(x) \neq \mathrm{sign}(p_\ell(x))$ only if for some $s > 0$ we have both

$$|\mathrm{Head}_t(p_\ell(x)) + \mathrm{Tail}_t(p_\ell(x))| \leq s\sqrt{\eta} \quad \text{and} \quad |\mathrm{Tail}_t(p_\ell(x)) - \mathbf{E}[\mathrm{Tail}_t(p_\ell(x))]| > s\sqrt{\eta}.$$

To bound the probability of the first event, recalling that $d_{\mathrm{K}}(p_\ell, q_\ell) \leq O(\epsilon')$ (by part (1) of Lemma 26) and that $\mathrm{Var}[q_\ell] = 1$, it easily follows that $\mathrm{Var}[p_\ell] = \Theta(1)$.Hence the Carbery-Wright inequality (Theorem 7) implies that

$$\mathbf{Pr}_x[|\mathrm{Head}_t(p_\ell(x)) + \mathrm{Tail}_t(p_\ell(x))|] \leq s\sqrt{\eta} \leq O(s^{1/2}\eta^{1/4}). \tag{12}$$

For the second event, we recall that $\mathrm{Var}[\mathrm{Tail}_t(p_\ell)] \leq \eta$, and hence for $s > e$ we may apply Theorem 38 to conclude that

$$\mathbf{Pr}_x[|\mathrm{Tail}_t(p_\ell(x)) - \mathbf{E}[\mathrm{Tail}_t(p_\ell(x))]| > s\sqrt{\eta}] \leq O(e^{-s}). \tag{13}$$

Choosing $s = \Theta(\log(1/\eta))$ we get that the RHS of (12) and (13) are both $\Theta(\sqrt{\log(1/\eta)} \cdot \eta^{1/4})$, and the claim is proved. □

It remains to handle the polynomials $r_{k_1+1}, \ldots, r_{k_2}$. For this we use the following claim:

Claim 29. *For each $\ell \in [k_1 + 1, \ldots, k_2]$ we have that the $r_\ell(x)$ defined in Step 2 of* **Process-Polys** *satisfies*

$$\mathbf{Pr}_{x \sim N(0,1)^n}[\mathrm{sign}(r_\ell(x) \neq \mathrm{sign}(p_\ell(x))] \leq O(\sqrt{\log(1/\eta)} \cdot \eta^{1/4}).$$

Proof. The proof is similar to Claim 28. Recall that for $\ell \in [k_1 + 1, k_2]$ we have $r_\ell = p_\ell - \mathrm{QuadTail}_t(p_\ell)$. Hence $\mathrm{sign}(r_\ell(x)) \neq \mathrm{sign}(p_\ell(x))$ only if for some $s > 0$ we have both

$$|p_\ell(x) + \mathbf{E}[\mathrm{QuadTail}_t(p_\ell)]| \leq s\sqrt{\eta} \quad \text{and} \quad |\mathrm{QuadTail}_t(p_\ell(x)) - \mathbf{E}[\mathrm{QuadTail}_t(p_\ell)]| > s\sqrt{\eta}.$$

For the first inequality, as above we have that $\mathrm{Var}[p_\ell] = \Theta(1)$ so as above we get that $\mathbf{Pr}_x[|p_\ell(x) + \mathbf{E}[\mathrm{QuadTail}_t(p_\ell)]| \leq s\sqrt{\eta}] \leq O(s^{1/2}\eta^{1/4})$. For the second inequality we have $\mathrm{Var}[\mathrm{QuadTail}_t(p_\ell)] \leq \eta/2$ so as above we get that $\mathbf{Pr}_x[|\mathrm{QuadTail}_t(p_\ell) - \mathbf{E}[\mathrm{QuadTail}_t(p_\ell)]| > s\sqrt{\eta}] \leq O(e^{-s})$. Choosing $s = \Theta(\log(1/\eta))$ as before the claim is proved.

□

Recalling that $\eta = \Theta((\epsilon/k)^4/(\log(k/\epsilon))^2)$, Claims 28 and 29, together with a union bound, give Lemma 27. □

4.3 Proof of Theorem 16. Given what we have done so far in this section with the **Change-Basis** and **Process-Polys** procedures, the proof of Theorem 16 is simple. Item (1) of Lemma 26 and Item (1) of Lemma 27 immediately give part (i) of Theorem 16. For part (ii), consider any restriction $\rho = (\rho_1, \ldots, \rho_t) \in \mathbb{R}^t$ fixing variables $y_1, \ldots, y_t$ of the polynomials $r_1, \ldots, r_k$.

We begin by observing that if the value k' returned by **Process-Polys** equals k, then Item (2) of Lemma 27 ensures that for all $1 \leq \ell \leq k$ the restricted polynomial $r_\ell|_\rho(y)$ has degree at most 1. In this case the distribution of $(r_1|_\rho(y), \ldots, r_k|_\rho(y))$ for $y \sim N(0,1)^n$ is precisely that of a multivariate Gaussian over $\mathbb{R}^k$. Since such a multivariate Gaussian is completely determined by its mean and covariance matrix, in this case we actually get that $d_{\mathrm{K}}(r|_\rho, N(\mu(r|_\rho), \Sigma(r_\rho))) = 0$. So for the rest of the argument we may assume that $k' < k$, and consequently that there is at least one polynomial r_k that has $\frac{\lambda_{\max}(\mathrm{Tail}_t(r_k))^2}{\mathrm{Var}[\mathrm{Tail}_t(r_k)]} \leq \epsilon'$ and $\mathrm{Var}[\mathrm{QuadTail}_t(r_k)] \geq \eta/2$.

First suppose that no restricted polynomial $r_\ell|_\rho$ has $\mathrm{Var}[r_\ell|_\rho] > 1$. Item (2) of Lemma 27 ensures that for $1 \leq \ell \leq k'$ the restricted polynomial $r_\ell|_\rho(y)$ has degree 1 (note that in terms of Theorem 8, this means that the maximum magnitude of any eigenvalue of $r_\ell|_\rho$ is zero). Now consider any $\ell \in [k' + 1, k]$. Recalling Remark 3, we have that the polynomial $r_\ell|_\rho$ equals $\mathrm{Tail}_t(r_\ell) + L$ for some affine form L. Hence

$$|\lambda_{\max}(r_\ell|_\rho)| = |\lambda_{\max}(\mathrm{Tail}_t(r_\ell))| \leq \sqrt{\mathrm{Var}[\mathrm{Tail}_t(r_\ell)] \cdot \epsilon'} \leq O(\sqrt{\epsilon'}).$$

where the first inequality is by Item (3). The second inequality holds because for $\ell \in [k' + 1, k]$, the polynomial r_ℓ output by **Process-Polys** is simply p_ℓ, so we have $\mathrm{Var}[\mathrm{Tail}_t(r_\ell)] \leq \mathrm{Var}[p_\ell] \leq \mathbf{E}[p_\ell^2]$. As in

the proof of Claim 28 we have that $\mathbf{E}[p_\ell^2] = O(1)$, giving the second inequality above. Item (4) ensures that that for $\ell \in [k'+1, k]$ we have $\mathrm{Var}[r_\ell|_\rho(y)] = \mathrm{Var}[\mathrm{Tail}_t(r_\ell) + L] \geq \mathrm{Var}[\mathrm{QuadTail}_t(r_\ell(y))] \geq \eta/2$. Thus we may apply Theorem 8 and conclude that the distribution of $(r_1|\rho, \ldots, r_k|\rho)$ is $O(k^{2/3}\epsilon'^{1/12}/\eta^{1/6})$-close (i.e. $O(\epsilon)$-close) in d_{K} to the distribution of the appropriate multivariate Gaussian, as claimed in the theorem.

Finally, consider the case that some restricted polynomial $r_\ell|_\rho$ has $\mathrm{Var}[r_\ell|_\rho] > 1$. In this case rescale each such restricted polynomial $r_\ell|_\rho$ to reduce its variance down to 1; let $\tilde{r}_1|_\rho, \ldots, \tilde{r}_k|_\rho$ be the restricted polynomials after this rescaling. As above for $1 \leq \ell \leq k'$ we have that each restricted polynomial $\tilde{r}_\ell|_\rho$ has $\lambda_{\max}(\tilde{r}_\ell|_\rho) = 0$, so consider any $\ell \in [k'+1, k]$. The rescaled polynomials $\tilde{r}_\ell$ satisfy $\tilde{r}_\ell|_\rho = \mathrm{Tail}_t(\tilde{r}_\ell) + \tilde{L}$, and we have

$$\frac{\lambda_{\max}(\mathrm{Tail}_t(\tilde{r}_\ell))^2}{\mathrm{Var}[\mathrm{Tail}_t(\tilde{r}_\ell)]} = \frac{\lambda_{\max}(\mathrm{Tail}_t(r_\ell))^2}{\mathrm{Var}[\mathrm{Tail}_t(r_\ell)]} \leq \epsilon',$$

so we get

$$|\lambda_{\max}(\tilde{r}_\ell|_\rho)| = |\lambda_{\max}(\mathrm{Tail}_t(\tilde{r}_\ell|_\rho))| \leq \sqrt{\mathrm{Var}[\mathrm{Tail}_t(\tilde{r}_\ell)] \cdot \epsilon'} \leq \sqrt{\mathrm{Var}[\mathrm{Tail}_t(r_\ell)] \cdot \epsilon'} \leq O(\sqrt{\epsilon'}),$$

where for the penultimate inequality we recall that $\tilde{r}_\ell$ is obtained by scaling r_ℓ down. By assumption we have that some ℓ has $\mathrm{Var}[\tilde{r}_\ell|_\rho] = 1$, so we can apply Theorem 8 and conclude that

$$d_{\mathrm{K}}(\tilde{r}_\ell|_\rho, N(\mu(\tilde{r}|_\rho), \Sigma(\tilde{r}|_\rho)) \leq O(k^{2/3}\epsilon'^{1/12}).$$

Un-rescaling to return to r_ℓ from $\tilde{r}_\ell$, we get that

$$d_{\mathrm{K}}(r_\ell|_\rho, N(\mu(r|_\rho), \Sigma(r|_\rho)) \leq O(k^{2/3}\epsilon'^{1/12}) = o(\epsilon),$$

and Theorem 16 is proved. □

5 Proof of Theorem 2: Efficient deterministic approximate counting using transformed degree-2 Gaussian polynomials

Throughout this section we focus on counting intersections of degree-2 PTFs. The proof for an arbitrary k-junta follows by expressing it as a disjunction of AND_k functions and a union bound.

Given Theorem 16, there is a natural approach for the counting algorithm **Count-Gauss**, corresponding to the following steps:

Count-Gauss

Input: k-tuple $p = (p_1, \ldots, p_k)$ of degree-2 polynomials $p_\ell(y_1, \ldots, y_n)$, $\ell \in [k]$, such that $\mathrm{Var}_{y \sim N(0,1)^n}[p_\ell(y)] = 1$; parameter $\epsilon > 0$.
Output: An $\pm O(\epsilon)$ additive approximation to the probability $\mathbf{Pr}_{x \sim N(0,1)^n}[\forall \ell \in [k], p_\ell(x) \geq 0]$.

