Electronic Colloquium on Computational Complexity, Report No. 173 (2013)

Efficient deterministic approximate counting for low-degree polynomial threshold functions

Anindya De*
Institute for Advanced Study

Rocco A. Servedio†
Columbia University

November 11, 2013

Abstract

We give a *deterministic* algorithm for approximately counting satisfying assignments of a degree-d polynomial threshold function (PTF). Given a degree-d input polynomial $p(x_1, \ldots, x_n)$ over $\mathbb{R}^n$ and a parameter $\epsilon > 0$, our algorithm approximates $\mathbf{Pr}_{x \sim \{-1,1\}^n}[p(x) \geq 0]$ to within an additive $\pm\epsilon$ in time $O_{d,\epsilon}(1) \cdot \mathrm{poly}(n^d)$. (Since it is NP-hard to determine whether the above probability is nonzero, any sort of efficient multiplicative approximation is almost certainly impossible even for randomized algorithms.) Note that the running time of our algorithm (as a function of n^d, the number of coefficients of a degree-d PTF) is a *fixed* polynomial. The fastest previous algorithm for this problem [Kan12b], based on constructions of unconditional pseudorandom generators for degree-d PTFs, runs in time $n^{O_{d,c}(1) \cdot \epsilon^{-c}}$ for all $c > 0$.

The key novel contributions of this work are

- A new multivariate central limit theorem, proved using tools from Malliavin calculus and Stein's Method. This new CLT shows that any collection of Gaussian polynomials with small eigenvalues must have a joint distribution which is very close to a multidimensional Gaussian distribution.
- A new decomposition of low-degree multilinear polynomials over Gaussian inputs. Roughly speaking we show that (up to some small error) any such polynomial can be decomposed into a bounded number of multilinear polynomials all of which have extremely small eigenvalues.

We use these new ingredients to give a deterministic algorithm for a Gaussian-space version of the approximate counting problem, and then employ standard techniques for working with low-degree PTFs (invariance principles and regularity lemmas) to reduce the original approximate counting problem over the Boolean hypercube to the Gaussian version.

As an application of our result, we give the first deterministic fixed-parameter tractable algorithm for the following moment approximation problem: given a degree-d polynomial $p(x_1, \ldots, x_n)$ over $\{-1, 1\}^n$, a positive integer k and an error parameter ϵ, output a $(1 \pm \epsilon)$-multiplicatively accurate estimate to $\mathbf{E}_{x \sim \{-1,1\}^n}[|p(x)|^k]$. Our algorithm runs in time $O_{d,\epsilon,k}(1) \cdot \mathrm{poly}(n^d)$.

*anindya@math.ias.edu. Work was partly done while the author was hosted by Oded Regev at NYU and partly while the author was a fellow at the Simons Institute, Berkeley. Partly supported by NSF grants CCF-1320188.

†rocco@cs.columbia.edu. Supported by NSF grants CCF-1115703 and CCF-1319788.

ISSN 1433-8092

1 Introduction

For decades a major research goal in computational complexity has been to understand the computational power of randomization – and perhaps to show that randomness does not actually augment the abilities of polynomial-time algorithms. Towards this end, an important research goal within unconditional derandomization has been the development of *deterministic approximate counting algorithms*. This line of research started with the work of Ajtai and Wigderson [AW85], who gave a sub-exponential time deterministic algorithm to approximately count the number of satisfying assignments of a constant-depth circuit. Since this early work many other classes of Boolean functions have been studied from this perspective, including DNF formulas, low-degree $GF[2]$ polynomials, linear threshold functions, and degree-2 polynomial threshold functions [LVW93, LV96, Tre04, GMR13, Vio09, GKM$^+$11, DDS13a, DDS13b].

In this paper we study the problem of deterministic approximate counting for *degree-d polynomial threshold functions* (PTFs). A degree-d PTF 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. In the special case where $d = 1$, degree-d PTFs are often referred to as *linear threshold functions* (LTFs). While LTFs and low-degree PTFs have been researched for decades (see e.g. [MK61, MTT61, MP68, Mur71, GHR92, Orp92, Hås94, Pod09] and many other works), they have recently been the focus of renewed research attention in fields such as concrete complexity theory [She08, She09, DHK$^+$10, Kan10, Kan12c, Kan12a, KRS12], learning theory [KKMS08, SSSS11, DOSW11, DDFS12], voting theory [APL07, DDS12] and others.

Our main result. The main contribution of this paper is to give a fixed polynomial time deterministic approximate counting algorithm for degree-d PTFs. We prove the following theorem:

Theorem 1. *There is a deterministic algorithm A with the following properties: Let A be given as input a degree-d polynomial p over $\{-1,1\}^n$ and an accuracy parameter $\epsilon > 0$. Algorithm A runs in time $O_{d,\epsilon}(1) \cdot \mathrm{poly}(n^d)$ and outputs a value $\tilde{v} \in [0,1]$ such that $\left|\tilde{v} - \mathbf{Pr}_{x \in \{-1,1\}^n}[p(x) \geq 0]\right| \leq \epsilon$.*

Note that the above result guarantees an *additive* approximation to the desired probability. While additive approximation is not as strong as multiplicative approximation, one should recall that the problem of determining whether $\mathbf{Pr}_{x \in \{-1,1\}^n}[p(x) \geq 0]$ is nonzero is well known to be NP-hard, even for degree-2 polynomials and even if all nonconstant monomials in p are restricted to have coefficients from $\{0,1\}$ (this can be shown via a simple reduction from Max-Cut). Thus no efficient algorithm, even allowing randomness, can give any multiplicative approximation to $\mathbf{Pr}_{x \sim \{-1,1\}^n}[p(x) \geq 0]$ unless NP $\subseteq$ RP. Given this, additive approximation is a natural goal.

Related work. Several previous works have given $\mathrm{poly}(n^d)$-time deterministic approximate counting algorithms for width-d DNF formulas (see e.g. [Tre04, LV96, GMR13] as well as the approach of [AW85] augmented with the almost t-wise independent distributions of [NN93], as discussed in [Tre04]). Degree-d PTFs are of course a broad generalization of width-d DNF formulas, and the algorithms for width-d DNFs referenced above do not extend to degree-d PTFs.

The $d = 1$ case for degree-d PTFs (i.e. LTFs) is qualitatively different from $d > 1$. For $d = 1$ the satisfiability problem is trivial, so one may reasonably hope for a multiplicatively $(1 \pm \epsilon)$-accurate deterministic approximate counting algorithm. Indeed such an algorithm, running in fully polynomial time $\mathrm{poly}(n, 1/\epsilon)$, was given by Gopalan *et al.* and Stefankovic *et al.* in [GKM$^+$11]. For $d \geq 2$, however, as noted above additive approximation is the best one can hope for, even for randomized algorithms. The only previous deterministic approximate counting results for degree-d PTFs for general d follow from known constructions of *unconditional pseudorandom generators* (PRGs) for degree-d PTFs. The first such construction was given by Meka and Zuckerman [MZ10], whose PRG yielded an $n^{O_d(1) \cdot \mathrm{poly}(1/\epsilon^d)}$-time deterministic approximate counting algorithm. Followup works by Kane [Kan11a, Kan11b, Kan12b] improved the parameters

of these PRGs, with the strongest construction from [Kan12b] (for PTFs over Gaussian inputs) giving a $n^{O_{d,c}(1)\cdot\epsilon^{-c}}$-time algorithm. Thus these prior works do not give a fixed polynomial-time algorithm.

For the special case of $d = 2$, in separate work [DDS13a] the authors have given a deterministic approximate counting algorithm for degree-2 PTFs that runs in time $\mathrm{poly}(n, 2^{\mathrm{poly}(1/\epsilon)})$. In [DDS13b] the authors extended the [DDS13a] result and gave an algorithm that does deterministic approximate counting for any $O(1)$-junta of degree-2 PTFs. As we explain in detail in the rest of this introduction, much more sophisticated techniques and analyses are required to obtain the results of the current paper for general d. These include a new central limit theorem in probability theory based on Malliavin calculus and Stein's method, and an intricate new decomposition procedure that goes well beyond the decomposition approaches employed in [DDS13a, DDS13b].

Our approach. The main step in proving Theorem 1 is to give a deterministic approximate counting algorithm for the *standard Gaussian distribution* $N(0,1)^n$ over $\mathbb{R}^n$ rather than the uniform distribution over $\{-1,1\}^n$. The key result that gives us Theorem 1 is the following:

Theorem 2. *There is a deterministic algorithm A with the following properties: Let A be given as input a degree-d polynomial p over $\mathbb{R}^n$ and an accuracy parameter $\epsilon > 0$. Algorithm A runs in time $O_{d,\epsilon}(1) \cdot \mathrm{poly}(n^d)$ and outputs a value $\tilde{v} \in [0,1]$ such that $\left|\tilde{v} - \mathbf{Pr}_{x\sim N(0,1)^n}[p(x) \geq 0]\right| \leq \epsilon$.*

Theorem 1 follows from Theorem 2 using the invariance principle of [MOO10] and the "regularity lemma" for polynomial threshold functions from [DSTW10]. The arguments that give Theorem 1 from Theorem 2 are essentially identical to the ones used in [DDS13a], so we omit them in this extended abstract (see the full version). In the rest of this introduction we describe the main ideas behind the proof of Theorem 2; as explained below, there are two main contributions.

First contribution: A new multivariate CLT. Our first contribution is a new multidimensional central limit theorem that we establish for r-tuples of degree-d Gaussian polynomials, i.e. r-tuples $(p_1(x), \ldots, p_r(x))$ where each p_i is a degree-d polynomial and $x \sim N(0,1)^n$. This CLT states that if each p_i has "small eigenvalues" (as defined at the start of Section 4), then the joint distribution converges to the multidimensional Normal distribution $\mathcal{G}$ over $\mathbb{R}^r$ whose mean and covariance match $(p_1, \ldots, p_r)$. The closeness here is with respect to "test functions" that have globally bounded second derivatives; see Theorem 19 for a detailed statement of our CLT. In Section 6 we use tools from mollification to go from the aforementioned kind of "closeness" to the kind of closeness which is required to analyze polynomial threshold functions.

Comparing with previous work, the degree-2 case [DDS13a] required a CLT for a single degree-2 Gaussian polynomial. The main technical ingredient of the [DDS13a] proof was a result of Chatterjee [Cha09]. [DDS13b] established the $d = 2$ case of our multidimensional CLT via a relatively straightforward analysis (requiring just basic linear algebra) of the central limit theorem from [NPR10]. We note that in the $d = 2$ case it is clear what is the correct notion of the eigenvalues of a degree-2 polynomial, namely the eigenvalues of the quadratic form. In contrast, it is far from clear what is the correct notion of the eigenvalues of a degree-d polynomial, especially since we require a notion that enables both a CLT and a decomposition as described later. (We note that the tensor eigenvalue definitions that are employed in [FW95, CS13, Lat06] do not appear to be suitable for our purposes.) Based on discussions with experts [Lat13, Nou13, Led13, Ole13], even the $d = 2$ version of our multidimensional CLT was not previously known, let alone the far more general version of the CLT which we establish in this work.

