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

Anindya De
School of Mathematics
IAS
Princeton, NJ U.S.A
Email: anindya@math.ias.edu

Ilias Diakonikolas
School of Informatics
University of Edinburgh
Edinburgh, Scotland
Email: ilias.d@ed.ac.uk

Rocco A. Servedio
Department of Computer Science
Columbia University
New York, NY U.S.A.
Email: rocco@cs.columbia.edu

***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 [1] 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 [1]. 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 [2], [3] (which apply to a single degree-d polynomial threshold function) and the regularity lemma of Gopalan et al [4] (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.

***Keywords*-Approximate counting; derandomization; polynomial threshold function**

I. Introduction

Unconditional derandomization has been an important research area in computational complexity theory over the past two decades [5]–[8]. 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 [1], [9]–[14]. 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 [15], 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., [16]–[23] and many other works) due to their importance in fields such as concrete complexity theory [24]–[30], learning theory [31]–[34], voting theory [35], [36], 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 [1] 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*.

A. 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*

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

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 [1], which establishes the special $k = 1$ case of the current result.

As noted in [1], 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 [14] 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 [1], 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 [1].

We remark that the only previous deterministic approximate counting algorithm for k-juntas of degree-2 PTFs follows from the *pseudorandom generators* (PRGs) of [37] (which are based on bounded independence). The running time of the resulting algorithm is $n^{\mathrm{poly}(1/\epsilon)}$, even for $k = 1$.

B. 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 [4], [28], [37]–[40], learning and property testing [41], [42], 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 [43] combined with appropriate mollification arguments [37]. 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 [1]. To explain our new contributions let us first briefly recall the [1] approach.

The main observation enabling the result in [1] 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 $p(x) = \sum_{i,j} a_{ij}x_ix_j$ (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, [1] employs a "critical-index" based analysis on the eigenvalues of the corresponding matrix. For the analysis of the [1] algorithm, [1] proves a CLT for a single degree-2 Gaussian polynomial with small eigenvalues (this CLT is based on a result of Chaterjee [44]). (We note that this informal description suppresses several non-trivial technical issues, see [1] for details.)

At a high level, the approach of the current paper builds on the approach of [1]. 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 3). 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 [45] (Theorem 6) 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 [45] then gives us a version of our desired CLT with respect to "test functions" with bounded second derivatives (Theorem 7). In the second step, we use tools from mollification [37] 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 IV).

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 [4]. However, there are significant complications that arise in our setting. In particular, in the [4] 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 [4] 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 [4]. Mimicking the [4] 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 [4] 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*

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

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 [1] (indeed, observe the above algorithm runs in time exponential in $1/\epsilon$ even for $k = 1$, whereas the algorithm of [1] 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 [46] 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 [2]–[4], [47], but none with all the properties that we require. [48] implicitly gave a regularity lemma for a single LTF, and [2], [3], [47] each gave (slightly different flavors of) regularity lemmas for a single degree-d PTF. Subsequently [4] 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 IV uses ideas from their work. However, as we describe in Section 7 of the full version, their approach does not seem to extend to degrees $d > 1$, so we must use a different approach to prove our regularity lemma.

C. Organization.

In Section II we state and prove our new multidimensional CLT. We give the transformation procedure that is at the heart of our decomposition approach in Section III, and present the actual deterministic counting algorithm for the Gaussian case that uses this transformation in Section IV. Section V shows how the new regularity lemma for k-tuples of Boolean PTFs gives the main Boolean counting result. Due to space limitations, the regularity lemma is proved in the full version.

II. A multidimensional CLT for degree-2 Gaussian polynomials

In this section we present 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.

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 $d_{\mathrm{K}}(X,Y) = \sup_{(\theta_1, \ldots, \theta_k) \in \mathbb{R}^k} |\mathbf{Pr}[\forall\, i \in [k]\ X_i \leq \theta_i] - \mathbf{Pr}[\forall\, i \in [k]\ Y_i \leq \theta_i]|$. We have the following:

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

Looking ahead to motivate this result for our ultimate purposes, Theorem 3 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 3 is a CLT due to Nourdin and Peccati [45] which gives a bound that involves the Malliavin derivative of the functions $q_1, \ldots, q_k$. In Section II-A we give the necessary background from Malliavin calculus and build on the [45] result to prove a result which is similar to Theorem 3 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 7 below). In Section II-B we show how Theorem 7 can be combined with standard "mollification" techniques to yield Theorem 3.