1. Run **Transform**(p, ϵ) to obtain a k-tuple of polynomials $r = (r_1, \ldots, r_k)$ each of unit variance and a value $0 \leq t \leq O(k \ln(1/\epsilon)/\epsilon^2)$.

2. Deterministically construct a product distribution $D^t = \otimes_{i=1}^t D_i$ supported on a set $S \subseteq \mathbb{R}^t$ of cardinality $(kt/\epsilon)^{O(t)}$ such that a t-tuple $\tau = (\tau_1, \ldots, \tau_t) \in \mathbb{R}^t$ drawn from D^t is "close" to a draw of $\rho = (\rho_1, \ldots, \rho_t)$ from $N(0,1)^t$. In particular, $D_i = D$ for all $i \in [t]$, where D is a sufficiently accurate discrete approximation to $N(0,1)$. (See the proof of Lemma 30 for a precise description of the construction and guarantee.)

3. For each $\tau \in S$, simplify the polynomials $r_1, \ldots, r_k$ by applying the restriction to obtain $(r_1|_\tau, \ldots, r_k|_\tau)$, and compute the vector of means $\mu(r_\tau)$ and matrix of covariances $\Sigma(r_\tau)$.

4. Finally, for each $\tau \in S$, deterministically compute a $\pm\epsilon$-accurate additive approximation to the probability $\mathbf{Pr}_{y\sim N(\mu(r_\tau),\Sigma(r_\tau))}[\forall i \in [k],\ y_i \geq 0]$; let p_τ be the value of the approximation that is computed. Average all the values of p_τ obtained for each value $\tau \in S$, and return the average.

Recall that the k-vector of polynomials $r = (r_1 \ldots, r_k)$ constructed in Step 1 satisfies the statement of Theorem 16. In particular, for every restriction of the first t variables, the restricted polynomials are ϵ-close in Kolmogorov distance to a Gaussian with the corresponding mean and covariance matrix. Hence, for each possible restriction ρ of these t variables, the probability that the restricted intersection of polynomials is satisfied is ϵ-close to the quantity $\mathbf{Pr}_{y\sim N(\mu(r_\rho),\Sigma(r_\rho))}[\forall i \in [k],\ y_i \geq 0]$. Hence, if we could take "all" possible restrictions of these t variables, compute the corresponding probabilities and "average" the outcomes, we would end up with an ϵ-approximation to the desired probability. To achieve this efficiently, in Step 2, we construct a sufficiently accurate discrete approximation to the normal distribution $N(0,1)^t$.

We have the following lemma:

Lemma 30. *Let $r_\ell : \mathbb{R}^n \to \mathbb{R}$, $\ell \in [k]$, be k unit variance degree-2 polynomials. There exists a discrete distribution $D^t = \otimes_{i=1}^t D_i$ supported on $(kt/\epsilon)^{O(t)}$ points that can be constructed explicitly in output polynomial time such that*

$$\left|\mathbf{Pr}_{x\sim N^t(0,1),y\sim N^{n-t}(0,1)}\left[\forall \ell \in [k], r_\ell(x,y) \geq 0\right] - \mathbf{Pr}_{\tilde{x}\sim D^t,y\sim N^{n-t}(0,1)}\left[\forall \ell \in [k], r_\ell(\tilde{x},y) \geq 0\right]\right| \leq O(\epsilon).$$

Proof. Before we proceed with the formal proof, we provide some intuition. The main technical point is how "fine" a discretization we need to guarantee an $\pm\epsilon$ approximation to the desired probability

$$\mathbf{Pr}_{z\sim N^n(0,1)}[\forall \ell \in [k], r_\ell(z) \geq 0].$$

Each component D_j, $j \in [t]$, of the product distribution D^t will be a discrete approximation to the standard Gaussian distribution $N(0,1)$. Consider a sample $x = (x_1, \ldots, x_t) \sim N^t(0,1)$ drawn from the standard Gaussian and its coordinate-wise closest discretized value $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_t)$. The main idea is to construct each D_j in such a way so that with probability at least $1 - O(\epsilon/k)$ over x, the absolute difference $\max_{j\in[t]} |x_j - \tilde{x}_j|$ is at most δ (where δ is a sufficiently small quantity). Conditioning on this event, the difference between the two probabilities $\mathbf{Pr}_{x\sim N^t(0,1),y\sim N^{n-t}(0,1)}\left[\forall \ell \in [k], r_\ell(x,y) \geq 0\right]$ and $\mathbf{Pr}_{\tilde{x}\sim D^t,y\sim N^{n-t}(0,1)}\left[\forall \ell \in [k], r_\ell(\tilde{x},y) \geq 0\right]$ can be bounded from above by the probability of the following event: there exists $\ell \in [k]$ such that the polynomial $r_\ell(x,y)$ is "close" to 0 or the difference between the two restricted polynomials $r_\ell(x,y) - r_\ell(\tilde{x},y)$ is "large". Each of these events can in turn be bounded by a combination of anti-concentration and concentration for degree-2 polynomials which completes the proof by a union bound.

Construction of the discrete distribution D^t. The distribution $D^t = \otimes_{j=1}^t D_j$ is a product distribution, whose individual marginals D_j, $j \in [t]$, are identical, i.e., $D_j = D$. The distribution D is a discrete approximation to $N(0,1)$. Intuitively, to construct D we proceed as follows. After truncating the "tails" of the Gaussian distribution, we partition the domain into a set of subintervals I_i. The distribution D will be supported on the leftmost points of the I_i's and the probability mass of each such point will be approximately equal to the mass the Gaussian distribution assigns to the corresponding interval. More specifically, let us denote $\epsilon' = \epsilon/(kt)$ and $M = \Theta(\sqrt{\log(1/\epsilon')})$. Then D is supported on the grid of points $s_i = i \cdot \delta$, where

i is an integer and δ is chosen (with foresight) to be $\delta \stackrel{\text{def}}{=} \Theta\left(\epsilon^2/(k^2\log(k/\epsilon))\right)$. The range of the index i is such that $|i|\cdot\delta \le M$, i.e. $i \in [-s, s]$, where $s \in \mathbb{Z}_+$ with $s = O((1/\delta)\cdot M)$.

The probability mass that D assigns to the point $s_i = i\cdot\delta$ is approximately equal to the probability that a standard Gaussian random variable assigns to the interval $I_i = [s_i, s_{i+1})$. In particular, if $\Phi(I)$ denotes the probability that a standard Gaussian puts in interval I, we will guarantee that

$$\sum_i |\Phi(I_i) - D(s_i)| \le \epsilon'. \tag{14}$$

To achieve this we make the error in each interval to be at most ϵ' divided by the number of intervals. It is clear that D can be constructed explicitly in time $\mathrm{poly}(tk/\epsilon)$. Note that, as a consequence of (14) we have that $d_{\mathrm{K}}(D, N(0,1)) \le \epsilon'$.

Properties of D. We define the natural coupling between $N(0,1)$ and D_j, $j \in [t]$: a sample $x_j \sim N(0,1)$ such that $x_j \in I_i$ is coupled to the point $\tilde{x}_j$ that corresponds to the left endpoint of the interval I_i. If x_j is such that $|x_j| > M$ we map x_j to an arbitrary point. This defines a coupling between the product distributions D^t and $N^t(0,1)$. The main property of this coupling is the following:

Fact 31. *With probability at least $1 - O(\epsilon/k)$ over a sample $x \sim N^t(0,1)$ its "coupled" version $\tilde{x}$ satisfies $\max_{j\in[t]}|x_j - \tilde{x}_j| \le \delta$.*

Proof. For each coordinate $j \in [t]$, it follows from Condition (14) and the concentration of the standard Gaussian random variable that with probability at least $1 - \epsilon'$ we have $|x_j - \tilde{x}_j| \le \delta$. The fact then follows by a union bound. □

We henceforth condition on this event. For technical reasons, we will further condition on the event that $\epsilon' \le |x_j| \le M$ for all $j \in [t]$. This event will happen with probability at least $1 - O(\epsilon/k)$, by Gaussian concentration and anti-concentration followed by a union bound. Note that the complementary event affects the desired probabilities by at most ϵ.

Fix an $x = (x_1, \ldots, x_t)$ with $\epsilon' \le |x_j| \le M$ for all $j \in [t]$ and a value $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_t)$ such that $\max_{j\in[t]} |x_j - \tilde{x}_j| \le \delta$. For $\ell \in [k]$, consider the difference $e_\ell(x, \tilde{x}, y) = r_\ell(x, y) - r_\ell(\tilde{x}, y)$ as a random variable in $y \sim N(0,1)^{n-t}$. We have the following claim:

Claim 32. *We have that $\mathrm{Var}_y[e_\ell] = O(\delta^2)$.*

Proof. Let $r_\ell(x_1, \ldots, x_n) = \sum_{i,j} a_{ij}x_ix_j + \sum_i b_ix_i + C$. By our assumption that $\mathrm{Var}[r_\ell] = 1$ and Claim 5, it follows that the sum of the squares of the coefficients of r_ℓ is in $[1/2, 1]$. A simple calculation yields that the difference between $r_\ell(x_1, x_2, \ldots, x_n)$ and $r_\ell(\tilde{x}_1, \ldots, \tilde{x}_t, x_{t+1}, \ldots, x_n)$ is at most

$$\sum_{1\le i\le j\le t} a_{ij}(x_ix_j - \tilde{x}_i\tilde{x}_j) + \sum_{i\le t, j\ge t+1} a_{ij}(x_i - \tilde{x}_i)x_j + \sum_{i\le t} b_i(x_i - \tilde{x}_i)$$

Taking into consideration our assumption that the sum of the squared coefficients of r_ℓ is at most 1 and that $|x_j - \tilde{x}_j| \le \delta$ for all $j \in [t]$, the variance of the above quantity term can be bounded from above by $O(\delta^2)$. □

Given a value of $\gamma > 0$, the two desired probabilities differ only if there exists $\ell \in [k]$ such that

$$\mathbf{Pr}_{x,\tilde{x},y}[|e_\ell(x, \tilde{x}, y)| \ge \gamma] \tag{15}$$

or

$$\mathbf{Pr}_{x,y}[|r_\ell(x, y)| \le \gamma]. \tag{16}$$

We will select the parameter γ appropriately so that for a given $\ell \in [k]$, both probabilities above are at most $O(\epsilon/k)$. The proof of the lemma will then follow by a union bound over ℓ.