It is instructive to consider our CLT in the context of a result of Latala [Lat06], which shows that (a somewhat different notion of) tensor eigenvalues can be used to bound the growth of moments of degree-d Gaussian polynomials. However, the moment bounds that can be obtained from this approach are too weak to establish asymptotic normality [Lat13].

Like [DDS13b], in this paper we also use the central limit theorem from [NPR10] as a starting point. However, our subsequent analysis crucially relies on the fact that there is a geometry-preserving isomor-

phism between the space of symmetric tensors and multivariate Gaussian polynomials. This allows us to view Gaussian polynomials in terms of the associated tensors and greatly facilitates the use of language and tools from tensor algebra. To establish our condition for asymptotic normality, we make significant use of tensor identities from Malliavin calculus which were developed in the context of application to Stein's method (see [NP09, Nou12, NPR10]).

Second contribution: Decomposition. The second main contribution of this paper is a novel decomposition that lets us transform a multilinear degree-d Gaussian polynomial p into a polynomial of the form $h(A_1, \ldots, A_r)$, where (informally)

1. p and $h(A_1, \ldots, A_r)$ are ϵ-close (i.e. $\mathbf{E}[p] = \mathbf{E}[h(A_1, \ldots, A_r)]$ and $\mathbf{Var}[p - h(A_1, \ldots, A - r)] \leq \epsilon$);
2. For each polynomial A_i, all of its eigenvalues are extremely small (at most η for some very small η); and
3. $r = r(\epsilon, d, \eta)$ is independent of n and depends only on the approximation parameter ϵ, the eigenvalue bound η, and the degree d of p.

This decomposition is useful for the following reasons: Property (1) ensures that the distributions of p and of $h(A_1, \ldots, A_r)$ are close in cdf-distance, and thus to in order to do approximate counting of Gaussian satisfying assignments for p, it suffices to do approximate counting of Gaussian satisfying assignments for $h(A_1, \ldots, A_r)$. Property (2) ensures that we may apply our new CLT to the r-tuple of polynomials $A_1, \ldots, A_r$, and thus we may approximately count satisfying assignments to $h(A_1, \ldots, A_r) \geq 0$ by approximating the fraction of assignments that satisfy $h(\mathcal{G}_1, \ldots, \mathcal{G}_r)$ where $\mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_r)$ is the multidimensional Normal distribution given by our CLT. Finally, by Property (3), approximating $\mathbf{Pr}[h(\mathcal{G}_1, \ldots, \mathcal{G}_r) \geq 0]$ is a "constant-dimensional problem" (independent of n) so it is straightforward for a deterministic algorithm to approximate this probability in time independent of n.

We note that there is a subtlety here which requires significant effort to overcome. As we discuss in Remark 20, in order for our CLT to give a nontrivial bound it must be the case that the eigenvalue bound η is much smaller than $1/r$. Mimicking decomposition approaches previously used in literature [Ser07, MZ09, DSTW10] has the problem that they will necessarily make $r \geq 1/\eta$, thus rendering such decompositions useless for our purposes. (One exception is the decomposition procedure from [Kan11a] where a similar problem arises, but since the desired target conditions there are different from ours, that work uses a different approach to overcome the difficulty; we elaborate on this below.) In our context, achieving a decomposition such that $\eta \ll 1/r$ requires ideas that go beyond those used in previous decompositions, and is responsible for the large "constant-factor" overhead (captured by $O_{d,\epsilon}(1)$) in the overall running time bound.

At a very high level our decomposition is reminiscent of the regularity lemma for degree-d polynomials over $\{-1, 1\}^n$ that was given in [DSTW10], in that both procedures break a given degree-d input polynomial into a collection of "regular" polynomials, but as we now explain, this resemblance is a superficial one as there are many significant differences. First, in the [DSTW10] setting the given input polynomials are over $\{-1, 1\}^n$ while here the polynomials are over Gaussian space; this is a major distinction since the geometry of Gaussian space plays a fundamental role in our proofs and techniques. Second, the notion of "regularity" that is used is quite different between the two works; in [DSTW10] a polynomial is regular if all variable influences are small whereas here a polynomial is "regular" if all its "tensor eigenvalues" are small. (We subsequently refer to this new notion of regularity which is introduced and used in our work as *eigenregularity.*) Third, in [DSTW10] each "atomic step" of the decomposition is simply to restrict an individual input variable to $+1$ or -1, whereas in this paper the atomic "decomposition step" now involves an eigenvalue computation (to identify two lower-degree polynomials whose product is nontrivially correlated with the polynomial being decomposed). Finally, the [DSTW10] decomposition produces a decision

tree over input variables with restricted polynomials at the leaves, whereas in this paper we produce a *single degree-d polynomial* $h(A_1, \ldots, A_r)$ as the output of our decomposition.

Our decomposition has some elements that are reminiscent of a decomposition procedure described in [Kan11a]. Kane's procedure, like ours, breaks a degree-d polynomial into a sum of product of lower degree polynomials. However, there are significant differences between the procedures. Roughly speaking, Kane's decomposition starts with a polynomial p and is aimed at upper bounding the higher moments of the resulting constituent polynomials, whereas our decomposition is aimed at upper bounding the eigenregularity (magnitude of the largest eigenvalues) of the constituent polynomials. To make sure that the number r of constituent polynomials compares favorably with the moment bounds, Kane divides these polynomials into several classes such that the number of polynomials in any class compares favorably with the moment bounds in that class (and some desired relation holds between the number of polynomials in the different classes). Instead, in our decomposition procedure, we want r to compare favorably with the eigenvalue bound η; given this requirement, it does not seem possible to mimic Kane's approach of splitting the constituent polynomials into several classes. Instead, through a rather elaborate decomposition procedure, we show that while it may not be possible to split the original polynomial p in a way so that r compares favorably with η, it is always possible to (efficiently) find a polynomial $\tilde{p}$ such that $p - \tilde{p}$ has small variance, and $\tilde{p}$ can be decomposed so that the number of constituent polynomials compare favorably with the eigenregularity parameter.

We note that it is possible for the polynomial $p - \tilde{p}$ to have small variance but relatively huge moments. Thus our decomposition procedure is not effective for the approach in [Kan11a] which is based on bounding moments. However, because $p - \tilde{p}$ has small variance, the distributions of p and $\tilde{p}$ are indeed close in cdf distance, which suffices for our purposes. Thus our decomposition procedure should be viewed as incomparable to that of [Kan11a].

We also remark that our decomposition is significantly more involved than the decompositions used in [DDS13a, DDS13b]. To see how this additional complexity arises, note that both these papers need to decompose either a single degree-2 Gaussian polynomial or a set of such polynomials; for simplicity assume we are dealing with a single degree-2 polynomial p. Then the [DDS13a] decomposition procedure splits p into a sum of products of linear functions plus a degree-2 polynomial which has small eigenvalues. Crucially, since a linear function of Gaussians is itself a Gaussian, this permits a change of basis in which these linear functions may be viewed as the new variables. By "restricting" these new variables, one is essentially left with a single degree-2 polynomial with a small eigenvalue. In contrast, if p has degree d greater than 2, then the [DDS13a] decomposition will split p into a sum of products of pairs of lower degree Gaussian polynomials plus a polynomial which has small eigenvalues. However, if $d > 2$ then some or all of the new constituent lower degree polynomials may have degree greater than 1. Since a polynomial of degree $d > 1$ cannot itself be viewed as a Gaussian, this precludes the possibility of "restricting" this polynomial as was done in [DDS13a]. Thus, one has to resort to an iterative decomposition, which introduces additional complications some of which were discussed above.

Organization. We begin in Section 2 by recording various useful preliminaries, including some basics from the study of isonormal Gaussian processes (in the context of finite-degree Gaussian polynomials) that are required for the rest of the paper. In Section 3 we show that it is sufficient to give an algorithm for deterministic approximate counting of degree-d polynomials in the special case where all the polynomials are multilinear. In Section 4 we prove our new CLT for k-tuples of degree-d Gaussian polynomials with "small eigenvalues." In Section 5 we describe our decomposition procedure that can be used to decompose a k-tuple of degree-d multilinear polynomials over Gaussian inputs into an essentially equivalent k-tuple of polynomials that have a highly structured "special form." In Section 6 we show how the CLT from Section 4 can be combined with the highly structured polynomials from Section 5 to prove Theorem 2. In Section 7 we sketch how Theorem 1 follows from Theorem 2. We close in Section 8 by briefly describing how

Theorem 1 can be applied to give the first deterministic fixed-parameter tractable algorithm for the problem of multiplicatively approximating the k-th absolute moment of a degree-d polynomial over $\{-1,1\}^n$.

2 Preliminaries

2.1 Basic Definitions, Notation and Useful Background

For A a real $N \times N$ matrix we write $\|A\|_2$ to denote the operator norm $\|A\|_2 = \max_{0 \neq x \in \mathbb{R}^N} \frac{\|Ax\|_2}{\|x\|_2}$. Throughout the paper we write $\lambda_{\max}(A)$ to denote the largest-magnitude eigenvalue of a symmetric matrix A.

We will need the following standard concentration bound for low-degree polynomials over independent Gaussians.