A. 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 [45]. 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. [45], [49] 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 6 below:

Claim 4 (see e.g. Equation (2.22) of [45])**.** *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 5. *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 6. *[[45], see also [49], 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*

$$|\mathbf{E}[h(p)] - \mathbf{E}[h(N)]| < \frac{1}{2} \|h''\|_\infty \cdot \mathrm{disc}(C, Y)$$

where $Y \in \mathbb{R}^{k \times k}$ such that $Y(i,j) = \langle Dp_i\ , -DL^{-1}p_j \rangle$ and $\mathrm{disc}(C, Y) = \sum_{i=1}^k \sum_{j=1}^k \mathbf{E}[|C(i,j) - Y(i,j)|]$.

We now use Theorem 6 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 3, which uses multidimensional Kolmogorov distance.)

Theorem 7. *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 6, 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

$$C(a,b) = \mathrm{Cov}(q_a, q_b) = \mathbf{E}[q_a q_b] = \mathbf{E}[Y(a,b)], \quad (2)$$

where the second equality is because q_a and q_b have mean 0 and the third equality is by Corollary 5. Since C is a covariance matrix and every covariance matrix is PSD, we may apply Theorem 6, and we get that

$$|\mathbf{E}[h(q)] - \mathbf{E}[h(N)]| < \frac{k^2}{2} \|h''\|_\infty \cdot \|C - Y\|_\infty$$

where $\|C - Y\|_\infty$ is the largest entry of the matrix $C - Y$. Further, the quantity on the right hand side is exactly

$$\|h''\|_\infty \cdot \max_{1 \le a,b \le k} \mathbf{E}[|Y(a,b) - \mathbf{E}[Y(a,b)]|$$

where we used (2) for the equality. By Jensen's inequality we have $\mathbf{E}[|Y(a,b) - \mathbf{E}[Y(a,b)]|] \le \sqrt{\mathrm{Var}[Y(a,b)]}$. Lemma 8 below gives us that $\mathrm{Var}[Y(a,b)] \le O(\epsilon^2)$, and the theorem is proved. ∎

It remains to establish the following lemma:

Lemma 8. *For each $1 \le a, b \le k$, we have that $\mathrm{Var}[Y(a,b)] = O(\epsilon^2)$.*

Proof: Fix $1 \le a, b \le k$, 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

$$q_a(x) = \sum_{i=1}^n \lambda_i x_i^2 + \sum_{i=1}^n \beta_i x_i + \gamma$$

and

$$q_b(x) = \sum_{i,j=1}^n \delta_{ij} x_i x_j + \sum_{i=1}^n \kappa_i x_i + \rho,$$

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

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

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

Hence

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

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

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

Given $p : \mathbb{R}^n \to \mathbb{R}$ defined by $p(x) = \sum_{1 \le i \le j \le n} a_{ij} x_i x_j + \sum_{1 \le i \le n} b_i x_i + C$, let us define $SS(p)$ as $SS(p) = \sum_{1 \le i \le j \le n} a_{ij}^2 + \sum_{1 \le i \le n} b_i^2$. We recall the following straightforward claim, which was established in [1]:

Claim 9. *[Claim 20 of [1]] Given $p : \mathbb{R}^n \to \mathbb{R}$, we have that $2\,SS(p) \ge \mathrm{Var}(p) \ge SS(p)$.*

By Claim 9, 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

$$\sum_{i=1}^n \sum_{\ell=1}^n (\lambda_\ell \delta_{i\ell})^2 + \sum_{\ell=1}^n (\kappa_\ell \lambda_\ell)^2 + \sum_{i=1}^n \left(\sum_{\ell=1}^n \beta_\ell \delta_{i\ell} \right)^2 \le O(\epsilon^2). \tag{3}$$

We bound each term of (3) 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 9 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

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

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 8. ∎

B. 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 7 to Theorem 3. 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 theorem from [37].