For fixed $x, \tilde{x}$, an application of the Chernoff bound (Theorem 38) in conjunction with Claim 32 implies that $\mathbf{Pr}_y[|e_\ell(x, \tilde{x}, y)| \geq \gamma]$ is at most $\tilde{\epsilon} = \epsilon/k$ as long as $\gamma = \Omega(\log(1/\tilde{\epsilon})\delta)$. By Fact 31 it thus follows that (15) is at most $O(\epsilon/k)$. Similarly, since $\mathrm{Var}[r_\ell] = 1$, by choosing $\gamma = \Theta(\tilde{\epsilon}^2)$, Carbery–Wright (Theorem 7) implies that (16) is at most $O(\tilde{\epsilon})$. By our choice of δ, it follows that for this choice of γ we indeed have that $\gamma = \Omega(\log(1/\tilde{\epsilon})\delta)$, which completes the proof. □

For Step 4 we note that the corresponding problem is that of counting an intersection of k halfspaces with respect to a Gaussian distribution over $\mathbb{R}^k$. We recall that, by Theorem 1.5 of [GOWZ10], $s = \tilde{O}(k^6/\epsilon^2)$-wise independence ϵ-fools such functions. Since we are dealing with a k-dimensional problem, any explicit construction of an s-wise independent distribution yields a deterministic ϵ-approximate counting algorithm that runs in time $k^{O(s)}$, completing the proof of Theorem 2.

6 Deterministic approximate counting for $g(\mathrm{sign}(q_1(x)), \ldots, \mathrm{sign}(q_k(x)))$ over $\{-1, 1\}^n$

In this section we extend the deterministic approximate counting result that we established for the Gaussian distribution on $\mathbb{R}^n$ to the uniform distribution over $\{-1, 1\}^n$, and prove Theorem 1. As discussed in the introduction, there are three main ingredients in the proof of Theorem 1. The first, of course, is the Gaussian counting result, Theorem 2, established earlier. The second is a deterministic algorithmic regularity lemma for k-tuples of low-degree polynomials:

Lemma 33. *[algorithmic regularity lemma, general k, general d] There is an algorithm* `ConstructTree` *with the following property:*

Let $p_1, \ldots, p_k$ be degree-d multilinear polynomials with b-bit integer coefficients over $\{-1, 1\}^n$. Fix $0 < \tau, \epsilon, \delta < 1/4$. Algorithm `ConstructTree` *(which is deterministic) runs in time $\mathrm{poly}(n, b, 2^{D_{d,k}(\tau,\epsilon,\delta)})$ and outputs a decision tree T of depth at most*

$$D_{d,k}(\tau, \epsilon, \delta) := \left(\frac{1}{\tau} \cdot \log \frac{1}{\epsilon} \cdot\right)^{(2d)^{\Theta(k)}} \cdot \log \frac{1}{\delta}.$$

Each internal node of the tree is labeled with a variable and each leaf ρ is labeled with a k-tuple of polynomials $((p_1)_\rho, \ldots, (p_k)_\rho)$ and with a k-tuple of labels $(\mathrm{label}_1(\rho), \ldots, \mathrm{label}_k(\rho))$. For each leaf ρ and each $i \in [k]$ the polynomial $(p_i)_\rho$ is the polynomial obtained by applying restriction ρ to polynomial p_i, and $\mathrm{label}_i(\rho)$ belongs to the set $\{+1, -1,$ "fail", "regular"$\}$. The tree T has the following properties:

1. *For each leaf ρ and index $i \in [k]$, if $\mathrm{label}_i(\rho) \in \{+1, -1\}$, then $\mathbf{Pr}_{x \in \{-1,1\}^n}[\mathrm{sign}((p_i)_\rho(x)) \neq \mathrm{label}_i(\rho)] \leq \epsilon$;*
2. *For each leaf ρ and index $i \in [k]$, if $\mathrm{label}_i(\rho) =$ "regular" then $(p_i)_\rho$ is τ-regular; and*
3. *With probability at least $1-\delta$, a random path from the root reaches a leaf ρ such that $\mathrm{label}_i(\rho) \neq$ "fail" for all $i \in [k]$.*

The third ingredient is the following version of the multidimensional invariance principle, which lets us move from the Gaussian to the Boolean domain:

Theorem 34. *Let $p_1(x), \ldots, p_k(x)$ be degree-d multilinear polynomials over $\{-1,1\}^n$, and let $P_i(x) = \mathrm{sign}(p_i(x))$ for $i = 1, \ldots, k$. Suppose that each p_i is τ-regular. Then for any $g : \{-1,1\}^k \to \{-1,1\}$, we have that*

$$\left|\mathbf{Pr}_{x\sim\{-1,1\}^n}[g(P_1(x), \ldots, P_k(x)) = 1] - \mathbf{Pr}_{\mathcal{G}\sim N(0,1)^n}[g(P_1(\mathcal{G}), \ldots, P_k(\mathcal{G})) = 1]\right| \leq \widetilde{\epsilon}(d, \tau, k),$$

where $\widetilde{\epsilon}(d, \tau, k) := 2^{O(k)} \cdot 2^{O(d)} \cdot \tau^{1/(8d)}$.

The regularity lemma for k-tuples of polynomials, Lemma 33, requires significant technical work; we prove it in Section 7. In contrast, Theorem 34 is a fairly direct consequence of the multidimensional invariance principle of Mossel [Mos08] . We explain how Theorem 34 follows from [Mos08] in Section 6.1. Before establishing the regularity lemma and the invariance principle that we will use, though, we first show how Theorem 1 follows from these results.

Proof of Theorem 1 using Theorem 2, Lemma 33 and Theorem 34: The algorithm for approximating $\mathbf{Pr}_{x\sim\{-1,1\}^n}[g(Q_1(x), \ldots, Q_k(x)) = 1]$ to within an additive $\pm\epsilon$ works as follows. It first runs algorithm `ConstructTree` from Lemma 33 with parameters d, k, τ_0, ϵ_0, and δ_0, where τ_0 satisfies $widetilde\epsilon(d, \tau_0, k) \leq \epsilon/4$, ϵ_0 equals $\epsilon/(4k)$, and δ_0 equals $\epsilon/4$, to construct the decision tree T. It initializes the value $\tilde{v}$ to be 0, and then iterates over all leaves ρ of the tree T, adding a contribution $\tilde{v}_\rho$ to $\tilde{v}$ at each leaf ρ according to the following rules: for a given leaf ρ at depth d_ρ,

- If any $i \in [k]$ has $\mathrm{label}_i(\rho) =$ "fail" then the contribution $\tilde{v}_\rho$ from that leaf is 0. Otherwise,
- Let $\kappa(\rho)$ be the restriction of variables $y_1, \ldots, y_k$ corresponding to the string $(\mathrm{label}_1(\rho), \ldots, \mathrm{label}_k(\rho)) \in \{+1, -1, \text{"regular"}\}$, so $\kappa(\rho)$ fixes variable y_i to $b \in \{+1, -1\}$ if $\mathrm{label}_i(\rho) = b$ and $\kappa(\rho)$ leaves variable y_i unfixed if $\mathrm{label}_i(\rho) =$"regular." Run the algorithm of Theorem 2, providing as input the k-tuple of polynomials $((p_1)_\rho, \ldots, (p_k)_\rho)$, the Boolean function $g_{\kappa(\rho)}$ (i.e. g with restriction $\kappa(\rho)$ applied to it), and the accuracy parameter $\epsilon/4$; let $\tilde{w}_\rho$ be the value thus obtained. The contribution from this leaf is $\tilde{v}_\rho := \tilde{w}_\rho \cdot 2^{-d_\rho}$.

Theorem 2 and Lemma 33 imply that the running time is as claimed; we now prove correctness. Let v denote the true value of $\mathbf{Pr}_{x\sim\{-1,1\}^n}[g(Q_1(x), \ldots, Q_k(x)) = 1]$. We may write v as $\sum_\rho v_\rho$, where the sum is over all leaves ρ of T and $v_\rho = w_\rho \cdot 2^{-d_\rho}$ where

$$w_\rho = \mathbf{Pr}_{x\sim\{-1,1\}^n}[g((Q_1)_\rho(x), \ldots, (Q_k)_\rho(x)) = 1].$$

We show that $|v - \tilde{v}| \leq \epsilon$ by showing that $\sum_\rho |\tilde{v}_\rho - v_\rho| \leq \epsilon$. To do this, let us partition the set of all leaves ρ of T into two disjoint subsets A and B, where a leaf ρ belongs to A if some $i \in [k]$ has $\mathrm{label}_i(\rho) =$"fail". Part (3) of Lemma 33 implies that $\sum_{\rho\in A} 2^{-d_\rho} \leq \delta_0 = \epsilon/4$, so we have that

$$\sum_{\rho\in A} |\tilde{v}_\rho - v_\rho| = \sum_{\rho\in A} v_\rho \leq \sum_{\rho\in A} 2^{-d_\rho} \leq \epsilon/4.$$

We bound $\sum_{\rho\in B} |\tilde{v}_\rho - v_\rho| \leq 3\epsilon/4$ by showing that each leaf $\rho \in B$ satisfies $|w_\rho - \tilde{w}_\rho| \leq 3\epsilon/4$; this is sufficient since

$$\sum_{\rho\in B} |\tilde{v}_\rho - v_\rho| = \sum_{\rho\in B} 2^{-d_\rho} |\tilde{w}_\rho - w_\rho| \leq \left(\max_{\rho\in B} |w_\rho - \tilde{w}_\rho|\right) \cdot \sum_{\rho\in B} 2^{-d_\rho} \leq \max_{\rho\in B} |w_\rho - \tilde{w}_\rho| \leq 3\epsilon/4.$$

So fix any leaf $\rho \in B$. Let $S_{\kappa(\rho)} \subseteq [k]$ be the subset of those indices i such that $\mathrm{label}_i(\rho) =$ "regular". By part (2) of Lemma 33 we have that $(p_i)_\rho$ is τ_0-regular for each $i \in S_{\kappa(\rho)}$. Hence we may apply Theorem 34 to the Boolean function $g_{\kappa(\rho)} : \{-1,1\}^{S_{\kappa(\rho)}} \to \{-1,1\}$, and we get that

$$\begin{aligned} & \left|\mathbf{Pr}_{x\sim\{-1,1\}^n}[g_{\kappa(\rho)}((Q_1)_\rho(x),\ldots,(Q_k)_\rho(x))=1] - \mathbf{Pr}_{\mathcal{G}\sim N(0,1)^n}[g_{\kappa(\rho)}((Q_1)_\rho(\mathcal{G}),\ldots,(Q_k)_\rho(\mathcal{G}))=1]\right| \\ \leq \;& \widetilde{\epsilon}(d,\tau_0,k) \leq \epsilon/4. \end{aligned} \tag{17}$$

By Theorem 2 we have that

$$\left|\tilde{w}_\rho - \mathbf{Pr}_{\mathcal{G}\sim N(0,1)^n}[g_{\kappa(\rho)}((Q_1)_\rho(\mathcal{G}),\ldots,(Q_k)_\rho(\mathcal{G}))=1]\right| \leq \epsilon/4. \tag{18}$$

Finally, part (1) of Lemma 33 and a union bound give that

$$\begin{aligned} \left|w_\rho - \mathbf{Pr}_{x\sim\{-1,1\}^n}[g_{\kappa(\rho)}((Q_1)_\rho(x),\ldots,(Q_k)_\rho(x))=1]\right| &\leq \sum_{i\in([k]\setminus S_{\kappa(\rho)})} \mathbf{Pr}_{x\sim\{-1,1\}^n}[(Q_i)_\rho(x) \neq \mathrm{label}_i(\rho)] \\ &\leq k\cdot\epsilon_0 = \epsilon/4. \end{aligned} \tag{19}$$

Combining (17), (18) and (19) with the triangle inequality we get that $|w_\rho - \tilde{w}_\rho| \leq 3\epsilon/4$, which concludes the proof of Theorem 1. □

6.1 Proof of Theorem 34 . We start by proving the theorem for the case that the k-junta g is the AND_k function. In fact, in this particular case the dependence of the error on the parameter k is polynomial. The generalization to an arbitrary k-junta follows using a union bound and the fact that any k-junta can be written as an OR of at most 2^k AND_k functions, each of which is satisfied by a different point in $\{-1,1\}^k$.