Theorem 3 ("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*

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

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

Theorem 4 ([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*

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

On several occasions we will require the following lemma, which provides a sufficient condition for two degree-d Gaussian polynomials to have approximately the same fraction of satisfying assignments:

Lemma 5. *Let $a(x), b(x)$ be degree-d polynomials over $\mathbb{R}^n$. For $x \sim N(0,1)^n$, if $\mathbf{E}[a(x) - b(x)] = 0$, $\mathbf{Var}[a] = 1$ and $\mathbf{Var}[a - b] \leq (\tau/d)^{3d}$, then $\mathbf{Pr}_{x \sim N(0,1)^n}[\mathrm{sign}(a(x)) \neq \mathrm{sign}(b(x))] \leq O(\tau)$.*

Proof. The argument is a straightforward consequence of Theorems 3 and 4. First note that we may assume τ is at most some sufficiently small positive absolute constant since otherwise the claimed bound is trivial. By Theorem 4, we have $\mathbf{Pr}[|a(x)| \leq (\tau/d)^d] \leq O(\tau)$. Since $\mathbf{Var}[a - b] \leq (\tau/d)^{3d}$ and $a - b$ has mean 0, applying the tail bound given by Theorem 3, we get $\mathbf{Pr}[|a(x) - b(x)| > (\tau/d)^d] \leq O(\tau)$ (with room to spare, recalling that τ is at most some absolute constant). Since $\mathrm{sign}(a(x))$ can disagree with $\mathrm{sign}(b(x))$ only if $|a(x)| \leq (\tau/d)^d$ or $|a(x) - b(x)| > (\tau/d)^d$, a union bound gives that $\mathbf{Pr}_{x \sim N(0,1)^{n'}}[\mathrm{sign}(a(x)) \neq \mathrm{sign}(b(x))] = O(\tau)$, and the lemma is proved. □

2.2 A linear algebraic perspective

We will often view polynomials over $N(0,1)^n$ as elements from a linear vector space. In this subsection we make this correspondence explicit and establish some simple but useful linear algebra background results. In particular, we consider the finite-dimensional real vector space of all degree d polynomials over n variables. This vector space is equipped with an inner product defined as follows: for polynomials P, Q, we have $\langle P, Q \rangle = \mathbf{E}_{x \sim N(0,1)^n}[P(x) \cdot Q(x)]$. For the rest of the paper we let V denote this vector space.

We will need to quantify the linear dependency between polynomials; it will be useful for us to do this in the following way. Let W be a subspace of V and $v \in V$. We write $v^{\|W}$ to denote the projection of v on W and $v^{\perp W}$ to denote the projection of v on the space orthogonal to W, so $v = v^{\|W} + v^{\perp W}$. Equipped with this notation, we define the following (somewhat non-standard) notion of linear dependency for an ordered set of vectors:

Definition 6. *Let V be defined as above and let $\mathcal{A} = \{v_1, \ldots, v_m\}$ be an ordered set of unit vectors belonging to V. Define $\mathcal{A}_i = \{v_1, \ldots, v_i\}$ and define V^i to be the linear span of $\mathcal{A}_i$. $\mathcal{A}$ is said to be ζ-far* from being linearly dependent *if for every $1 < i \leq m$, we have $\|(v_i)^{\perp V_{i-1}}\|_2 \geq \zeta$.*

Note that viewing the vector space V as $\mathbb{R}^t$ (for some t), we can associate a matrix $M_{\mathcal{A}} \in \mathbb{R}^{t \times m}$ with $\mathcal{A}$ where the i^{th} column of $M_{\mathcal{A}}$ is v_i. The smallest non-zero singular value of this matrix is another measure of the dependency of the vectors in $\mathcal{A}$. Observe that this value (denoted by $\sigma_{\min}(M_{\mathcal{A}})$) can alternately be characterized as

$$\sigma_{\min}(M_{\mathcal{A}}) = \inf_{\alpha \in \mathbb{R}^m : \|\alpha\|_2 = 1} \left\| \sum_{i=1}^{m} \alpha_i v_i \right\|_2 .$$

We next have the following lemma which shows that if $\mathcal{A}$ is ζ-far from being linearly dependent, then the smallest non-zero singular value of $M_{\mathcal{A}}$ is noticeably large.

Lemma 7. *If $\mathcal{A}$ is ζ-far from being linearly dependent (where $\zeta \leq 1/4$), then $\sigma_{\min}(M_{\mathcal{A}}) \geq \zeta^{2m-2}$.*

Proof. We will prove this by induction on m, by proving a lower bound on $\| \sum_{i=1}^{m} \alpha_i v_i \|_2$ for any unit vector $\alpha \in \mathbb{R}^m$. For $m = 1$, the proof is obvious by definition. For the induction step, observe that

$$\left\| \sum_{i=1}^{m} \alpha_i v_i \right\|_2 \geq |\alpha_m| \cdot \left\| v_m^{V_{m-1}} \right\|_2$$

where we use the notation from Definition 6. If $|\alpha_m| \geq \zeta^{2m-3}$, then we get the stated bound on $\| \sum_{i=1}^{m} \alpha_i v_i \|_2$. In the other case, since $|\alpha_m| < \zeta^{2m-3}$, we have

$$\left\| \sum_{i=1}^{m} \alpha_i v_i \right\|_2 \geq \left\| \sum_{i=1}^{m-1} \alpha_i v_i \right\|_2 - |\alpha_m| \geq \left\| \sum_{i=1}^{m-1} \alpha_i v_i \right\|_2 - \zeta^{2m-3}.$$

However, by the induction hypothesis, we get

$$\left\| \sum_{i=1}^{m-1} \alpha_i v_i \right\|_2 \geq (1 - \zeta^{2m-3})\zeta^{2m-4} \geq \frac{\zeta^{2m-4}}{2}.$$

Thus, $\| \sum_{i=1}^{m} \alpha_i v_i \|_2 \geq \zeta^{2m-4}/2 - \zeta^{2m-3} \geq \zeta^{2m-2}$ (provided $\zeta \leq 1/4$). □

The next simple claim says that if $\mathcal{A}$ is ζ-far from being linearly dependent and v lies in the linear span of $\mathcal{A}$, then we can upper bound the size of the coefficients used to represent v.

Claim 8. *Let v be a unit vector which lies in the span of $\mathcal{A}$ and let $\mathcal{A}$ be ζ-far from being linearly dependent. Then, if $v = \sum_{i=1}^{m} \beta_i v_i$ is the unique representation of v as a linear combination of v_i's, we have $\sqrt{\sum_{i=1}^{m} \beta_i^2} \leq (1/\zeta)^{2m-2}$.*

Proof. Let $\gamma_i = \beta_i / \sqrt{\sum_{i=1}^{m} \beta_i^2}$. Since γ is a unit vector, by Lemma 7 we have that

$$\left\| \sum_{i=1}^{m} \gamma_i v_i \right\|_2 \geq \zeta^{2m-2}.$$

Thus $\zeta^{2m-2} \cdot \sqrt{\sum_{i=1}^{m} \beta_i^2} \leq 1$, giving the claimed upper bound. □

We will also need another simple fact which we state below.

Fact 9. *Let $\mathcal{A}_i$ be ζ-far from being linearly dependent. Let v_{i+1} and v be unit vectors such that $|\langle v, v_{i+1}\rangle| \geq \zeta$ and v is orthogonal to V_i. Then $\mathcal{A}_{i+1}$ is ζ-far from being linearly dependent.*

Proof. Note that $v_{i+1} = v_{i+1}^{\|V_i} + v_{i+1}^{\perp V_i}$ where $V_i = span(\mathcal{A}_i)$ (following Definition 6). Hence we have

$$\|v_{i+1}^{\perp V_i}\|_2 \geq |\langle v, v_{i+1}^{\perp V_i}\rangle| = |\langle v, v_{i+1}^{\|V_i}\rangle + \langle v, v_{i+1}^{\perp V_i}\rangle| = |\langle v, v_{i+1}\rangle| \geq \zeta,$$

where the first inequality is by Cauchy-Schwarz and the first equality uses that v is orthogonal to V_i. □

2.3 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 $|\tilde{\sigma}_1 - \sigma_1| \leq \epsilon$, 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 $|u^{(i)} \cdot u^{(j)} - \tilde{u}^{(i)} \cdot \tilde{u}^{(j)}| \leq 2^{-\mathrm{poly}(b)}$ 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. Note that the final guarantee we require from Theorem 2 is only that $\tilde{p}$ is an additively accurate approximation to the unknown probability. Note further that our Lemma 5 gives the following: for $p(x)$ a degree-d Gaussian polynomial with $\mathbf{Var}[p] = 1$, and $\tilde{p}(x)$ a degree-d polynomial so that for each fixed monomial the coefficients of p and $\tilde{p}$ differ by at most κ, then taking $\kappa = (\epsilon^3/(d^3 \cdot n))^d$, we have that $|\mathbf{Pr}[p(x) \geq 0] - \mathbf{Pr}[\tilde{p}(x) \geq 0]| \leq \epsilon$.

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 $\mathrm{poly}(n^d)$ many such approximation steps are performed by $\mathrm{poly}(n^d)$-time algorithms, the resulting approximate implementations of our algorithms in a bit-complexity model also achieve the guarantee of Theorem 2, at the cost of a fixed $\mathrm{poly}(n^d)$ overhead in the running time. Since working through the details of such an analysis is as tedious for the reader as it is for the authors, we content ourselves with this brief discussion.

2.4 Polynomials and tensors: some basics from isonormal Gaussian processes

We start with some basic background; a more detailed discussion of the topics we cover here can be found in [NP09, Nou12, NPR10].

Gaussian processes. A *Gaussian process* is a collection of jointly distributed random variables $\{X_t\}_{t\in T}$, where T is an index set, such that if $S \subset T$ is finite then the random variables $\{X_s\}_{s\in S}$ are distributed

as a multidimensional normal. Throughout this paper we will only deal with *centered* Gaussian processes, meaning that $\mathbf{E}[X_t] = 0$ for every $t \in T$. It is well known that a centered Gaussian process is completely characterized by its set of covariances $\{\mathbf{E}[X_s X_t]\}_{s,t \in T}$. An easy but important observation is that the function $d : T \times T \to \mathbb{R}^+$ defined by $d(s,t) = \sqrt{\mathbf{E}[(X_s - X_t)^2]}$ forms a (pseudo)-metric on the set T.

Isonormal Gaussian processes. For $\mathcal{H}$ any separable Hilbert space, the Gaussian process $\{X(h)\}_{h \in \mathcal{H}}$ over $\mathcal{H}$ is *isonormal* if it is centered and $\mathbf{E}[X(h) \cdot X(h')] = \langle h, h' \rangle$ for every $h, h' \in \mathcal{H}$. It is easy to see that for an isonormal Gaussian process, the metric d induced by the process $\{X(h)\}_{h \in \mathcal{H}}$ is the same as the metric on the Hilbert space $\mathcal{H}$, and thus there is an isomorphism between the Gaussian process $\{X(h)\}$ and the Hilbert space $\mathcal{H}$. This isomorphism allows us to reason about the process $\{X(h)\}$ using the geometry of $\mathcal{H}$.

Throughout this paper the Hilbert space $\mathcal{H}$ will be the n-dimensional Hilbert space $\mathbb{R}^n$, and we will consider the isonormal Gaussian process $\{X(h)\}_{h \in \mathcal{H}}$. This Gaussian process can be described explicitly as follows: We equip $\mathbb{R}^n$ with the standard normal measure $N(0,1)^n$, and corresponding to every element $h \in \mathbb{R}^n$ we define $X(h) = h \cdot x$, where $x \sim N(0,1)^n$. The resulting $\{X(h)\}$ is easily seen to be a Gaussian process with the property that $\mathbf{E}[X(h) \cdot X(g)] = \langle h, g \rangle$.