Theorem 10. *[[37], 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*

- $|I(x) - \tilde{I}_c(x)| \leq \min\{1, O((\frac{k}{c \cdot d(x,\partial R)})^2)\}$ *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 11. *Let* $p_1(x), \ldots, p_k(x) : \mathbb{R}^n \to \mathbb{R}$ *be degree-2 polynomials with* $\max_{i \in [k]} \mathrm{Var}(p_i) \geq \lambda$*, 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* $\max_i \mathrm{Var}(X_i) \geq \lambda$*. Suppose that for all functions* $h : \mathbb{R}^k \to \mathbb{R}, h \in \mathcal{C}^2$*, it holds that* $|\mathbb{E}[h(X)] - \mathbb{E}[h(Y)]| \leq \|h''\|_\infty \cdot \eta$*. Then we have*

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

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

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

where we used the first item of Theorem 10 to bound the first term. We proceed to bound the other two terms. For the first one, choose $\delta > 0$ and now note that

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

The second inequality above used $0 \leq I, \tilde{I}_c \leq 1$ and the second item of Theorem 10. The final inequality used the Carbery-Wright anti-concentration bound 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

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

(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

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

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

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

which concludes the proof of Lemma 11. ■

With Lemma 11 and Theorem 7 in hand we are ready to prove Theorem 3:

Proof of Theorem 3: 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 7 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 11, 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. □

III. 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 12, 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 12. *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 $0 \le t \le O(k \ln(1/\epsilon)/\epsilon^2)$ such that both of the following hold:*

(i) *$d_{\mathrm{K}}(q, r) \le O(\epsilon)$, 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 $d_{\mathrm{K}}(r|_\rho, N(\mu(r|_\rho), \Sigma(r_\rho))) \le \epsilon$. Here "$r_\rho$" denotes the random variable $(r_1|_\rho(y), \ldots, r_k|_\rho(y))$ where $y \sim N(0,1)^n$ and $r_i|_\rho(y) \stackrel{def}{=} r_i(\rho_1, \ldots, \rho_t, y_{t+1}, \ldots, y_n)$; $\mu(r|_\rho)$ 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 $0 \le t \le O(k \ln(1/\epsilon)/\epsilon^2)$.

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 III-A below gives a description and analysis sketch of **Change-Basis**. Further details are postponed to the full version.

A. The Change-Basis procedure.

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 $r \le n$ 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 $1 \le i, j \le r$ (equivalently, $v^{(1)}, \ldots, v^{(r)}$ are orthonormal vectors).

Definition 13. *Let $q : \mathbb{R}^n \to \mathbb{R}$ be a degree-2 polynomial $q(x) = \sum_{1\le i\le j\le n} a_{ij}x_ix_j + \sum_{1\le i\le n} b_ix_i + c$ 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.*

$$q = \sum_{1\le i\le j\le n} \alpha_{ij}L_i(x)L_j(x) + \sum_{1\le i\le n} \beta_iL_i(x) + c. \quad (4)$$

Define

$$\mathrm{Res}(q, \mathcal{A}, \mathcal{B}) \stackrel{def}{=} \sum_{r<i\le j\le n} \alpha_{ij}L_i(x)L_j(x) + \sum_{r<i\le n} \beta_iL_i(x) + c \quad (5)$$

and

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

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 14. *There is an efficient algorithm* **Rewrite** *which, given as input q 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 (4) holds.*

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 15. *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 $|\lambda_{\max}(\mathrm{Res}(q, \mathcal{A}, \mathcal{B}))| \le |\lambda_{\max}(q)|$.*

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 coefficients, 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

$$p_\ell(y) = \sum_{1 \le i \le j \le n} \alpha_{ij}^{(\ell)} y_i y_j + \sum_{1 \le i \le n} \beta_i^{(\ell)} y_i + c^{(\ell)}, \quad (6)$$

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 (4).

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 16 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 (6)) satisfying the guarantees of Lemma 16, 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 (4)). 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 $q_\ell(x)$ equals
$$\sum_{1 \le i \le j \le n} \alpha_{ij}^{(\ell)} L_i(x) L_j(x) + \sum_{1 \le i \le n} \beta_i^{(\ell)} L_i(x) + c^{(\ell)}.$$
Output the polynomials $p_1(y), \ldots, p_k(y)$ defined by these coefficients as in (6), 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 $\lambda_{\max}(\tilde{q}_\ell)$.