The proof has two steps: In the first step we prove the theorem for "smooth" functions; in the second step we use FT-mollification to reduce the theorem to the smooth case. The first step is an immediate application of Theorem 4.1 in [Mos10]. In particular, the following statement is a corollary of his statement to our setting:

Theorem 35 ([Mos10], Corollary of Theorem 4.1)**.** *Let $p_1(x), p_2(x), \ldots, p_k(x)$ be degree-d multilinear polynomials (where either $x \in \{-1,1\}^n$ or $x \in \mathbb{R}^n$) such that $\mathrm{Var}[p_i] = 1$ and $\max_j \mathrm{Inf}_j(p_i) \leq \tau$ for all $i = 1,\ldots,k$. Let $\Psi : \mathbb{R}^k \to \mathbb{R}$ be a C^3 function with $\|\Psi^{(\mathbf{i})}\|_\infty \leq B$ for every vector $\mathbf{i} \in (\mathbb{Z}_{\geq 0})^n$ with $\|\mathbf{i}\|_1 \leq 3$, where $\Psi^{(\mathbf{i})}$ denotes the $\mathbf{i}$-th iterated partial derivative of Ψ. Then,*

$$\left|\mathbf{E}_{x\sim\{-1,1\}^n}\left[\Psi\left(p_1(x),\ldots,p_k(x)\right)\right] - \mathbf{E}_{G\sim N(0,1)^n}\left[\Psi\left(p_1(G),\ldots,p_k(G)\right)\right]\right| \leq \epsilon := 2Bk^{9/2}(8\sqrt{2})^d \cdot d\sqrt{\tau}.$$

Remark 36. We now briefly explain how the above is obtained from Theorem 4.1 of [Mos10]. Theorem 4.1 considers a k-dimensional multi-linear polynomial $q = (q_1,\ldots,q_k)$. The variance of the k-vector q is defined to be the sum of the variances of the individual components, i.e., $\mathrm{Var}[q] = \sum_{j\in[k]} \mathrm{Var}[q_j]$. Similarly, the influence of the i-th variable on q is defined as the sum of the influences of the components, i.e., $\mathrm{Inf}_i[q] = \sum_{j\in[k]} \mathrm{Inf}_i[q_j]$. The degree of q is the maximum of the degree of the components. Note that when we apply Theorem 4.1 to our setting, the corresponding k-dimensional multi-linear polynomial $p = (p_1,\ldots,p_k)$ has variance equal to k. Similarly, the influence of each variable in p is at most $k\tau$. Finally, the value α in the notation of [Mos10] is by definition equal to $1/2$. (See the derivation on top of p. 21 of the ArXiV version of [Mos10].)

Note that in Theorem 35 the error parameter ϵ depends polynomially on k and exponentially on d. As we now show, when the k-junta g is the AND function, the second step (FT-mollifcation) also results in a polynomial dependence on k.

Let g be the AND function on k variables. We assume (wlog) that the range of g is $\{0,1\}$ as opposed to $\{-1,1\}$. Let $p = (p_1, \ldots, p_k)$ be our k-vector of degree-d multilinear polynomials satisfying the assumptions of Theorem 35. Denote by θ_i and p'_i the constant and non-constant parts of p_i respectively, for $i = 1, \ldots, k$, so $p_i(x) = p'_i(x) + \theta_i$ for $i = 1, \ldots, k$, where $p'_i(x)$ is a degree-d polynomial with constant term 0 and variance 1.

Consider the region $R = \{y_i + \theta_i \geq 0, i \in [k]\} \subseteq \mathbb{R}^k$. We claim that, in order to prove Theorem 34 for g being the AND_k function, it suffices to establish the existence of a smooth function Ψ such that the following two bounds hold:

$$\mathbf{E}_{x\sim\mathcal{D}}\left[\Psi\left(p'_1(x), \ldots, p'_k(x)\right)\right] \approx_\delta \mathbf{E}_{x\sim\mathcal{D}}\left[I_R\left(p'_1(x), \ldots, p'_k(x)\right)\right], \tag{20}$$

where $\mathcal{D}$ is taken either to be the uniform distribution over $\{-1,1\}^n$ or to be $N(0,1)^n$, for an appropriately small value of δ. Indeed, given these two versions of Equation 20, Theorem 34 follows from Theorem 35 and the triangle inequality with $\widetilde{\epsilon} = 2\delta + \epsilon$.

To establish the existence of a smooth approximation Ψ to I_R satisfying 20, we appeal to Theorem 14. In particular, the smooth function Ψ will be the function $\tilde{I}_c$ of that theorem, for an appropriately large value of the parameter $c > 0$. Note that there is a tradeoff between the relevant parameters: On the one hand, the higher the value of c, the better an approximation $\tilde{I}_c$ will be to I_R, and hence the smaller the parameter δ will be. On the other hand, when c increases, so does the upper bound on the magnitude of the derivatives of $\tilde{I}_c$ (see the first condition of Theorem 14). This in turn places a lower bound on the value of B (the maximum value of the third derivative) in Theorem 35 – hence, the parameter ϵ increases. As a consequence of this tradeoff, one needs to select the parameter c carefully to minimize the total error of $\widetilde{\epsilon} = O(\delta + \epsilon)$.

We will additionally need to use the fact that the random vector $p' = (p'_1, \ldots, p'_k)$ is sufficiently anti-concentrated (so that the contribution to the error from the region where I_R and its FT-mollified version differ by a lot is sufficiently small). For the case of the Gaussian distribution, this follows immediately from the Carbery-Wright inequality (Theorem 7). For the case of the uniform distribution over the cube, this follows (as usual), by a combination of the "basic" invariance principle of [MOO10] combined with Theorem 7.

We perform the calculation for the regular boolean case below. It turns out that this is the bottleneck quantitatively – and it subsumes the Gaussian case (since the corresponding anti-concentration bound holds for the Gaussian case as well). We start by recording the following fact, which is a corollary of [MOO10] combined with Theorem 7:

Fact 37. *Let $q : \{-1,1\}^n \to \mathbb{R}$ be a τ-regular degree-d polynomial with $\mathrm{Var}[q] = 1$ and $\rho > 0$. Then, for all $\theta \in \mathbb{R}$ we have*

$$\mathbf{Pr}_{x\in\{-1,1\}^n}\left[|p(x) - \theta| \leq \rho\right] \leq O(d\tau^{1/(8d)}) + O(d\rho^{1/d}).$$

Choice of Parameters: We set $\rho \stackrel{\text{def}}{=} O(\tau^{1/8})$ and choose the parameter c in Theorem 14 equal to $c \stackrel{\text{def}}{=} k/\rho$. We proceed to bound from above the quantity

$$\left|\mathbf{E}_{x\sim\{-1,1\}^n}\left[I_R\left(p'_1(x), \ldots, p'_k(x)\right)\right] - \mathbf{E}_{x\sim\{-1,1\}^n}\left[\tilde{I}_c\left(p'_1(x), \ldots, p'_k(x)\right)\right]\right|.$$

We start by observing that for any $y \in \mathbb{R}^k$, the Euclidean distance $\|y - \partial R\|$ is at least $\min_i |y_i + \theta_i|$. Hence by a union bound combined with the above fact we obtain

$$\mathbf{Pr}_x[\|p'(x) - \partial R\| \leq \rho] \leq \mathbf{Pr}_x[\min_i\{|p'_i(x) + \theta_i|\} \leq \rho] \leq \sum_{i=1}^k \mathbf{Pr}_x[|p'_i(x) + \theta_i| \leq \rho] = O(kd\tau^{1/(8d)}).$$

Similarly, for $w \geq \rho$ we have

$$\mathbf{Pr}_x[\|p'(x) - \partial R\| \leq w] = O(kdw^{1/d}).$$

Using these inequalities and Theorem 14 we bound from above the desired quantity as follows:

$$\begin{aligned}
& \left|\mathbf{E}_x\left[I_R\left(p'(x)\right)\right] - \mathbf{E}_x\left[\tilde{I}_c\left(p'(x)\right)\right]\right| \\
\leq\ & \mathbf{E}_x\left[\left|I_R(p'(x)) - \tilde{I}_c(p'(x))\right|\right] \\
\leq\ & \mathbf{Pr}_x[\|p'(x) - \partial R\| \leq \rho] + \sum_{s=0}^{\infty}\left(\frac{k^2}{c^2 2^{2s}\rho^2}\right)\mathbf{Pr}_x[\|p'(x) - \partial R\| \leq 2^{s+1}\rho] \\
\leq\ & O(kd\tau^{1/(8d)}) + O(kd\rho^{1/d})\sum_{s=0}^{\infty} 2^{-2s}2^{s/d} \quad \text{(by our choice of } c = k/\rho) \\
=\ & O(kd\tau^{1/(8d)}).
\end{aligned}$$

Hence we obtain Equation 20 for $\delta = O(kd\tau^{1/(8d)})$. It remains to determine the corresponding value of ϵ in Theorem 35. Note that, by Theorem 14, the value of the third derivative of the FT-mollified function $\tilde{I}_c$ will be at most $(2c)^3 = O(k/\rho)^3$. This is the value of B, which determines the value of ϵ. The total error ϵ is roughly

$$\epsilon = B \cdot \mathrm{poly}(k) \cdot 2^{O(d)} \cdot \sqrt{\tau} = \mathrm{poly}(k) \cdot 2^{O(d)} \cdot \sqrt{\tau}/\rho^3 = \mathrm{poly}(k) \cdot 2^{O(d)} \cdot \tau^{1/8}.$$

Therefore, the total error is $\widetilde{\epsilon} = 2\delta + \epsilon$ which is at most $\mathrm{poly}(k) \cdot 2^{O(d)} \cdot \tau^{1/(8d)}$. This completes the proof for the case of the AND function. The general case follows via a union bound by viewing an arbitrary k-junta as a disjunction of 2^k AND_k functions.