Tensors. We write $\mathcal{H}^{\otimes q}$ to denote the q-tensor product of $\mathcal{H}$. Fixing a basis $e_1, \ldots, e_n$ of $\mathcal{H}$, recall that every element of $\mathcal{H}^{\otimes q}$ may be uniquely written as

$$f = \sum_{i_1, \ldots, i_q = 1}^{n} f(i_1, \ldots, i_q) \cdot e_{i_1} \otimes \cdots \otimes e_{i_q} \quad \text{where } f(i_1, \ldots, i_q) \in \mathbb{R}. \tag{1}$$

We write $\mathcal{H}^{\odot q}$ to denote the subspace of symmetric q-tensors. These are the elements $f \in \mathcal{H}^{\otimes q}$ such that, writing f as in (1) above, we have $f(i_1, \ldots, i_q) = f(i_{\sigma(1)}, \ldots, i_{\sigma(q)})$ for every permutation $\sigma \in S_q$. Given a tensor $f \in \mathcal{H}^{\otimes q}$ written as in (1), we write $\tilde{f}$ to denote the symmetrization of f,

$$\tilde{f} \stackrel{\text{def}}{=} \frac{1}{q!} \sum_{\sigma \in S_q} \sum_{i_1, \ldots, i_q = 1}^{n} f(i_{\sigma(1)}, \ldots, i_{\sigma(q)}) \cdot e_{i_1} \otimes \cdots \otimes e_{i_q},$$

which lies in $\mathcal{H}^{\odot q}$; we alternately write $\mathrm{Sym}(f)$ for $\tilde{f}$ if the latter is typographically inconvenient.

Given an element $h \in \mathcal{H}$, we write $h^{\odot q}$ to denote the q-th tensor product of h with itself, $h \otimes \cdots \otimes h$. Note that this is a symmetric tensor; we sometimes refer to $h^{\odot q}$ as a *pure symmetric q-tensor*. Note that every symmetric q-tensor can be expressed as a finite linear combination of pure symmetric q-tensors $h_i^{\odot q}$ where each $h_i \in \mathcal{H}$ has $\|h_i\| = 1$ (but this representation is not in general unique).

We say that a tensor f as in (1) is *multilinear* if $f(i_1, \ldots, i_q) = 0$ whenever $i_a = i_b$ for any $a \neq b$, $a, b \in [q]$ (i.e. all the diagonal entries of f are zero).

From tensors to Gaussian polynomials and back. We write $H_q(x)$ to denote the q^{th} Hermite polynomial; this is the univariate polynomial

$$H_0(x) \equiv 1, H_1(x) = x, \quad H_q(x) = \frac{(-1)^q}{q!} e^{x^2/2} \frac{d^q}{dx^q} e^{-x^2/2}.$$

Note that these have been normalized so that that $\mathbf{E}_{x \sim N(0,1)}[H_q(x)^2] = 1/q!$. We write $\mathcal{W}^q$ to denote the q-th *Wiener chaos*; this is the space spanned by all random variables of the form $H_q(X(h))$ (intuitively, this is the span of all homogenous degree-q multivariate Hermite polynomials.). We note that it can be shown (see e.g. Section 2.2 of [NP09]) that for $h, g \in \mathcal{H}$ with $\|h\| = \|g\| = 1$ we have

$$\mathbf{E}[H_q(X(h)) \cdot H_q(X(g))] = \frac{1}{q!} \langle h^{\odot q}, g^{\odot q} \rangle. \tag{2}$$

The *iterated Ito integral* is a map which takes us from symmetric tensors to Gaussian polynomials as follows. Given $q \geq 1$ and $h \in \mathcal{H}$ which satisfies $\|h\| = 1$, we define

$$I_q(h^{\odot q}) = q! \cdot H_q(X(h)). \tag{3}$$

(We define $I_0(c) = c$ for $c \in \mathbb{R}$.) Note that with the definition of Ito integral, we can rephrase the guarantee of (2) as saying

$$\mathbf{E}[I_q(h^{\odot q}) \cdot I_q(g^{\odot q})] = q! \cdot \langle h^{\odot q}, g^{\odot q} \rangle.$$

So far, the map $I_q(\cdot)$ has been defined only for pure symmetric q-tensors of unit norm. However, equipped with the fact that every $x \in \mathcal{H}^{\odot q}$ can be written as a linear combination of such tensors, the map I_q can be linearly extended to the whole space $\mathcal{H}^{\odot q}$. Using the multiplication formula for Ito integrals (equation (5) below) and standard identities for Hermite polynomials, it can be shown (see again Section 2.2 of [NP09]) that such an extension is consistent and unambiguous, i.e. it does not depend on the particular linear combination of the pure symmetric tensors we use to represent $x \in \mathcal{H}$. Thus every element of $\mathcal{H}^{\odot q}$ maps to an element of $\mathcal{W}^q$ and further this mapping can be shown to be bijective. In fact, the map I_q is an isomorphism (up to scaling) between the space $\mathcal{W}^q$ of Gaussian chaoses and the Hilbert space $\mathcal{H}^{\odot q}$, as is shown by the following relation: for $f, g \in \mathcal{H}^{\odot q}$, we have

$$\mathbf{E}[I_q(f) \cdot I_q(g)] = q! \cdot \langle f, g \rangle$$

(ee Claim 11 below for a proof). This relation forges a connection between the q-th Wiener chaos $\mathcal{W}^q$ and the geometry of the space $\mathcal{H}^{\odot q}$. This connection is crucial for us as we extensively use operations in the space $\mathcal{H}^{\odot q}$ to reason about the Wiener chaos $\mathcal{W}^q$.

Let $F = F(x_1, \ldots, x_n)$ be any degree-d Gaussian polynomial over $\mathcal{H} = \mathbb{R}^n$. Since $\mathbf{E}[F^2] < \infty$, the Wiener chaos decomposition implies that there exists a unique sequence $f_0, \ldots, f_d$ where $f_q \in \mathcal{H}^{\odot q}$ such that

$$F = \sum_{q=0}^{d} I_q(f_q), \tag{4}$$

where by convention $I_0(f_0) = \mathbf{E}[F]$. Moreover, this decomposition is easily seen to be effective, in the sense that given F we can deterministically construct the tensors $f_0, \ldots, f_d$ in time $n^{O(d)}$. In a bit more detail, let J_q be the operator which maps $F : \mathbb{R}^n \to \mathbb{R}$ (with $\mathbf{E}[F^2] < \infty$) to its projection on the space $\mathcal{W}^q$. If F is explicitly presented as a polynomial of degree d, then the action of each operator $J_1, \ldots, J_d$ can easily be computed in time $n^{O(d)}$, and given an explicit representation (as a polynomial) of any point F_q in the image of J_q, it is straightforward to compute f_q (in time $n^{O(q)}$) such that $I_q(f_q) = F_q$. (In the other direction it is also straightforward, given $f_q \in \mathcal{H}^{\odot q}$ for $q = 0, \ldots, d$, to output the degree-d polynomial F satisfying (4).)

Remark 10. *If F is a multilinear degree-d Gaussian polynomial over $\mathcal{H}$, then it can be shown that in the Wiener chaos decomposition $F = I_0(f_0) + \cdots + I_d(f_d)$, each f_q is a multilinear symmetric tensor. Conversely, if $f_q \in \mathcal{H}^{\odot q}$ is a multilinear symmetric tensor then it can be shown that $I_q(f_q)$ is a multilinear Gaussian polynomial.*

2.5 Some background results from isonormal Gaussian processes.

We begin with a simple claim about the inner product between Ito integrals:

Claim 11. ***[Inner product between Ito integrals.]*** *Let $f \in \mathcal{H}^{\odot p}$ and $g \in \mathcal{H}^{\odot q}$. Then*

$$\mathbf{E}[I_p(f) \cdot I_q(g)] = \begin{cases} 0 & \text{if } p \neq q, \\ p! \cdot \langle f, g \rangle & \text{if } p = q. \end{cases}$$

Proof. For $p \neq q$ the claim follows from the fact that different levels of Wiener chaos are orthogonal to each other. For $p = q$, we may write f, g in terms of pure symmetric tensors as $f = \sum_{i=1}^{t} \alpha_i \cdot f_i^{\odot p}$, $g = \sum_{j=1}^{t} \beta_j \cdot g_j^{\odot p}$, and hence

$$\begin{aligned}
\mathbf{E}[I_p(f) \cdot I_p(g)] &= \sum_{i,j=1}^{t} \alpha_i \beta_j \, \mathbf{E}[I_p(f_i^{\odot p}) \cdot I_p(g_j^{\odot p})] \\
&= (p!)^2 \sum_{i,j=1}^{t} \alpha_i \beta_j \, \mathbf{E}[H_p(X(f_i)) \cdot H_p(X(g_j))] \\
&= p! \sum_{i,j=1}^{t} \alpha_i \beta_j \langle f_i^{\odot p}, g_j^{\odot p} \rangle \\
&= p! \langle f, g \rangle,
\end{aligned}$$

where the first equality is by linearity of $I_p(\cdot)$, the second is by (3), the third is by (2), and the fourth is by the bilinearity of $\langle \cdot, \cdot \rangle$. □

As a consequence of this we get the following useful fact:

Fact 12. *Let $f_1, \ldots$ be symmetric tensors where $f_i \in \mathcal{H}^{\odot i}$. Then we have $\mathbf{Var}[\sum_i I_i(f_i)] = \sum_i \mathbf{Var}[I_i(f_i)]$.*

Proof. All the random variables $I_i(f_i)$ are centered for $i \geq 1$, so it suffices to show that $\mathbf{E}[(\sum_i I_i(f_i))^2] = \sum_i \mathbf{E}[I_i(f_i)^2]$. This follows directly from Claim 11. □

Contraction products. Consider symmetric tensors $f \in \mathcal{H}^{\odot q}$ and $g \in \mathcal{H}^{\odot r}$. For $0 \leq s \leq \min\{q, r\}$ we define the *s-th contraction product* $f \otimes_s g \in \mathcal{H}^{\otimes q+r-2s}$ as follows:

$$(f \otimes_s g)_{(t_1, t_2, \ldots, t_{q+r-2s})} = \sum_{i_1, \ldots, i_s}^{\infty} \langle f, e_{i_1} \otimes \ldots \otimes e_{i_s} \rangle \otimes \langle g, e_{i_1} \otimes \ldots \otimes e_{i_s} \rangle.$$

One way to visualize the contraction product is as a matrix multiplication. We may view $f \in \mathcal{H}^q$ as a matrix $f_{q-s,s}$ where the rows of $f_{q-s,s}$ are identified with the elements of $[n]^{q-s}$ and the columns with the elements of $[n]^s$, and we may likewise view $g \in \mathcal{H}^r$ as an $[n]^s \times [n]^{r-s}$ matrix. A matrix multiplication between $f_{q-s,s}$ and $g_{s,r-s}$ results in a matrix of dimension $[n]^{q-s} \times [n]^{r-s}$, which can be viewed as an element of $\mathcal{H}^{\otimes q+r-2s}$; this element is the s-th contraction product $f \otimes_s g$.