In the full version, we prove the following:

Lemma 16. *(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 (6)) and a value $0 \le t \le k \ln(1/\eta)/\epsilon'^2$ 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) $\mathrm{Var}_{y \sim N(0,1)^n}[\mathrm{Tail}_t(p_\ell(y))] \le \eta$, *or*
   (b) $\frac{(\lambda_{\max}(\mathrm{Tail}_t(p_\ell))^2}{\mathrm{Var}[\mathrm{Tail}_t(p_\ell)]} \le \epsilon'$.

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

$$d_{\mathrm{K}}(p, q) \le O(\epsilon'). \quad (7)$$

In a real number model of computation, the "Idealized Assumptions" hold true and we obtain the "idealized" version of Lemma 16. In a bit complexity model the "Idealized Assumptions" do not hold in that we cannot exactly compute the desired quantities, and instead high-accuracy approximations must be used. See the end of Section 2 (Preliminaries) of the full version for a discussion of approximation issues and error bounds.

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

Given Theorem 12, 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 17 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 12. 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 17. *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*

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

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 [4], $s = \tilde{O}(k^6/\epsilon^2)$-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.

V. 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 18. *[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 19. *Let $p_1(x), \dots, p_k(x)$ be degree-d multilinear polynomials over $\{-1,1\}^n$, and let $P_i(x) = \mathrm{sign}(p_i(x))$ for $i = 1, \dots, k$. Suppose that each p_i is τ-regular. Then for any $g : \{-1,1\}^k \to \{-1,1\}$, we have that*

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

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 18, requires significant technical work; we prove it in Section 7 of the full version. In contrast, Theorem 19 is a fairly direct consequence of the multidimensional invariance principle of Mossel [46] (see full version). In the full version we show how these results can be combined to prove Theorem 1.

ACKNOWLEDGMENT

A.D. was supported in part by Umesh Vazirani's Templeton Foundation Grant 21674, by the Simons Institute at UC Berkeley, and by the National Science Foundation under agreements Princeton University Prime Award No. CCF-0832797 and Sub-contract No. 00001583. Some of this work was done while A.D. was at the Simons Institute, UC Berkeley. I.D. was supported in part by a SICSA PECE grant and a Carnegie research grant. R.A.S. was supported by NSF grant CCF-1115703.

REFERENCES

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

[2] 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*, 2010, pp. 211–222.

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

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

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

[6] N. Nisan, "Pseudorandom bits for constant depth circuits," *Combinatorica*, vol. 11, no. 1, pp. 63–70, 1991.

[7] ——, "Pseudorandom generators for space-bounded computations," *Combinatorica*, vol. 12, no. 4, pp. 449–461, 1992.

[8] N. Nisan and A. Wigderson, "Hardness vs randomness," *J. Computer & Systems Sciences*, vol. 49, no. 2, pp. 149–167, 1994.

[9] M. Luby, B. Velickovic, and A. Wigderson, "Deterministic approximate counting of depth-2 circuits," in *Proceedings of the 2nd ISTCS*, 1993, pp. 18–24.

[10] M. Luby and B. Velickovic, "On deterministic approximation of DNF," *Algorithmica*, vol. 16, no. 4/5, pp. 415–433, 1996.

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

[12] P. Gopalan, R. Meka, and O. Reingold, "DNF sparsification and a faster deterministic counting algorithm," *Computational Complexity*, vol. 22, no. 2, pp. 275–310, 2013.

[13] E. Viola, "The Sum of d Small-Bias Generators Fools Polynomials of Degree d," *Computational Complexity*, vol. 18, no. 2, pp. 209–217, 2009.

[14] P. Gopalan, A. Klivans, R. Meka, D. Stefankovic, S. Vempala, and E. Vigoda, "An fptas for #knapsack and related counting problems," in *FOCS*, 2011, pp. 817–826.

[15] V. Kabanets and R. Impagliazzo, "Derandomizing polynomial identity tests means proving circuit lower bounds," Electronic Colloquium on Computational Complexity (ECCC), Tech. Rep. 9(55), 2002.