7 An algorithmic regularity lemma: Proof of Lemma 33

7.1 Useful definitions and tools For $p(x_1, \ldots, x_n) = \sum_{S \subset [n], |S| \leq d} \widehat{p}(S) \prod_{i \in S} x_i$ a multilinear degree-d polynomial over $\{-1,1\}^n$, recall that

$$\mathrm{Inf}_i(p) = \sum_{S \ni i} \widehat{p}(S)^2 = \mathbf{E}_{x_i \in \{-1,1\}}[\mathrm{Var}_{x \setminus x_i \in \{-1,1\}^{n-1}}[p(x)]]$$

and that

$$\sum_{0 \neq S} \widehat{p}(S)^2 = \mathrm{Var}[p] \leq \sum_{i=1}^{n} \mathrm{Inf}_i(p) \leq d \cdot \mathrm{Var}[p]. \tag{21}$$

We say that p is τ-*regular* if for all $i \in [n]$ we have

$$\mathrm{Inf}_i(p) \leq \tau \cdot \mathrm{Var}[p].$$

We will use the following standard tail bound on low-degree polynomials over $\{-1,1\}^n$, see e.g. Theorem 2.12 of [AH11] for a proof. (Here and throughout this section unless otherwise indicated, we write $\mathbf{Pr}[\cdot]$, $\mathbf{E}[\cdot]$ and $\mathrm{Var}[\cdot]$ to indicate probability, expectation, and variance with respect to a uniform draw of x from $\{-1,1\}^n$.)

Theorem 38 ("degree-d Chernoff bound", [AH11])**.** *Let $p : \{-1,1\}^n \to \mathbb{R}$ be a degree-d polynomial. For any $t > e^d$, we have*

$$\mathbf{Pr}[|p(x) - \mathbf{E}[p]| > t \cdot \sqrt{\mathrm{Var}[p]}] \leq de^{-\Omega(t^{2/d})}.$$

As a corollary we have:

Corollary 39. *There is an absolute constant C such that the following holds:*

Let $p : \{-1,1\}^n \to \mathbb{R}$ be a degree-d multilinear polynomial that has

$$|\widehat{p}(\emptyset)| = |\mathbf{E}[p]| \geq (C \log(d/\epsilon))^{d/2} \cdot \mathrm{Var}[p]. \tag{22}$$

*Then $\mathbf{Pr}[\mathrm{sign}(p(x)) \neq \mathrm{sign}(\widehat{p}(\emptyset))] \leq \epsilon$. We say that a polynomial p satisfying (22) is ϵ-*skewed*.*

The following terminology will be convenient for us:

Definition 40. *Fix $0 < \epsilon, \tau < 1/4$ and let $q(x_1, \ldots, x_n)$ be a multilinear degree-d polynomial. We say that q is (τ, ϵ)-*good *if at least one of the following two conditions holds:*

1. *q is τ-regular; or*
2. *q is ϵ-skewed.*

Using this terminology we can give a concise statement of the regularity lemma for a single degree-d polynomial as follows:

Lemma 41. *[regularity lemma, $k = 1$] [DSTW10, Kan13] There is a positive absolute constant A such that the following holds:*

Let p be a degree-d multilinear polynomial over $\{-1,1\}^n$ and fix $0 < \tau, \epsilon, \delta < 1/4$. Then there is a decision tree T of depth at most

$$D_{d,1}(\tau, \epsilon, \delta) := \frac{1}{\tau}\left(d \log \frac{1}{\tau} \log \frac{1}{\epsilon}\right)^{Ad} \cdot \log \frac{1}{\delta},$$

[1]*such that with probability at least $1 - \delta$, at a random leaf ρ the restricted polynomial p_ρ is (τ, ϵ)-good.*

(We note that [DSTW10] states the regularity lemma in a form which is slightly weaker than this because it only claims that for almost every leaf the restricted PTF at that leaf is τ-close to τ-regular. However, inspection of the [DSTW10] proof shows that it actually gives the above result: at almost every leaf the restricted polynomial is either regular or skewed. Proposition 15 of [Kan13] gives a statement equivalent to Lemma 41 above, along with a streamlined proof. We further note that [HKM09] independently established a very similar regularity lemma, although with slightly different parameters, that could also be used in place of Lemma 41.)

7.2 The structural result The main structural result we prove is the following extension of Lemma 41 to k-tuples of degree-d polynomials:

Lemma 42. *[regularity lemma, general k, general d] Let $p_1, \ldots, p_k$ be degree-d multilinear polynomials over $\{-1,1\}^n$. Fix $0 < \tau, \epsilon, \delta < 1/4$. Then there is a decision tree T of depth at most*

$$D_{d,k}(\tau, \epsilon, \delta) \leq \left(\frac{1}{\tau} \cdot \log \frac{1}{\epsilon}\right)^{(2d)^{\Theta(k)}} \cdot \log \frac{1}{\delta}$$

such that with probability at least $1 - \delta$, at a random leaf ρ all the restricted polynomials $(p_1)_\rho, \ldots, (p_k)_\rho$ are (τ, ϵ)-good.

[1]Throughout the paper we write "$D_{d,k}(\tau, \epsilon, \delta)$" to denote the depth bound of the decision tree given by a regularity lemma for k-tuples of degree-d polynomials in which the regularity parameter is τ, the skew parameter is ϵ, and the "probability that a leaf is not (τ, ϵ)-good" parameter is δ.

Remark 43. *It is easy to verify (see Theorem 52 of [DDS13]) that there is an efficient deterministic algorithm that constructs the decision tree whose existence is asserted by the original $k = 1$ regularity lemma for degree-d polynomials, Lemma 41. Given this, inspection of the proof of Lemma 42 shows that the same is true for the decision tree whose existence is asserted by Lemma 42. (The key observation, in both cases, is that given a degree-d polynomial q, it is easy to efficiently deterministically compute the values $|\widehat{q}(\emptyset)|$, $\mathrm{Inf}_i(q)$ and $\mathrm{Var}[q]$, and thus to determine whether or not q is τ-regular and whether or not it is ϵ-skewed.) Thus in order to establish the algorithmic regularity lemma, Lemma 33, it is sufficient to prove Lemma 42.*

Remark 44. *Suppose that we prove a result like Lemma 42 but with a bound of $\gamma(d, k, \tau, \epsilon, \delta)$ on the RHS upper bounding $D_{d,k}(\tau, \epsilon, \delta)$. Then it is easy to see that we immediately get a bound of $\gamma(d, k, \tau, \epsilon, 1/2) \cdot O(\log \frac{1}{\delta})$, simply by repeating the construction $2 \ln \frac{1}{\delta}$ times on leaves that do not satisfy the desired (τ, ϵ)-good condition. Thus to prove Lemma 42 it suffices to prove a bound of the form $\gamma(d, k, \tau, \epsilon, \delta)$ and indeed this is what we do below, by showing that*

$$\gamma(d, k, \tau, \epsilon, \delta) = \left(\frac{1}{\tau} \cdot \log \frac{1}{\epsilon} \cdot \log \frac{1}{\delta}\right)^{(2d)^{\Theta(k)}}$$

is an upper bound on the solution of the equations (24) and (25) given below; see Section 7.7.

7.3 Previous results and our approach. As noted earlier, Gopalan et al. prove a regularity lemma for k-tuples of linear forms in [GOWZ10]. While their lemma is phrased somewhat differently (they prove it in a more general setting of product probability spaces), it yields a result that is qualitatively similar to the special $d = 1$ case of Lemma 42. Indeed, the quantitative bound (i.e. the number of variables that are restricted) in the [GOWZ10] lemma is better than the quantitative bounds we achieve in the case $d = 1$. However, there seem to be significant obstacles in extending the [GOWZ10] approach from linear forms to degree-d polynomials; we discuss their approach, and contrast it with our approach, in the rest of this subsection.

The [GOWZ10] regularity lemma works by "collecting variables" in a greedy fashion. Each of the k linear forms has an initial "budget" of at most B (the exact value of B is not important for us), meaning that at most B variables will be restricted "on its behalf". The lemma iteratively builds a set S where each linear form gets to contribute up to B variables to the set. At each step in building S, if some linear form ℓ_i (a) has not yet exceeded its budget of B variables and (b) is not yet regular, then a variable that has high influence in ℓ_i (relative to the total influence of all variables in ℓ_i) is put into S and the "budget" of ℓ_i is decreased by one. If no such linear form exists then the process ends. It is clear that the process ends after at most kB variables have been added into S. At the end of the process, each linear form ℓ_i is either regular, or else there have been B occasions when ℓ_i contributed a high-influence variable to S. This ensures that if ρ is a random restriction fixing the variables in S, then with high probability the restricted $(\ell_i)_\rho$ will be skewed. (The argument for this goes back to [Ser07, DGJ$^+$10] and employs a simple anti-concentration bound for linear forms with super-increasing weights.)

While these arguments work well for $d = 1$ (linear forms), it is not clear how to extend them to $d > 1$. One issue is that in a linear form, any restriction of a set S of "head" variables leaves the same "tail" linear form (changing only the constant term), while this is not true for higher-degree polynomials. A more significant obstacle is that for $d > 1$, restricted variables can interact with each other "in the head" of the polynomial p_i, and we do not have a degree-d analogue of the simple anti-concentration bound for linear forms with super-increasing weights that is at the heart of the $d = 1$ argument. (This anti-concentration bound uses independence between variables in a linear form to enable a restriction argument saying that regardless of the existence of other variables "between" the variables with super-increasing weights, a linear form containing super-increasing weights must have good anti-concentration. This no longer holds in the higher degree setting.)

Our approach. The idea behind our approach is extremely simple. Consider first the case of $k = 2$ where there are two polynomials p_1 and p_2. For carefully chosen parameters $\tau' \ll \tau$ and $\epsilon' \ll \epsilon$ we first use the usual regularity lemma (for a single polynomial) on p_1 to construct a decision tree such that at a random leaf ρ', the polynomial $(p_1)_{\rho'}$ is with high probability (τ', ϵ')-good. Then at each leaf ρ', we use the usual regularity lemma (for a single polynomial) on $(p_2)_{\rho'}$ to construct a decision tree such that at a random leaf ρ_2 of the tree, the polynomial $((p_2)_{\rho'})_{\rho_2}$ is with high probability (τ, ϵ)-good.

The only thing that can go wrong in the above scheme is that $(p_1)_{\rho'}$ is (τ', ϵ')-good, but as a result of subsequently applying the restriction ρ_2, the resulting polynomial $((p_1)_{\rho'})_{\rho_2}$ is not (τ, ϵ)-good. However, if $(p_1)_{\rho'}$ is τ'-regular, then exploiting the fact that $\tau' \ll \tau$, it can be shown that $((p_1)_{\rho'})_{\rho_2}$ will at least be τ-regular – intuitively this is because restricting the (relatively few) variables ρ_2 required to ensure that $(p_2)_{\rho'}$ becomes (τ, ϵ)-good, cannot "damage" the τ'-regularity of $(p_1)_{\rho'}$ by too much. And similarly, if $(p_1)_{\rho'}$ is ϵ'-skewed, then exploiting the fact that $\epsilon' \ll \epsilon$) it can be shown that $((p_1)_{\rho'})_{\rho_2}$ will at least be ϵ-skewed, for similar reasons. Thus, we can bound the overall failure probability that either polynomial fails to be (τ, ϵ)-good as desired. The general argument for $k > 2$ is an inductive extension of the above simple argument for $k = 2$. [2]

7.4 Proof of Lemma 42 In this section we prove Lemma 42. The argument is an inductive one using the result for $(k-1)$-tuples of degree-d polynomials. As discussed in Remark 44, to establish Lemma 42 it suffices to prove the following:

Lemma 45. *[regularity lemma, general k, general $d > 1$] Let $p_1, \ldots, p_k$ be multilinear degree-d polynomials over $\{-1,1\}^n$. Fix $0 < \tau, \epsilon, \delta < 1/4$. Then there is a decision tree T of depth at most*

$$D_{d,k}(\tau, \epsilon, \delta) \leq \left(\frac{1}{\tau} \cdot \log \frac{1}{\epsilon} \cdot \log \frac{1}{\delta}\right)^{(2d)^{\Theta(k)}} \tag{23}$$

such that with probability at least $1 - \delta$, at a random leaf ρ all of $(p_1)_\rho, \ldots, (p_k)_\rho$ are (τ, ϵ)-good.