Note that the contraction product $f \otimes_s g$ of two symmetric tensors may not be a symmetric tensor. We write $f \tilde{\otimes}_s g$ to denote $\mathrm{Sym}(f \otimes_s g)$; the resulting symmetric tensor is an element of $\mathcal{H}^{\odot q+r-2s}$.

We will make heavy use of the following multiplication formula for Ito integrals (see p. 4 of [NPR10]):

Theorem 13. ***[Multiplication formula for Ito integrals.]*** *If $f \in \mathcal{H}^{\odot p}$ and $g \in \mathcal{H}^{\odot q}$, then*

$$I_p(f) \cdot I_q(g) = \sum_{r=0}^{\min\{p,q\}} r! \cdot \binom{p}{r} \binom{q}{r} I_{p+q-2r}(f \tilde{\otimes}_r g). \tag{5}$$

3 Dealing with non-multilinear polynomials

The decomposition procedure that we use relies heavily on the fact that the input polynomials p_i are multilinear. To handle general (non-multilinear) degree-d polynomials, the first step of our algorithm is to transform them to (essentially) equivalent multilinear degree-d polynomials. This is accomplished by a simple procedure whose performance is described in the following theorem.[1] Note that given Theorem 14, in subsequent sections we can (and do) assume that the polynomial p given as input in Theorem 2 is multilinear.

Theorem 14. *There is a deterministic procedure* **Linearize** *with the following properties: The algorithm takes as input a (not necessarily multilinear) variance-1 degree-d polynomial p over $\mathbb{R}^n$ and an accuracy parameter $\delta > 0$. It runs in time $O_{d,\delta}(1) \cdot poly(n^d)$ and outputs a multilinear degree-d polynomial q over $R^{n'}$, with $n' \leq O_{d,\delta}(1) \cdot n$, such that*

$$\left|\mathbf{Pr}_{x\sim N(0,1)^n}[p(x) \geq 0] - \mathbf{Pr}_{x\sim N(0,1)^{n'}}[q(x) \geq 0]\right| \leq O(\delta).$$

Proof. The procedure **Linearize** is given below. (Recall that a diagonal entry of a q-tensor $f = \sum_{i_1,\ldots,i_q=1} f(i_1,\ldots,i_q) \cdot e_{i_1} \otimes \cdots \otimes e_{i_q}$ is a coefficient $f(i_1,\ldots,i_q)$ that has $i_a = i_b$ for some $a \neq b$.)

Linearize

Input: A degree-d polynomial $p : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbf{Var}(p) = 1$.
Output: A degree-d multilinear polynomial $q : \mathbb{R}^{n'} \to \mathbb{R}$ such that

$$\left|\mathbf{Pr}_{x\sim N(0,1)^n}[p(x) \geq 0] - \mathbf{Pr}_{x\sim N(0,1)^{n'}}[q(x) \geq 0]\right| \leq \delta.$$

where $n' = n \cdot K$.

1. Let $p(x_1,\ldots,x_n) = \sum_{j=0}^{d} I_j(f_j)$ be the Wiener chaos decomposition of p. Let $K = d^2 \cdot (d/\delta)^{3d}$.
2. Construct polynomial $\tilde{q} : \mathbb{R}^{n'} \to \mathbb{R}$ from p by replacing each x_i by $(y_{i,1} + \ldots + y_{i,K})/\sqrt{K}$.
3. Let $\tilde{q} = \sum_{j=0}^{d} I_j(\tilde{f}_j)$ be the Wiener chaos decomposition of $\tilde{q}$, where $\tilde{f}_j \in \mathcal{H}_{n'}^{\odot q}$ and $\mathcal{H}_{n'} = \mathbb{R}^{n'}$.
4. For each $0 \leq j \leq d$, obtain g_j from $\tilde{f}_j$ by zeroing out all the diagonal entries from $\tilde{f}_j$.
5. Output $q = \sum_{j=0}^{d} I_j(g_j)$.

It is clear that all the tensors g_j are multilinear, so by Remark 10 the polynomial q that the procedure **Linearize** outputs is multilinear. The main step in proving Theorem 14 is to bound the variance of $\tilde{q} - q$, so we will establish the following claim:

Claim 15. $\mathbf{Var}[\tilde{q} - q] \leq \frac{d}{K} \cdot \mathbf{Var}[\tilde{q}]$.

Proof. We first observe that

$$\mathbf{Var}[\tilde{q} - q] = \mathbf{Var}\left[\sum_{j=1}^{d}(I_j(\tilde{f}_j) - I_j(g_j))\right] = \sum_{j=1}^{d}\mathbf{E}\left[(I_j(\tilde{f}_j) - I_j(g_j))^2\right], \tag{6}$$

[1] A similar "multilinearization" procedure is analyzed in [Kan11a], but since the setting and required guarantees are somewhat different here we give a self-contained algorithm and analysis.

where the first equality is because $g_0 = f_0$ and the second is by Claim 11 and and the fact that each $I_j(\tilde{f}_j), I_j(g_j)$ has mean 0 for $j \geq 1$. Now fix a value $1 \leq j \leq d$. Since each g_j is obtained from $\tilde{f}_j$ by zeroing out diagonal elements, again using Claim 11 we see that

$$\mathbf{E}\left[(I_j(\tilde{f}_j) - I_j(g_j))^2\right] = j! \cdot \|\tilde{f}_j - g_j\|_F^2, \tag{7}$$

where the squared Frobenius norm $\|\tilde{f}_j - g_j\|_F^2$ equals the sum of squared entries of the tensor $\tilde{f}_j - g_j$. Now,observe that the entry $\alpha_{i_1,\ldots,i_j} = f_j(i_1,\ldots,i_j)$ of the tensor f_j maps to the entry

$$\alpha_{i_1,\ldots,i_j} \frac{(e_{i_1,1} + \ldots + e_{i_1,K})}{\sqrt{K}} \otimes \ldots \otimes \frac{(e_{i_j,1} + \ldots + e_{i_j,K})}{\sqrt{K}} = \alpha_{i_1,\ldots,i_j} \cdot \sum_{(\ell_1,\ldots,\ell_j)\in[K]^j} \frac{1}{K^{j/2}} \otimes_{a=1}^{j} e_{i_a,\ell_a}$$

when $\tilde{q}$ is constructed from p. Further observe that all K^j outcomes of $\otimes_{a=1}^{j} e_{i_a,\ell_a}$ are distinct. Since g_j is obtained by zeroing out the diagonal entries of $\tilde{f}_j$, we get that

$$\|\tilde{f}_j - g_j\|_F^2 = \sum_{(i_1,\ldots,i_j)\in[n]^j} (\alpha_{i_1,\ldots,i_j})^2 \cdot \frac{1}{K^j} \cdot |\mathcal{S}_{K,j}|$$

where the set $\mathcal{S}_{K,j} = \{(\ell_1,\ldots,\ell_j) \in [K]^j : \ell_1,\ldots,\ell_j \text{ are not all distinct}\}$. It is easy to see that $|\mathcal{S}_{K,j}| \leq (j^2 \cdot K^j)/K$, so we get

$$\|\tilde{f}_j - g_j\|_F^2 \leq \sum_{(i_1,\ldots,i_j)\in[n]^j} (\alpha_{i_1,\ldots,i_j})^2 \cdot \frac{j^2}{K}.$$

Returning to (6) and (7), this yields

$$\mathbf{Var}[q - \tilde{q}] \leq \sum_{j=1}^{d} j! \cdot \sum_{(i_1,\ldots,i_j)\in[n]^j} (\alpha_{i_1,\ldots,i_j})^2 \cdot \frac{j^2}{K} \leq \frac{d^2}{K} \cdot \left(\sum_{j=1}^{d} j! \cdot \sum_{(i_1,\ldots,i_j)\in[n]^j} (\alpha_{i_1,\ldots,i_j})^2\right).$$

Using Fact 12 and Claim 11, we see that

$$\mathbf{Var}[p] = \sum_{j=1}^{d} \mathbf{Var}[I_j(f_j)] = \sum_{j=1}^{d} \mathbf{E}[I_j(f_j)^2] = \sum_{j=1}^{d} j! \cdot \sum_{(i_1,\ldots,i_j)\in[n]^j} (\alpha_{i_1,\ldots,i_j})^2.$$

It is easy to see that $\mathbf{Var}[\tilde{q}] = \mathbf{Var}[p]$, which equals 1 by assumption, so we have that $\mathbf{Var}[q - \tilde{q}] \leq \frac{d^2}{K} \cdot \mathbf{Var}[\tilde{q}]$ as desired. □

To finish the proof of Theorem 14, observe that by our choice of K we have $\mathbf{Var}[q - \tilde{q}] \leq (\delta/d)^{3d} \cdot \mathbf{Var}[\tilde{q}]$. Since $q - \tilde{q}$ has mean 0 and $\mathbf{Var}[\tilde{q}] = 1$ we may apply Lemma 5, and we get that $|\mathbf{Pr}_{x\sim N(0,1)^{n'}}[q(x) \geq 0] - \mathbf{Pr}_{x\sim N(0,1)^{n'}}[\tilde{q}(x) \geq 0]| \leq O(\delta)$. The theorem follows by observing that the two distributions $p(x)_{x\sim N(0,1)^n}$ and $\tilde{q}(x)_{x\sim N(0,1)^{n'}}$ are identical. □

4 A multidimensional CLT for low-degree Gaussian polynomials

Our goal in this section is to prove a CLT (Theorem 19 below) which says, roughly, the following: Let $F_1,\ldots,F_r$ be eigenregular low-degree Gaussian polynomials over $\mathbb{R}^n$ (here the meaning of "eigenregular" is that the polynomial has "small eigenvalues"; more on this below). Then the distribution of $(F_1,\ldots,F_r)$ is close — as measured by test functions with absolutely bounded second derivatives — to the r-dimensional Normal distribution with matching mean and covariance.

To make this statement more precise, let us begin by explaining what exactly is meant by the eigenvalues of a polynomial – this is clear enough for a quadratic polynomial, but not so clear for degrees 3 and higher.

Eigenvalues of tensors and polynomials. We begin by defining the largest eigenvalue of a symmetric tensor.