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

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

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

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

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

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

[22] J. Håstad, "On the size of weights for threshold gates," *SIAM Journal on Discrete Mathematics*, vol. 7, no. 3, pp. 484–492, 1994.

[23] V. V. Podolskii, "Perceptrons of large weight," *Problems of Information Transmission*, vol. 45, no. 1, pp. 46–53, 2009.

[24] A. A. Sherstov, "Halfspace matrices," *Computational Complexity*, vol. 17, no. 2, pp. 149–178, 2008.

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

[26] I. Diakonikolas, P. Harsha, A. Klivans, R. Meka, P. Raghavendra, R. A. Servedio, and L.-Y. Tan, "Bounding the average sensitivity and noise sensitivity of polynomial threshold functions," in *STOC*, 2010, pp. 533–542.

[27] D. Kane, "The Gaussian surface area and noise sensitivity of degree-d polynomial threshold functions," in *CCC*, 2010, pp. 205–210.

[28] D. M. Kane, "A structure theorem for poorly anticoncentrated gaussian chaoses and applications to the study of polynomial threshold functions," in *FOCS*, 2012, pp. 91–100.

[29] ——, "The correct exponent for the gotsman-linial conjecture," *CoRR*, vol. abs/1210.1283, 2012.

[30] Z. S. Karnin, Y. Rabani, and A. Shpilka, "Explicit dimension reduction and its applications," *SIAM J. Comput.*, vol. 41, no. 1, pp. 219–249, 2012.

[31] A. Kalai, A. Klivans, Y. Mansour, and R. Servedio, "Agnostically learning halfspaces," *SIAM Journal on Computing*, vol. 37, no. 6, pp. 1777–1805, 2008.

[32] S. Shalev-Shwartz, O. Shamir, and K. Sridharan, "Learning kernel-based halfspaces with the 0-1 loss," *SIAM J. Comput.*, vol. 40, no. 6, pp. 1623–1646, 2011.

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

[34] 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)*, 2012, pp. 729–746.

[35] H. Aziz, M. Paterson, and D. Leech, "Efficient algorithm for designing weighted voting games," in *IEEE Intl. Multitopic Conf.*, 2007, pp. 1–6.

[36] A. De, I. Diakonikolas, and R. A. Servedio, "The inverse shapley value problem," in *ICALP (1)*, 2012, pp. 266–277.

[37] I. Diakonikolas, D. M. Kane, and J. Nelson, "Bounded independence fools degree-2 threshold functions," in *Proc. 51st IEEE Symposium on Foundations of Computer Science (FOCS)*, 2010, pp. 11–20.

[38] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. Servedio, and E. Viola, "Bounded independence fools halfspaces," *SIAM J. on Comput.*, vol. 39, no. 8, pp. 3441–3462, 2010.

[39] R. Meka and D. Zuckerman, "Pseudorandom generators for polynomial threshold functions," in *STOC*, 2010, pp. 427–436.

[40] D. M. Kane, "k-independent gaussians fool polynomial threshold functions," in *IEEE Conference on Computational Complexity*, 2011, pp. 252–261.

[41] K. Matulef, R. O'Donnell, R. Rubinfeld, and R. Servedio, "Testing halfspaces," *SIAM J. on Comput.*, vol. 39, no. 5, pp. 2004–2047, 2010.

[42] R. O'Donnell and R. Servedio, "The Chow Parameters Problem," *SIAM J. on Comput.*, vol. 40, no. 1, pp. 165–199, 2011.

[43] E. Mossel, "Gaussian bounds for noise correlation of functions," *GAFA*, vol. 19, pp. 1713–1756, 2010.

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

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

[46] E. Mossel, "Gaussian bounds for noise correlation of functions and tight analysis of long codes," *FOCS*, pp. 156–165, 2008.

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

[48] R. Servedio, "Every linear threshold function has a low-weight approximator," *Comput. Complexity*, vol. 16, no. 2, pp. 180–209, 2007.

[49] I. Nourdin, "Lectures on gaussian approximations with malliavin calculus," Tech. Rep. http://arxiv.org/abs/1203.4147v3, 28 June 2012.