Proof. The proof is by induction on k. The base case $k = 1$ is given by Lemma 41; we have that $D_{d,1}(\tau, \epsilon, \delta)$ satisfies the claimed bound (23). So we may suppose that $k \geq 2$ and that Lemma 42 holds for $1, 2, \ldots, k-1$.

Here is a description of how the tree for $p_1, \ldots, p_k$ is constructed.

(a) Let

$$\tau' = \frac{\tau^{\Theta(d)}}{\left(d \log \frac{1}{\tau} \log \frac{1}{\epsilon} \log \frac{1}{\delta}\right)^{\Theta(d^2)}}, \quad \epsilon' = \left(\frac{\epsilon}{d}\right)^{\frac{1}{\tau^2}\left(d \log \frac{1}{\tau} \log \frac{1}{\epsilon}\right)^{\Theta(d)} \cdot (\log \frac{1}{\delta})^2}. \tag{24}$$

Let T' be the depth-$D_{d,k-1}(\tau', \epsilon', \delta/2)$ decision tree obtained by inductively applying the "$k-1$" case of Lemma 45 to the polynomials $p_1(x), \ldots, p_{k-1}$ with parameters τ', ϵ', and $\delta/2$.

(b) For each leaf ρ' in T' such that all of $(p_1)_{\rho'}, \ldots, (p_{k-1})_{\rho'}$ are (τ', ϵ')-good:

– Apply the "$k = 1$" case of Lemma 45 to the polynomial $(p_k)_{\rho'}$ with parameters τ, ϵ, and $\delta/2$. (We say that a leaf/restriction obtained in this second phase, which we denote ρ_k, *extends* ρ'.)

[2] As suggested by the sketch given above, we choose τ' relative to τ so that if $(p_1)_{\rho'}$ is τ'-regular then $((p_1)_{\rho'})_{\rho_2}$ will be τ-regular with probability 1 (and similarly for ϵ' and ϵ). A natural idea is to weaken this requirement so that $((p_1)_{\rho'})_{\rho_2}$ will be τ-regular only with high probability over a random choice of ρ_2. It is possible to give an analysis following this approach, but the details are significantly more involved and the resulting overall bound that we were able to obtain is not significantly better than the bound we achieve with our simpler "probability-1" approach. Very roughly speaking the difficulties arise because it is non-trivial to give a strong tail bound over the choice of a random restriction sampled from a decision tree in which different sets of variables may be queried on different paths.

– Replace the leaf ρ' with the depth-$D_{d,1}(\tau, \epsilon, \delta/2)$ tree (call it $T_{\rho'}$) thus obtained.

(c) Output the resulting tree T.

It is clear that the decision tree T has depth at most

$$D_{d,k}(\tau, \epsilon, \delta) \stackrel{\text{def}}{=} D_{d,k-1}(\tau', \epsilon', \delta/2) + D_{d,1}(\tau, \epsilon, \delta/2). \tag{25}$$

In Section 7.7 we shall show that the quantity $D_{d,k}(\tau, \epsilon, \delta)$ that is defined by (24) and (25) later indeed satisfies (23).

For a given leaf ρ of T, let ρ' be the restriction corresponding to the variables fixed in step (a), and let ρ_k be the restriction that extends ρ' in step (b), so $\rho = \rho'\rho_k$.

In order for it not to be the case that all of $(p_1)_\rho, \ldots, (p_k)_\rho$ are (τ, ϵ)-good at a leaf $\rho = \rho'\rho_k$, one of the following must occur:

(i) one of $(p_1)_{\rho'}, \ldots, (p_{k-1})_{\rho'}$ is not (τ', ϵ')-good;

(ii) all of $(p_1)_{\rho'}, \ldots, (p_{k-1})_{\rho'}$ are (τ', ϵ')-good but $(p_k)_{\rho'\rho_k}$ is not (τ, ϵ)-good;

(iii) all of $(p_1)_{\rho'}, \ldots, (p_{k-1})_{\rho'}$ are (τ', ϵ')-good but one of $(p_1)_{\rho'\rho_k}, \ldots, (p_{k-1})_{\rho'\rho_k}$. is not (τ, ϵ)-good.

By step (a), we have $\mathbf{Pr}[(i)] \leq \delta/2$. Given any fixed ρ' such that all of $(p_1)_{\rho'}, \ldots, (p_{k-1})_{\rho'}$ are (τ', ϵ')-good, by step (b) we have $\mathbf{Pr}_{\rho_k}[(p_k)_{\rho'\rho_k}$ is not (τ, ϵ)-good$] \leq \delta/2$, and hence $\mathbf{Pr}[(ii)] \leq \delta/2$. So via a union bound, the desired probability bound (that with probability $1 - \delta$, all of $(p_1)_{\rho'\rho_k}, \ldots, (p_k)_{\rho'\rho_k}$ are (τ, ϵ)-good at a random leaf $\rho = \rho'\rho_k$) follows from the following claim, which says that (iii) above cannot occur:

Claim 46. *Fix any $i \in \{1, \ldots, k-1\}$. Fix ρ' to be any leaf in T' such that $(p_i)_{\rho'}$ is (τ', ϵ')-good. Then $(p_i)_{\rho'\rho_k}$ is (τ, ϵ)-good.*

To prove Claim 46, let us write $a(x)$ to denote $(p_i)_{\rho'}(x)$, so the polynomial a is (τ', ϵ')-good. There are two cases depending on whether a is τ'-regular or ϵ'-skewed.

Case I: a is τ'-regular. In this case the desired bound is given by the following lemma which we prove in Section 7.5. (Note that the setting of τ' given in Equation (24) is compatible with the setting given in the lemma below.)

Lemma 47. *Let $a(x)$ be a degree-d τ'-regular polynomial, where*

$$\tau' = \frac{1}{2}\left(\frac{d-1}{eD}\right)^{d-1} \cdot \frac{1}{16D^2} \quad \text{and } D = D_{d,1}(\tau, \epsilon, \delta/2).$$

Let T be a depth-D decision tree. Then for each leaf ρ of T, the polynomial a_ρ is τ-regular.

Case II: a is ϵ'-skewed. In this case the desired bound is given by the following lemma which we prove in Section 7.6. (Note that the setting of ϵ' given in Equation (24) is compatible with the setting given in the lemma below.)

Lemma 48. *Let $a(x)$ be a degree-d ϵ'-skewed polynomial, where*

$$\epsilon' = \left(\frac{\epsilon}{d}\right)^{\Theta((eD/d)^2)} \quad \text{and } D = D_{d,1}(\tau, \epsilon, \delta/2).$$

Let T be a depth-D decision tree. Then for each leaf ρ of T, the polynomial a_ρ is ϵ-skewed.

These lemmas, together with the argument (given in Section 7.7) showing that $D_{d,k}(\tau, \epsilon, \delta) = \left(\frac{1}{\tau} \cdot \log \frac{1}{\epsilon} \cdot \log \frac{1}{\delta}\right)^{(2d)^{\Theta(k)}}$ satisfies equations (24) and (25), yield Claim 46. □

7.5 Proof of Lemma 47 The key to proving Lemma 47 is establishing the following claim. (Throughout this subsection the expression "$\left(\frac{d-1}{es}\right)^{d-1}$" and its multiplicative inverse should both be interpreted as 1 when $d = 1$.)

Claim 49. *Let $p(x_1, \ldots, x_n)$ be a multilinear degree-d polynomial which is τ'-regular. Let $S \subset [n]$ be a set of at most s variables and let ρ be a restriction fixing precisely the variables in S. Suppose that*

$$\tau' \leq \frac{1}{2}\left(\frac{d-1}{es}\right)^{d-1} \cdot \min\left\{\frac{1}{16s^2}, \tau\right\}.$$

Then we have that p_ρ is τ-regular.

Proof of Claim 49: Since p is τ'-regular, for each $i \in [n]$ we have that $\mathrm{Inf}_i(p) \leq \tau' \cdot \mathrm{Var}[p]$. Let T denote $[n] \setminus S$, the set of variables that "survive" the restriction. The high level idea of the proof is to show that both of the following events take place:

(i) No variable $j \in T$ has $\mathrm{Inf}_j(p_\rho)$ "too much larger" than $\tau' \cdot \mathrm{Var}[p]$, i.e. all $j \in T$ satisfy $\mathrm{Inf}_j(p_\rho) \leq \alpha\tau' \mathrm{Var}[p]$ for some "not too large" $\alpha > 1$; and

(ii) The variance $\mathrm{Var}[p_\rho]$ is "not too much smaller" than $\mathrm{Var}[p]$, i.e. $\mathrm{Var}[p_\rho] \geq (1-\beta)\mathrm{Var}[p]$ for some "not too large" $0 < \beta < 1$.

Given (i) and (ii), the definition of regularity implies that p_ρ is $\left(\frac{\alpha}{1-\beta} \cdot \tau'\right)$-regular.

Event (i): Upper bounding influences in the restricted polynomial. We use the following simple claim, which says that even in the worst case influences cannot grow too much under restrictions fixing "few" variables in low-degree polynomials.