Definition 16. *For any $p \geq 2$ and $g \in \mathcal{H}^{\odot p}$, define $\lambda_{\max}(g)$, the* largest-magnitude eigenvalue *of g, as follows. Consider a partition of $[p]$ into S and $\overline{S} = [p] \setminus S$ where both S and $\overline{S}$ are non-empty.*[2] *We define*

$$\lambda_{S,\overline{S}}(g) = \sup_{x \in \mathcal{H}^S, y \in \mathcal{H}^{\overline{S}}} \frac{\langle g, x \otimes y \rangle}{\|x\|_F \cdot \|y\|_F} \qquad \text{and} \qquad \lambda_{\max}(g) = \max_{S,\overline{S} \neq \emptyset} \lambda_{S,\overline{S}}(g).$$

(Here $\|x\|_F$ denotes the Frobenius norm of x.) For $p \in \{0,1\}$ and $g \in \mathcal{H}^{\odot p}$ we say that $\lambda_{\max}(g) = 0$.

Fix a Gaussian polynomial F of degree d, and recall that F has a unique Wiener chaos decomposition as $F = \sum_{q=0}^{d} I_q(f_q)$ with $f_q \in \mathcal{H}^{\odot q}$. The following definition plays a crucial role in the rest of the paper.

Definition 17. *We define the* largest-magnitude eigenvalue *of F to be*

$$\lambda_{\max}(F) = \max\{\lambda_{\max}(f_2), \ldots, \lambda_{\max}(f_d)\}.$$

*We say that F is ϵ-*eigenregular *if $\frac{\lambda_{\max}(F)}{\sqrt{\mathbf{Var}[F]}} \leq \epsilon$, and we sometimes refer to $\frac{\lambda_{\max}(F)}{\sqrt{\mathbf{Var}[F]}}$ as the* eigenregularity of F.

Remark 18. *If F is a polynomial of degree at most 1 then we say that the polynomial F is 0-eigenregular (and hence ϵ-eigenregular for every $\epsilon > 0$).*

Now we can give a precise statement of our new CLT:

Theorem 19. *Fix $d \geq 2$ and let $F = (F_1, \ldots, F_r)$ be Gaussian polynomials over $\mathbb{R}^n$, each of degree at most d, such that for each i we have $\mathbf{E}[F_i] = 0$, $\mathbf{Var}[F_i] \leq 1$ and F_i is ϵ-eigenregular. Let C denote the covariance matrix of F, so $C(i,j) = \mathrm{Cov}(F_i, F_j) = \mathbf{E}_{x \sim N(0,1)^n}[F_i(x)F_j(x)]$. Let $\mathcal{G}$ be a mean-zero r-dimensional Gaussian random variable with covariance matrix C. Then for any $\alpha : \mathbb{R}^r \to \mathbb{R}, \alpha \in \mathcal{C}^2$ such that all second derivatives of α are at most $\|\alpha''\|_\infty < \infty$, we have*

$$|\mathbf{E}[\alpha(F_1, \ldots, F_r)] - \mathbf{E}[\alpha(\mathcal{G})]| < 2^{O(d \log d)} \cdot r^2 \cdot \sqrt{\epsilon} \cdot \|\alpha''\|_\infty.$$

The rest of Section 4 is dedicated to the proof of Theorem 19. The proof of Theorem 19 is somewhat involved, using Malliavin calculus in the context of Stein's method; it builds on recent work by Nourdin, Peccati and Réveillac [NP09, Nou12, NPR10]. In Section 4.1 we first give the necessary background ingredients from Malliavin calculus which will serve as the tools in our proof, and in Section 4.2 we give our proof of Theorem 19.

Remark 20. *It is clear from the statement of Theorem 19 that in order for the theorem to yield a meaningful bound, it must be the case that the number of polynomials r is small compared to $1/\sqrt{\epsilon}$. Looking ahead, in our eventual application of Theorem 19, the r polynomials $F_1, \ldots, F_r$ will be obtained by applying the decomposition procedure described in Section 5 to the original degree-d input polynomial. Thus it will be crucially important for our decomposition procedure to decompose the original polynomial into r polynomials all of which are* extremely *eigenregular, in particular ϵ-eigenregular for a value $\epsilon \ll 1/r^2$. Significant work will be required in Section 5 to surmount this challenge.*

[2] (Note that since we are only dealing with symmetric tensors we could equivalently have considered only partitions into $[1, \ldots, k]$, $[k+1, \ldots, p]$ where $1 \leq k \leq p-1$.)

4.1 Background from Malliavin calculus.

Malliavin derivatives. Let $F = f(X(h_1), \ldots, X(h_m))$ where $h_1, \ldots, h_m \in \mathcal{H}$ and f is a differentiable function. The *Malliavin derivative* is a $\mathcal{H}$ valued random variable defined as

$$DF \stackrel{\text{def}}{=} \sum_{i=1}^{n} \frac{\partial f(X(h_1), \ldots, X(h_n))}{\partial x_i} h_i.$$

Note that if $F = f(x_1, \ldots, x_n)$ (i.e. $m = n$ and h_i is the canonical basis vector $e_i \in \mathbb{R}^n$) then we have

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

where as usual we have $x \sim N(0,1)^n$.

Our proof of Theorem 19 will involve a lot of manipulation of inner products of Malliavin derivatives. The following results will be useful:

Claim 21. *[[NPR10]] Let $q \geq p$ and $f \in \mathcal{H}^{\odot p}$ and $g \in \mathcal{H}^{\odot q}$. Let $F = I_p(f)$ and $G = I_q(g)$.*

$$\langle DF, DG \rangle = pq \sum_{r=1}^{\min\{p,q\}} (r-1)! \binom{p-1}{r-1} \binom{q-1}{r-1} I_{p+q-2r}(f \widetilde{\otimes}_r g)$$

Theorem 22. *[[NPR10]] Let $q \geq p$ and $f \in \mathcal{H}^{\odot p}$ and $g \in \mathcal{H}^{\odot q}$. Let $F = I_p(f)$ and $G = I_q(g)$.*

$$\mathbf{E}[\langle DF, DG \rangle^2] = p^2 q^2 \sum_{r=1}^{p} (r-1)!^2 \binom{p-1}{r-1}^2 \binom{q-1}{r-1}^2 (p+q-2r)! \|f \widetilde{\otimes}_r g\|^2 \qquad \text{if } p < q$$

$$\mathbf{E}[\langle DF, DG \rangle^2] = p^2 p!^2 \langle f, g \rangle^2 + p^4 \sum_{r=1}^{p-1} (r-1)!^2 \binom{p-1}{r-1}^4 (2p-2r)! \|f \widetilde{\otimes}_r g\|^2 \qquad \text{if } p = q$$

$$\mathbf{E}[\langle DF, DG \rangle^2] = p^2 q^2 \sum_{r=1}^{q} (r-1)!^2 \binom{p-1}{r-1}^2 \binom{q-1}{r-1}^2 (p+q-2r)! \|f \widetilde{\otimes}_r g\|^2 \qquad \text{if } p > q.$$

(The last equality above is not explicitly stated in [NPR10] but it follows easily from their proof of the first equality; see Equation 3.12 in the proof of Lemma 3.7 of [NPR10].)

We recall (see [NPR10, NP09]) that the operator L (which is called the generator of the Ornstein-Uhlenbeck semigroup) is defined by

$$LF = \sum_{q=0}^{\infty} -q J_q(F).$$

We also recall the that the *pseudo-inverse of L* is defined to be the operator

$$L^{-1} F = \sum_{q=1}^{\infty} -J_q(F)/q.$$

Both operators are well-defined for all finite-degree Gaussian polynomials F.

We recall the following key identity which provides the fundamental connection between Malliavin Calculus and Stein's method:

Claim 23 (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 24. *Let p and q be finite degree polynomials over $\mathcal{X}$ with $\mathbf{E}[q] = 0$. Then, $\mathbf{E}[qp] = \mathbf{E}[\langle Dp\,,\, -DL^{-1}q \rangle]$.*

4.2 Proof of Theorem 19

We recall the following CLT due to Nourdin and Peccati:

Theorem 25. *[[NP09], see also [Nou12], Theorem 6.1] Let $F = (F_1, \ldots, F_r)$ where each F_i is a Gaussian polynomial with $\mathbf{E}[F_i] = 0$. Let C be a symmetric PSD matrix in $\mathbb{R}^{r\times r}$ and let $\mathcal{G}$ be a mean-0 r-dimensional Gaussian random variable with covariance matrix C. Then for any $\alpha : \mathbb{R}^r \to \mathbb{R}, \alpha \in \mathcal{C}^2$ such that $\|\alpha''\|_\infty < \infty$, we have*

$$|\mathbf{E}[\alpha(F)] - \mathbf{E}[\alpha(\mathcal{G})]| < \frac{1}{2}\|\alpha''\|_\infty \cdot \left(\sum_{i=1}^{r}\sum_{j=1}^{r} \mathbf{E}[|C(i,j) - Y(i,j)|]\right)$$

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

We now use Theorem 25 to prove Theorem 19.

Proof. As in Theorem 25, we write $Y(a,b)$ to denote $\langle DF_a, -DL^{-1}F_b\rangle$. For any $1 \le a, b \le r$, we have

$$C(a,b) = \mathrm{Cov}(F_a, F_b) = \mathbf{E}[F_aF_b] = \mathbf{E}[Y(a,b)], \tag{8}$$

where the second equality is because F_a and F_b have mean 0 and the third equality is by Corollary 24. Since C is a covariance matrix and every covariance matrix is PSD, we may apply Theorem 25, and we get that

$$|\mathbf{E}[\alpha(F)] - \mathbf{E}[\alpha(\mathcal{G})]| < \frac{r^2}{2}\|\alpha''\|_\infty \cdot \max_{1\le a,b\le r} \mathbf{E}[|C(a,b) - Y(a,b)|] = \frac{r^2}{2}\|\alpha''\|_\infty \cdot \max_{1\le a,b\le r} \mathbf{E}[|Y(a,b) - \mathbf{E}[Y(a,b)]|],$$

where we used (8) for the equality. By Jensen's inequality we have $\mathbf{E}[|Y(a,b) - \mathbf{E}[Y(a,b)]|] \le \sqrt{\mathbf{Var}[Y(a,b)]}$. Lemma 26 below gives us that $\mathbf{Var}[Y(a,b)] \le 2^{O(d\log d)}\epsilon$, and the theorem is proved. □

So to prove Theorem 19, it remains to establish the following lemma:

Lemma 26. *For each $1 \le a, b \le k$, we have that $\mathbf{Var}[Y(a,b)] = 2^{O(d\log d)}\epsilon$ where $Y(a,b) = \langle DF_a, -DL^{-1}F_b\rangle$.*

4.2.1 Proof of Lemma 26

We begin with the following useful facts about contraction products:

Fact 27. *Let $h \in \mathcal{H}^{\odot q_1}$, $g \in \mathcal{H}^{\odot q_2}$ where $q_1 \ge q_2$. Then for $1 \le r \le \min\{q_1 - 1, q_2\}$, we have $\|h\widetilde{\otimes}_r g\| \le \lambda_{\max}(h)\|g\|$.*

Proof. We first observe that the range of allowed values on r ensures that the contraction product $h\widetilde{\otimes}_r g$ is well defined. Next, we note that since symmetrizing can never increase the norm of a tensor, we have $\|h\widetilde{\otimes}_r g\|^2 \le \|h\otimes_r g\|^2$. As mentioned in our earlier discussion about contraction products we may view h as an $[n]^{q_1-r} \times [n]^r$ matrix H and g as an $[n]^r \times [n]^{q_2-r}$ matrix G with columns G_i. Since $1 \le r \le q_1 - 1$ the matrix H is non-degenerate (neither a single row nor a single column), and we have

$$\|h\otimes_r g\|^2 = \|HG\|_F^2 = \sum_i \|HG_i\|_2^2 \le \sum_i \lambda_{\max}(h)^2\|G_i\|_2^2 = \lambda_{\max}(h)^2\|G\|_2^2 = \lambda_{\max}(h)^2\|g\|^2$$

as claimed. □

Fact 28. *Fix $a \in \mathcal{H}^{\odot q_1}$, $b \in \mathcal{H}^{\odot q_2}$ where $q_1 \ge q_2$ and $c \in \mathcal{H}^{\odot q_3}$, $d \in \mathcal{H}^{\odot q_4}$ where $q_3 \ge q_4$. Then for $1 \le r_1 \le \min\{q_1 - 1, q_2\}$ and $1 \le r_2 \le \{q_3 - 1, q_4\}$ satisfying $q_1 + q_2 - 2r_1 = q_3 + q_4 - 2r_2$, we have $\langle a\widetilde{\otimes}_{r_1} b, c\widetilde{\otimes}_{r_2} d\rangle \le \lambda_{\max}(a)\lambda_{\max}(c) \cdot \|b\| \cdot \|d\|$.*

Proof. By Cauchy-Schwarz we have that

$$\langle a\widetilde{\otimes}_{r_1} b, c\widetilde{\otimes}_{r_2} d\rangle \leq \|a\widetilde{\otimes}_{r_1} b\| \cdot \|c\widetilde{\otimes}_{r_2} d\|,$$

and using Fact 27 twice this is at most the claimed bound. □

Fix $a, b \in [k]$. We may write

$$F_a = \sum_{q=1}^{d} I_q(a_q) \qquad \text{and} \qquad F_b = \sum_{q=1}^{d} I_q(b_q)$$

where each $a_q, b_q \in \mathcal{H}^{\odot q}$, and by assumption each $2 \leq q \leq d$ has $\lambda_{\max}(a_q), \lambda_{\max}(b_q) \leq \epsilon$. (Note that there is no contribution of the form $I_0(a_0)$ because by assumption we have $\mathbf{E}[F_a] = 0$ and $\mathbf{E}[I_q(a_q)] = 0$ for $q > 0$, and likewise for b.) Recall also that by assumption we have $\mathbf{Var}[F_a] \leq 1$, and hence $\mathbf{E}[F_a^2] \leq 1$. Using Claim 11, we have that

$$\mathbf{E}[F_a^2] = \mathbf{E}\left[\left(\sum_{q=1}^{d} I_q(a_q)\right)^2\right] = \sum_{q=1}^{d} \mathbf{E}[I_q(a_q)^2] = \sum_{q=1}^{d} q! \cdot \langle a_q, a_q\rangle \leq 1,$$

which immediately implies that

$$\|a_q\|^2 \leq \frac{1}{q!} \quad \text{for all } q \in [d] \text{ (and likewise } \|b_q\|^2 \leq \frac{1}{q!}\text{).} \tag{9}$$

Recall that $\mathbf{Var}[Y(a,b)] = \mathbf{E}[Y(a,b)^2] - \mathbf{E}[Y(a,b)]^2$. We begin by giving a compact expression for $\mathbf{E}[Y(a,b)]^2$ as follows:

$$\begin{aligned}
\mathbf{E}[Y(a,b)]^2 &= \mathbf{E}[\langle DF_a, -DL^{-1}F_b\rangle]^2 = \mathbf{E}[F_a F_b]^2 \quad \text{(by Corollary 24)} \\
&= \mathbf{E}\left[\left(\sum_{q=1}^{d} I_q(a_q)\right)\left(\sum_{q=1}^{d} I_q(b_q)\right)\right]^2 \\
&= \left(\sum_{q=1}^{d} q!\langle a_q, b_q\rangle\right)^2 \quad \text{(by linearity of expectation and Claim 11).}
\end{aligned} \tag{10}$$

Thus to prove Lemma 26 it suffices to show that

$$\mathbf{E}[Y(a,b)^2] \leq \left(\sum_{q=1}^{d} q!\langle a_q, b_q\rangle\right)^2 + 2^{O(d\log d)}\epsilon; \tag{11}$$

we do this below. We begin by writing

$$Y(a,b) = \langle DF_a, -DL^{-1}F_b\rangle = \left\langle \sum_{q=1}^{d} DI_q(a_q), \sum_{q=1}^{d} DI_q(b_q)/q \right\rangle = X + Y, \tag{12}$$

where

$$X = \sum_{q=1}^{d} \frac{1}{q}\langle DI_q(a_q), DI_q(b_q)\rangle \quad \text{and} \quad Y = \sum_{d \geq q_1 > q_2 \geq 1} \left(\frac{1}{q_1} + \frac{1}{q_2}\right)\langle DI_{q_1}(a_{q_1}), DI_{q_2}(b_{q_2})\rangle. \tag{13}$$

Thus our goal is to upper bound $\mathbf{E}[(X+Y)^2]$ by the RHS of (11); we do this via the following two claims.

Claim 29. *We have*

$$\mathbf{E}[X^2] \leq \left(\sum_{q=1}^{d} q!\langle a_q, b_q\rangle\right)^2 + 2^{O(d\log d)}\epsilon^2. \tag{14}$$

Claim 30. *We have*

$$\mathbf{E}[Y^2] \leq 2^{O(d\log d)}\epsilon^2. \tag{15}$$

Given Claims 29 and 30 we have

$$\begin{aligned}
\mathbf{E}[(X+Y)]^2 &= \mathbf{E}[X^2] + 2\,\mathbf{E}[XY] + \mathbf{E}[Y^2] \leq \mathbf{E}[X]^2 + \mathbf{E}[Y^2] + 2\sqrt{\mathbf{E}[X^2]\,\mathbf{E}[Y]^2} &(16)\\
&\leq \left(\sum_{q=1}^{d} q!\langle a_q, b_q\rangle\right)^2 + 2^{O(d\log d)}\epsilon^2 + 2\sqrt{\mathbf{E}[X^2]\,\mathbf{E}[Y]^2}. &(17)
\end{aligned}$$

Now note that

$$\begin{aligned}
\sqrt{\mathbf{E}[X^2]\,\mathbf{E}[Y]^2} &= \sqrt{\left(\left(\sum_{q=1}^{d} q!\langle a_q, b_q\rangle\right)^2 + 2^{O(d\log d)}\epsilon^2\right)\cdot 2^{O(d\log d)}\epsilon^2} \quad \text{(by Claims 29 and 30)}\\
&\leq 2^{O(d\log d)}\epsilon^2 + 2^{O(d\log d)}\epsilon\cdot\sum_{q=1}^{d} q!\langle a_q, b_q\rangle \qquad (\text{by } \sqrt{x+y}\leq\sqrt{x}+\sqrt{y})\\
&\leq 2^{O(d\log d)}\epsilon^2 + 2^{O(d\log d)}\epsilon\cdot\sum_{q=1}^{d}\left(\sqrt{q!}\|a_q\|\right)\cdot\left(\sqrt{q!}\|b_q\|\right) \quad \text{(by Cauchy-Schwarz)}\\
&= 2^{O(d\log d)}\epsilon \quad \text{(by (9))}.
\end{aligned}$$

Combining this with (17) we indeed get (11) as desired. Thus it remains only to prove Claims 29 and 30.

Proof of Claim 29. We may write X^2 as $A + B$, where

$$A = \sum_{q=1}^{d}\frac{1}{q^2}\langle DI_q(a_q), DI_q(b_q)\rangle^2$$

and

$$B = \sum_{d\geq q_1>q_2\geq 1}\frac{2}{q_1 q_2}\langle DI_{q_1}(a_{q_1}), DI_{q_1}(b_{q_1})\rangle\cdot\langle DI_{q_2}(a_{q_2}), DI_{q_2}(b_{q_2})\rangle.$$

First we analyze $\mathbf{E}[A]$. Using Theorem 22 we have that

$$\begin{aligned}
\mathbf{E}[A] &= \sum_{q=1}^{d}\frac{1}{q^2}\,\mathbf{E}[\langle DI_q(a_q), DI_q(b_q)\rangle^2]\\
&= \sum_{q=1}^{d}(q!)^2\langle a_q, b_q\rangle^2 + q^2\sum_{r=1}^{q-1}((r-1)!)^2\binom{q-1}{r-1}^4\cdot(2q-2r)\|a_q\widetilde{\otimes}_r b_q\|^2. &(18)
\end{aligned}$$

Now observe that for $1\leq r\leq q-1$, we have

$$\|a_q\widetilde{\otimes}_r b_q\|^2 \leq \lambda_{\max}(a_q)^2\cdot\|b_q\|^2 \leq \epsilon^2/q!$$

where we have used Fact 27 for the first inequality and the eigenvalue bound and (9) for the second. Hence from (18) we obtain

$$\mathbf{E}[A] \leq \sum_{q=1}^{d} (q!)^2 \langle a_q, b_q \rangle^2 + 2^{O(d \log d)} \epsilon^2. \tag{19}$$