Claim 50. *Let $p(x_1, \ldots, x_n)$ be a degree-d polynomial and $S \subset [n]$ a set of at most s variables. Then for any $j \in [n] \setminus S$ and any $\rho \in \{-1, 1\}^S$, we have $\mathrm{Inf}_j(p_\rho) \leq \left(\frac{es}{d-1}\right)^{d-1} \cdot \mathrm{Inf}_j(p)$.*

Proof. Let T denote $[n] \setminus S$. Fix any $j \in T$ and any $U \subseteq T$ such that $j \in U$. The Fourier coefficient $\widehat{p_\rho}(U)$ equals $\sum_{S' \subseteq S} \widehat{p}(S' \cup U) \prod_{i \in S'} \rho_i$. Recalling that p has degree d, we see that in order for a subset S' to make a nonzero contribution to the sum it must be the case that $|S'| \leq d - |U| \leq d - 1$, so we have that $\widehat{p_\rho}(U)$ is a (± 1)-weighted sum of at most $\sum_{j=0}^{d-1} \binom{s}{j} \leq \left(\frac{es}{d-1}\right)^{d-1}$ Fourier coefficients of p. It follows from Cauchy-Schwarz that

$$\widehat{p_\rho}(U)^2 = \left(\sum_{S' \subseteq S} \widehat{p}(S' \cup U) \prod_{i \in S'} \rho_i\right)^2 \leq \left(\sum_{S' \subseteq S} \widehat{p}(S' \cup U)^2\right) \cdot \left(\frac{es}{d-1}\right)^{d-1}.$$

Summing this inequality over all $U \subseteq T$ such that $j \in U$, we get that

$$\mathrm{Inf}_j(p_\rho) = \sum_{j \in U \subseteq T} \widehat{p_\rho}(U)^2 \leq \left(\sum_{j \in V \subseteq [n]} \widehat{p}(V)^2\right) \cdot \left(\frac{es}{d-1}\right)^{d-1} = \left(\frac{es}{d-1}\right)^{d-1} \mathrm{Inf}_j(p).$$

□

In the context of event (i), since $\mathrm{Inf}_j(p) \leq \tau' \cdot \mathrm{Var}[p]$, we get that $\mathrm{Inf}_j(p_\rho) \leq \left(\frac{es}{d-1}\right)^{d-1} \cdot \tau' \cdot \mathrm{Var}[p]$, i.e. the "$\alpha$" parameter of (i) is $\left(\frac{es}{d-1}\right)^{d-1}$.

Event (ii): Lower bounding the variance of the restricted polynomial. The following simple claim says that restricting a single variable in a regular polynomial cannot decrease the variance by too much:

Claim 51. *For $p(x_1, \ldots, x_n)$ any multilinear degree-d κ-regular polynomial and ρ any restriction that fixes a single variable to a value in $\{-1, 1\}$, the restricted polynomial p_ρ satisfies $\mathrm{Var}[p_\rho] \geq (1 - 2\sqrt{\kappa})\,\mathrm{Var}[p]$.*

Proof. Let κ be a restriction that fixes x_1 to either $+1$ or -1. For a set $U \subset [n]$, $1 \notin U$ we have that the sets U and $U \cup \{1\}$ together contribute $\widehat{p}(U)^2 + \widehat{p}(U \cup \{1\})^2$ to $\mathrm{Var}[p] = \sum_{0 \neq V} \widehat{p}(V)^2$. In p_ρ, we have $\widehat{p_\rho}(U \cup \{1\}) = 0$ and $\widehat{p_\rho}(U) = \widehat{p}(U) \pm \widehat{p}(U \cup \{1\})$, so the sets U and $U \cup \{1\}$ together contribute $(\widehat{p}(U) \pm \widehat{p}(U \cup \{1\}))^2$ to $\mathrm{Var}[p_\rho]$. Hence the difference between the contributions in p versus in p_ρ is at most $2|\widehat{p}(U)\widehat{p}(U \cup \{1\})|$ in magnitude. Summing over all $U \subset [n], 1 \notin U$ we get that

$$\begin{aligned}
\mathrm{Var}[p] - \mathrm{Var}[p_\rho] &\leq 2 \sum_{1 \notin U \subset [n]} |\widehat{p}(U)\widehat{p}(U \cup \{1\})| \\
&\leq 2 \cdot \sqrt{\sum_{1 \notin U \subset [n]} \widehat{p}(U)^2} \cdot \sqrt{\sum_{1 \notin U \subset [n]} \widehat{p}(U \cup \{1\})^2} \\
&\leq 2 \cdot \sqrt{\mathrm{Var}[p]} \cdot \sqrt{\mathrm{Inf}_1(p)} \\
&\leq 2 \cdot \sqrt{\mathrm{Var}[p]} \cdot \sqrt{\kappa \cdot \mathrm{Var}[p]} \quad \text{(because } p \text{ is } \kappa\text{-regular)} \\
&= 2\sqrt{\kappa} \cdot \mathrm{Var}[p].
\end{aligned}$$

□

To establish part (ii), we consider the restriction ρ fixing all variables in S as being built up by restricting one variable at a time. We must be careful in doing this, because the variance lower bound of Claim 51 depends on the regularity of the current polynomial, and this regularity changes as we successively restrict variables (indeed this regularity is what we are trying to bound). Therefore, for $0 \leq t \leq s$, let us define reg_t as the "worst-case" (largest possible) regularity of the polynomial p after t variables have been restricted (so we have $\mathrm{reg}_0 = \tau'$ since by assumption p is initially τ'-regular); our goal is to upper bound reg_s. For $0 \leq t \leq s$, let ρ_t denote a restriction that fixes exactly t of the s variables in S (so p_{ρ_0} is simply p). By repeated applications of Claim 51 we have

$$\begin{aligned}
\mathrm{Var}[p_{\rho_t}] &\geq \left(1 - 2\sqrt{\mathrm{reg}_{t-1}}\right) \mathrm{Var}[p_{\rho_{t-1}}] \\
&\geq \left(1 - 2\sqrt{\mathrm{reg}_{t-1}}\right)\left(1 - 2\sqrt{\mathrm{reg}_{t-2}}\right) \mathrm{Var}[p_{\rho_{t-2}}] \\
&\geq \cdots \\
&\geq (1 - 2\sqrt{\mathrm{reg}_{t-1}}) \cdots (1 - 2\sqrt{\mathrm{reg}_0})\,\mathrm{Var}[p],
\end{aligned}$$

and by Claim 50 we have that every j satisfies $\mathrm{Inf}_j(p_{\rho_t}) \leq \left(\frac{es}{d-1}\right)^{d-1} \cdot \max_{i \in [n]} \mathrm{Inf}_i(p)$. We shall set parameters so that $\sum_{r=0}^{s-1} \sqrt{\mathrm{reg}_r} \leq \frac{1}{4}$; since

$$(1 - 2\sqrt{\mathrm{reg}_{t-1}}) \cdots (1 - 2\sqrt{\mathrm{reg}_0}) \geq 1 - 2\sum_{r=0}^{s-1} \sqrt{\mathrm{reg}_r},$$

this means that for all $0 \leq t \leq s$ we shall have $\mathrm{Var}[p_{\rho_t}] \geq \frac{1}{2}\,\mathrm{Var}[p]$. We therefore have that every t satisfies

$$\frac{\mathrm{Inf}_j(p_{\rho_t})}{\mathrm{Var}[p_{\rho_t}]} \leq \frac{\left(\frac{es}{d-1}\right)^{d-1} \cdot \max_{i \in [n]} \mathrm{Inf}_i(p)}{\frac{1}{2}\,\mathrm{Var}[p]} \leq 2\left(\frac{es}{d-1}\right)^{d-1} \tau',$$

and therefore $\mathrm{reg}_t \leq 2\left(\frac{es}{d-1}\right)^{d-1}\tau'$. Finally, to confirm that $\sum_{r=0}^{s-1}\sqrt{\mathrm{reg}_r} \leq \frac{1}{4}$ as required, we observe that we have

$$\sum_{r=0}^{s-1}\sqrt{\mathrm{reg}_r} \leq s\sqrt{\max_{0\leq r\leq s-1}\mathrm{reg}_r} \leq s\sqrt{2\left(\frac{es}{d-1}\right)^{d-1}\tau'}$$

which is at most $\frac{1}{4}$ by the conditions that Claim 49 puts on τ'. So we indeed have that

$$\mathrm{reg}_s \leq 2\left(\frac{es}{d-1}\right)^{d-1}\tau' \leq \tau,$$

again by the conditions that Claim 49 puts on τ'. This concludes the proof of Claim 49. □

With Claim 49 in hand we are ready to prove Lemma 47. As stated in the lemma, let $a(x)$ be a degree-d τ'-regular polynomial, where

$$\tau' = \frac{1}{2}\left(\frac{d-1}{eD}\right)^{d-1}\cdot\frac{1}{16D^2} \quad \text{and } D = D_{d,1}(\tau,\epsilon,\delta/2)$$

(note that by the definition of the $D_{d,1}(\cdot,\cdot,\cdot)$ function we have that $\frac{1}{16D^2} < \tau$). Claim 49 gives that at every leaf ρ of T the polynomial a_ρ is τ-regular, and Lemma 47 is proved. □

7.6 Proof of Lemma 48 We may suppose w.l.o.g. that $\mathrm{Var}[a] = 1$. Since a is ϵ'-skewed, we may suppose that $\widehat{p}(\emptyset) \geq (C\log(d/\epsilon'))^{d/2}$.

Let ρ be any restriction fixing up to D variables. The idea of the proof is to show that (i) $\widehat{p_\rho}(\emptyset) > 0$ is still "fairly large", and (ii) $\mathrm{Var}[p]$ is "not too large"; together these conditions imply that p_ρ is skewed. We get both (i) and (ii) from the following claim which is quite similar to Claim 50:

Claim 52. *Let $p(x_1,\dots,x_n)$ be a degree-d polynomial with $\mathrm{Var}[p] = 1$ and $\widehat{p}(\emptyset) = 0$. Let $S \subset [n]$ be a set of at most s variables. Then for any $\rho \in \{-1,1\}^S$, we have that (i) $|\widehat{p_\rho}(\emptyset)| \leq \left(\frac{es}{d}\right)^{d/2}$, and (ii) $\mathrm{Var}[p_\rho] \leq \left(\frac{es}{d-1}\right)^{d-1}\mathrm{Var}[p]$.*

Proof. Let T denote $[n]\setminus S$ and let us write x_S to denote the vector of variables $(x_i)_{i\in S}$ and likewise x_T denotes $(x_i)_{i\in T}$. We may write $p(x)$ as $p(x_S,x_T) = p'(x_S) + q(x_S,x_T)$ where $p'(x_S)$ is the truncation of p comprising only the monomials all of whose variables are in S, i.e. $p'(x_S) = \sum_{U\subseteq S}\widehat{p}(U)\prod_{i\in U}x_i$.

For part (i), it is clear that for $\rho\in\{-1,1\}^S$ we have that $\widehat{p_\rho}$ equals $p'(\rho)$. Since

$$\mathrm{Var}[p'] = \sum_{U\subseteq S}\widehat{p}(U)^2 \leq \sum_{U\subseteq[n]}\widehat{p}(U)^2 = \mathrm{Var}[p] = 1,$$

we have

$$|\widehat{p_\rho}(\emptyset)| = \left|\sum_{U\subseteq S}\widehat{p}(U)\prod_{i\in U}\rho_i\right| \leq \sqrt{\sum_{U\subseteq S}\widehat{p}(U)^2}\cdot\sqrt{\sum_{j=0}^{d}\binom{s}{j}} \leq \binom{es}{d}^{d/2}.$$

For (ii), as in the proof of Claim 50 we get that any nonempty $U\subseteq T$ has

$$\widehat{p_\rho}(U)^2 = \left(\sum_{S'\subseteq S}\widehat{p}(S'\cup U)\prod_{i\in S'}\rho_i\right)^2 \leq \left(\sum_{S'\subseteq S}\widehat{p}(S'\cup U)^2\right)\cdot\left(\frac{es}{d-1}\right)^{d-1}.$$

Summing this inequality over all nonempty $U \subseteq T$,we get that

$$\mathrm{Var}[p_\rho] = \sum_{\emptyset \neq U \subseteq T} \widehat{p_\rho}(U)^2 \leq \left(\sum_{\emptyset \neq V \subseteq [n]} \widehat{p}(V)^2 \right) \cdot \left(\frac{es}{d-1} \right)^{d-1} = \left(\frac{es}{d-1} \right)^{d-1} \mathrm{Var}[p].$$

This concludes the proof of Claim 52. □

Proof of Lemma 48: Fix any leaf ρ in the decision tree T from the statement of Lemma 48. As noted at the start of this subsection we may suppose w.l.o.g. that $\mathrm{Var}[a] = 1$ and $\widehat{a}(\emptyset) \geq (C \log(d/\epsilon'))^{d/2}$. Claim 52 gives us that $\widehat{p_\rho}(\emptyset) \geq (C \log(\frac{d}{\epsilon'}))^{d/2} - (\frac{eD}{d})^{d/2}$ and that $\mathrm{Var}[p_\rho] \leq (\frac{eD}{d-1})^{d-1}$, so p_ρ must be ϵ-skewed as long as the following inequality holds:

$$\left(C \log \left(\frac{d}{\epsilon'} \right) \right)^{d/2} \geq \left(\frac{eD}{d} \right)^{d/2} + \left(\frac{eD}{d-1} \right)^{d-1} \cdot \left(C \log \left(\frac{d}{\epsilon} \right) \right)^{d/2}. \tag{26}$$

Simplifying the above inequality we find that taking ϵ' as specified in Lemma 48 satisfies the inequality, and Lemma 48 is proved.

7.7 The solution to the equations To complete the proof of Lemma 42 it suffices to show that the quantity $D_{d,k}(\tau, \epsilon, \delta)$ that is defined by (24) and (25) indeed satisfies (23). It is clear from Lemma 41 that (23) holds when $k = 1$. A tedious but straightforward induction using (24) and (25) shows that (23) gives a valid upper bound. (To verify the inductive step it is helpful to note that for $k > 1$, equations (24) and (25) together imply that $D_{d,k}(\tau, \epsilon, \delta) \leq 2D_{d,k-1}(\tau', \epsilon', \delta/2)$.)

References

[AH11] Per Austrin and Johan Håstad. Randomly supported independence and resistance. *SIAM J. Comput.*, 40(1):1–27, 2011.

[APL07] H. Aziz, M. Paterson, and D. Leech. Efficient algorithm for designing weighted voting games. In *IEEE Intl. Multitopic Conf.*, pages 1–6, 2007.

[AW85] M. Ajtai and A. Wigderson. Deterministic simulation of probabilistic constant depth circuits. In *Proc. 26th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 11–19, 1985.

[BELY09] I. Ben-Eliezer, S. Lovett, and A. Yadin. Polynomial Threshold Functions: Structure, Approximation and Pseudorandomness. Available at http://arxiv.org/abs/0911.3473, 2009.

[Cha09] S. Chatterjee. Fluctuations of eigenvalues and second-order Poincaré inequalities. *Probability Theory and Related Fields*, 143:1–40, 2009.

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

[DDFS12] A. De, I. Diakonikolas, V. Feldman, and R. Servedio. Near-optimal solutions for the Chow Parameters Problem and low-weight approximation of halfspaces. In *Proc. 44th ACM Symposium on Theory of Computing (STOC)*, pages 729–746, 2012.

[DDS12] Anindya De, Ilias Diakonikolas, and Rocco A. Servedio. The inverse shapley value problem. In *ICALP (1)*, pages 266–277, 2012.

[DDS13] A. De, I. Diakonikolas, and R. Servedio. Deterministic approximate counting for degree-2 polynomial threshold functions. manuscript, 2013.

[DGJ+10] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. Servedio, and E. Viola. Bounded independence fools halfspaces. *SIAM J. on Comput.*, 39(8):3441–3462, 2010.

[DHK+10] Ilias Diakonikolas, Prahladh Harsha, Adam Klivans, Raghu Meka, Prasad Raghavendra, Rocco A. Servedio, and Li-Yang Tan. Bounding the average sensitivity and noise sensitivity of polynomial threshold functions. In *STOC*, pages 533–542, 2010.

[DKN10] Ilias Diakonikolas, Daniel M. Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *Proc. 51st IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 11–20, 2010.

[DOSW11] I. Diakonikolas, R. O'Donnell, R. Servedio, and Y. Wu. Hardness results for agnostically learning low-degree polynomial threshold functions. In *SODA*, pages 1590–1606, 2011.

[DSTW10] I. Diakonikolas, R. Servedio, L.-Y. Tan, and A. Wan. A regularity lemma, and low-weight approximators, for low-degree polynomial threshold functions. In *CCC*, pages 211–222, 2010.

[GHR92] M. Goldmann, J. Håstad, and A. Razborov. Majority gates vs. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992.

[GKM+11] Parikshit Gopalan, Adam Klivans, Raghu Meka, Daniel Stefankovic, Santosh Vempala, and Eric Vigoda. An fptas for #knapsack and related counting problems. In *FOCS*, pages 817–826, 2011.

[GL96] Gene Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, 1996.

[GMR13] P. Gopalan, R. Meka, and O. Reingold. DNF sparsification and a faster deterministic counting algorithm. *Computational Complexity*, 22(2):275–310, 2013.

[GOWZ10] P. Gopalan, R. O'Donnell, Y. Wu, and D. Zuckerman. Fooling functions of halfspaces under product distributions. In *IEEE Conf. on Computational Complexity (CCC)*, pages 223–234, 2010.

[Hås94] J. Håstad. On the size of weights for threshold gates. *SIAM Journal on Discrete Mathematics*, 7(3):484–492, 1994.

[HKM09] P. Harsha, A. Klivans, and R. Meka. Bounding the sensitivity of polynomial threshold functions. Available at http://arxiv.org/abs/0909.5175, 2009.

[Jan97] S. Janson. *Gaussian Hilbert Spaces*. Cambridge University Press, Cambridge, UK, 1997.

[Kan10] D.M. Kane. The Gaussian surface area and noise sensitivity of degree-d polynomial threshold functions. In *CCC*, pages 205–210, 2010.

[Kan11] Daniel M. Kane. k-independent gaussians fool polynomial threshold functions. In *IEEE Conference on Computational Complexity*, pages 252–261, 2011.

[Kan12a] Daniel M. Kane. The correct exponent for the gotsman-linial conjecture. *CoRR*, abs/1210.1283, 2012.

[Kan12b] Daniel M. Kane. A structure theorem for poorly anticoncentrated gaussian chaoses and applications to the study of polynomial threshold functions. In *FOCS*, pages 91–100, 2012.

[Kan13] Daniel M. Kane. The correct exponent for the gotsman-linial conjecture. In *Proc. 28th Annual IEEE Conference on Computational Complexity (CCC)*, 2013.

[KI02] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. Technical Report 9(55), Electronic Colloquium on Computational Complexity (ECCC), 2002.

[KKMS08] A. Kalai, A. Klivans, Y. Mansour, and R. Servedio. Agnostically learning halfspaces. *SIAM Journal on Computing*, 37(6):1777–1805, 2008.

[KRS12] Zohar Shay Karnin, Yuval Rabani, and Amir Shpilka. Explicit dimension reduction and its applications. *SIAM J. Comput.*, 41(1):219–249, 2012.

[LV96] M. Luby and B. Velickovic. On deterministic approximation of DNF. *Algorithmica*, 16(4/5):415–433, 1996.

[LVW93] Michael Luby, Boban Velickovic, and Avi Wigderson. Deterministic approximate counting of depth-2 circuits. In *Proceedings of the 2nd ISTCS*, pages 18–24, 1993.

[MK61] J. Myhill and W. Kautz. On the size of weights required for linear-input switching functions. *IRE Trans. on Electronic Computers*, EC10(2):288–290, 1961.

[MOO10] E. Mossel, R. O'Donnell, and K. K. Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Annals of Mathematics*, 171:295–341, 2010.

[MORS10] K. Matulef, R. O'Donnell, R. Rubinfeld, and R. Servedio. Testing halfspaces. *SIAM J. on Comput.*, 39(5):2004–2047, 2010.

[Mos08] Elchanan Mossel. Gaussian bounds for noise correlation of functions and tight analysis of long codes. *FOCS*, pages 156–165, 2008.

[Mos10] E. Mossel. Gaussian bounds for noise correlation of functions. *GAFA*, 19:1713–1756, 2010.

[MP68] M. Minsky and S. Papert. *Perceptrons: an introduction to computational geometry*. MIT Press, Cambridge, MA, 1968.

[MTT61] S. Muroga, I. Toda, and S. Takasu. Theory of majority switching elements. *J. Franklin Institute*, 271:376–418, 1961.

[Mur71] S. Muroga. *Threshold logic and its applications*. Wiley-Interscience, New York, 1971.

[MZ10] Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. In *STOC*, pages 427–436, 2010.

[Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991.

[Nis92] N. Nisan. Pseudorandom generators for space-bounded computations. *Combinatorica*, 12(4):449–461, 1992.

[Nou12] I. Nourdin. Lectures on gaussian approximations with malliavin calculus. Technical Report http://arxiv.org/abs/1203.4147v3, 28 June 2012.

[NP09] I. Nourdin and G. Peccati. Stein's method meets malliavin calculus: a short survey with new estimates. Technical Report http://arxiv.org/abs/0906.4419v2, 17 Sep 2009.

[NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Computer & Systems Sciences*, 49(2):149–167, 1994.

[Orp92] P. Orponen. Neural networks and complexity theory. In *Proceedings of the 17th International Symposium on Mathematical Foundations of Computer Science*, pages 50–61, 1992.

[OS11] R. O'Donnell and R. Servedio. The Chow Parameters Problem. *SIAM J. on Comput.*, 40(1):165–199, 2011.

[Pod09] V. V. Podolskii. Perceptrons of large weight. *Problems of Information Transmission*, 45(1):46–53, 2009.

[Ser07] R. Servedio. Every linear threshold function has a low-weight approximator. *Comput. Complexity*, 16(2):180–209, 2007.

[She08] Alexander A. Sherstov. Halfspace matrices. *Computational Complexity*, 17(2):149–178, 2008.

[She09] A. Sherstov. The intersection of two halfspaces has high threshold degree. In *Proc. 50th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2009.

[SSSS11] Shai Shalev-Shwartz, Ohad Shamir, and Karthik Sridharan. Learning kernel-based halfspaces with the 0-1 loss. *SIAM J. Comput.*, 40(6):1623–1646, 2011.

[Tre04] L. Trevisan. A note on approximate counting for k-DNF. In *Proceedings of the Eighth International Workshop on Randomization and Computation*, pages 417–426, 2004.

[Vio09] E. Viola. The Sum of d Small-Bias Generators Fools Polynomials of Degree d. *Computational Complexity*, 18(2):209–217, 2009.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de