We turn now to bound $\mathbf{E}[B]$. Using Claim 21 we get

$$\begin{aligned}
\mathbf{E}[B] &= \sum_{d \geq q_1 > q_2 \geq 1} \frac{2}{q_1 q_2} \mathbf{E}\left[\left(q_1^2 \sum_{r_1=1}^{q_1} (r_1-1)! \binom{q_1-1}{r_1-1}^2 I_{2q_1-2r_1}(a_{q_1} \widetilde{\otimes}_{r_1} b_{q_1}) \right) \cdot \right. \\
&\qquad \left. \left(q_2^2 \sum_{r_2=1}^{q_2} (r_2-1)! \binom{q_2-1}{r_2-1}^2 I_{2q_2-2r_2}(a_{q_2} \widetilde{\otimes}_{r_2} b_{q_2}) \right) \right] \\
&= \sum_{d \geq q_1 > q_2 \geq 1} 2 q_1 q_2 \sum_{r_1=1}^{q_1} \sum_{r_2=1}^{q_2} (r_1-1)!(r_2-1)! \binom{q_1-1}{r_1-1}^2 \binom{q_2-1}{r_2-1}^2 \cdot \\
&\qquad \mathbf{E}[I_{2q_1-2r_1}(a_{q_1} \widetilde{\otimes}_{r_1} b_{q_1}) I_{2q_2-2r_2}(a_{q_2} \widetilde{\otimes}_{r_2} b_{q_2})].
\end{aligned} \tag{20}$$

Let us fix a given outcome of $q_1 > q_2$. Recalling Claim 11, we see that the only (r_1, r_2) pairs that will give a nonzero expectation are those such that $2q_1 - 2r_1 = 2q_2 - 2r_2$, i.e. $r_2 = q_2 - q_1 + r_1$. For such an (r_1, r_2) pair, by Claim 11 we get that $\mathbf{E}[I_{2q_1-2r_1}(a_{q_1} \widetilde{\otimes}_{r_1} b_{q_1}) \cdot I_{2q_2-2r_2}(a_{q_2} \widetilde{\otimes}_{r_2} b_{q_2})]$ equals $(2q_1 - 2r_1)! \langle a_{q_1} \widetilde{\otimes}_{r_1} b_{q_1}, a_{q_2} \widetilde{\otimes}_{r_2} b_{q_2} \rangle$, which in turn satisfies

$$\begin{aligned}
\langle a_{q_1} \widetilde{\otimes}_{r_1} b_{q_1}, a_{q_2} \widetilde{\otimes}_{r_2} b_{q_2} \rangle &= \langle a_{q_1}, b_{q_1} \rangle \cdot \langle a_{q_2}, b_{q_2} \rangle \text{ if } r_1 = q_1, \\
\langle a_{q_1} \widetilde{\otimes}_{r_1} b_{q_1}, a_{q_2} \widetilde{\otimes}_{r_2} b_{q_2} \rangle &\leq \epsilon^2 \cdot \frac{1}{\sqrt{q_1!}} \cdot \frac{1}{\sqrt{q_2!}} \text{ if } 1 \leq r_1 \leq q_1 - 1,
\end{aligned}$$

where the inequality follows from Fact 28, the eigenvalue bound, and (9). We thus arrive at

$$\begin{aligned}
E[B] &\leq \sum_{d \geq q_1 > q_2 \geq 1} 2(q_1)!(q_2)! \langle a_{q_1}, b_{q_1} \rangle \cdot \langle a_{q_2}, b_{q_2} \rangle \\
&\quad + \sum_{d \geq q_1 > q_2 \geq 1} 2 q_1 q_2 \sum_{r_1=1}^{q_1-1} (r_1-1)!(q_2 - q_1 + r_1 - 1)! \binom{q_1-1}{r_1-1}^2 \binom{q_2-1}{q_2-q_1+r_1-1}^2 \cdot \epsilon^2 \cdot \frac{(2q_1-2r_1)!}{\sqrt{(q_1!)(q_2!)}} \\
&\leq \sum_{d \geq q_1 > q_2 \geq 1} 2(q_1)!(q_2)! \langle a_{q_1}, b_{q_1} \rangle \cdot \langle a_{q_2}, b_{q_2} \rangle + 2^{O(d \log d)} \epsilon^2.
\end{aligned}$$

Combining this with (18), we get that

$$\mathbf{E}[X^2] \leq \left(\sum_{q=1}^{d} q! \langle a_q, b_q \rangle \right)^2 + 2^{O(d \log d)} \epsilon^2.$$

This concludes the proof of Claim 29. □

Proof of Claim 30. We have

$$\begin{aligned}
\mathbf{E}[Y^2] &= \sum_{d \geq q_1 > q_2 \geq 1} \sum_{d \geq q_3 > q_4 \geq 1} \left(\frac{1}{q_1} + \frac{1}{q_2} \right) \left(\frac{1}{q_3} + \frac{1}{q_4} \right) \cdot \mathbf{E}[\langle DI_{q_1}(a_{q_1}), DI_{q_2}(b_{q_2}) \rangle \cdot \langle DI_{q_3}(a_{q_3}), DI_{q_4}(b_{q_4}) \rangle] \\
&< 4 \sum_{1 \leq q_1 < q_2 \leq d} \sum_{1 \leq q_3 < q_4 \leq d} \sqrt{\mathbf{E}[\langle DI_{q_1}(a_{q_1}), DI_{q_2}(b_{q_2}) \rangle^2]} \cdot \sqrt{\mathbf{E}[\langle DI_{q_3}(a_{q_3}), DI_{q_4}(b_{q_4}) \rangle^2]},
\end{aligned}$$

where we have used Cauchy-Schwarz and the fact that $(\frac{1}{q_1}+\frac{1}{q_2})(\frac{1}{q_3}+\frac{1}{q_4})$ is always strictly less than 4. Fix any $d \geq q_1 > q_2 \geq 1$; to prove Claim 30 it suffices to show that $\mathbf{E}[\langle DI_{q_1}(a_{q_1}), DI_{q_2}(b_{q_2})\rangle^2] \leq 2^{O(d\log d)}\epsilon^2$. For this we use the third bound of Theorem 22, which gives

$$\mathbf{E}[\langle DI_{q_1}(a_{q_1}), DI_{q_2}(b_{q_2})\rangle^2] \;=\; (q_1)^2(q_2)^2\sum_{r=1}^{q_2}((r-1)!)^2\binom{q_1-1}{r-1}^2\binom{q_2-1}{r-1}^2(q_1+q_2-2r)\|a_{q_1}\widetilde{\otimes}_r b_{q_2}\|^2.$$

For any $1 \leq r \leq q_2$ we have that $r \leq q_1 - 1$ (since $q_1 > q_2$), and hence by Fact 27, the eigenvalue bound and (9) we get that $\|a_{q_1}\widetilde{\otimes}_r b_{q_2}\|^2 \leq \epsilon^2/q_2!$. Thus each summand in the previous expression is at most $2^{O(q_1\log q_1)}\epsilon^2 = 2^{O(d\log d)}\epsilon^2$, as required. This concludes the proof of Claim 30, and with it the proof of Lemma 26 and of Theorem 19. □

5 Decomposing k-tuples of multilinear degree-d Gaussian polynomials

In this section we prove our main decomposition result for k-tuples of multilinear Gaussian polynomials, Theorem 31. We begin by giving a precise statement of the result, followed by a discussion of how the result fits into our broader context.

Theorem 31. *Fix $d \geq 2$ and fix any non-increasing computable function $\beta : [1,\infty) \to (0,1)$ that satisfies $\beta(x) \leq 1/x$. There is a procedure* **Regularize-Poly**$_\beta$ *with the following properties. The procedure takes as input a degree-d multilinear Gaussian polynomial p with $\mathbf{Var}[p] = 1$ and a parameter $\tau > 0$. It runs in $\mathrm{poly}(n^d)\cdot O_{d,\tau}(1)$ time and outputs a collection of polynomials $\{h_q\}_{q=0,\ldots,d}$ and $\{A_{q,\ell}\}_{q=0,\ldots,d,\ell=1,\ldots,m_q}$.*

Write $p(x)$ as $\sum_{q=0}^d c_q p_q(x)$ where $p_q \in \mathcal{W}^q$ for all q and $\mathbf{Var}[p_q] = 1$ for $1 \leq q \leq d$. For $0 \leq q \leq d$ and $x \in \mathbb{R}^n$, let

$$\tilde{p}_q(x) = c_q h_q(A_{q,1}(x),\ldots,A_{q,m_q}(x)), \qquad \text{and let} \qquad \tilde{p}(x) = \sum_{q=0}^d \tilde{p}_q(x).$$

The following conditions hold:

1. *For each $q \in \{0,\ldots,d\}$ the polynomial $\tilde{p}_q$ belongs to $\mathcal{W}^q$. Moreover, for $q \in \{1,\ldots,d\}$, each polynomial $A_{q,\ell}$ belongs to $\mathcal{W}^j$ for some $1 \leq j \leq d$ and has $\mathbf{Var}[A_{q,\ell}] = 1$.*
2. *We have $\left|\mathbf{Pr}_{x\sim N(0,1)^n}[p(x)\geq 0] - \mathbf{Pr}_{x\sim N(0,1)^n}[\tilde{p}(x)\geq 0]\right| \leq O(\tau)$, and moreover $\mathbf{Var}_{x\sim N(0,1)^n}[p(x)-\tilde{p}(x)] \leq (\tau/d)^{3d}$.*
3. *Each polynomial h_q is a multilinear polynomial in its m_q arguments. Moreover, there exist functions $N_\beta(d,\tau)$ and $M_\beta(d,\tau)$ such that if $\mathrm{Coeff}(h_q)$ denotes the sum of the absolute values of the coefficients of h_q, then $\sum_{q=1}^d \mathrm{Coeff}(h_q) \leq M_\beta(d,\tau)$ and the number of arguments to all h_q's, $\sum_{q=1}^d m_q$, is at most $N_\beta(d,\tau)$. Also, the degree of the polynomial h_q (for all $1 \leq q \leq d$) is upper bounded by d.*
4. *Further, let $\mathrm{Num} = \sum_{q=1}^d m_q$ and $\mathrm{Coeff} = \sum_{q=1}^d \mathrm{Coeff}(h_q)$. Then each polynomial $A_{q,\ell}(x)$ is $\beta(\mathrm{Num}+\mathrm{Coeff})$-eigenregular.*

Discussion. Intuitively, Condition (2) means that it will be sufficient to do deterministic approximate counting for the polynomial $\tilde{p}$ rather than the original polynomial p. Condition (4) ensures that the eigenregularity of each polynomial $A_{q,\ell}$ compares favorably both with the number of polynomials produced and with the size of the coefficients in the "outer" polynomials h_q. As discussed earlier, having the eigenregularity be small relative to the number of polynomials is crucial since it is required in order for our CLT, Theorem 19,

to yield a good bound. We need to deal with the size of the coefficients for technical reasons – as we will see in Section 6, we will apply our CLT where its "test function" α is a smooth approximator to the 0/1-valued function which, on input $\{A_{j,\ell}\}_{j,\ell}$, outputs 1 if and only if $\text{sign}\left(\sum_{q=0}^{d} h_q(A_{q,1},\ldots,A_{q,m_q})\right) = 1$. Our CLT's quantitative bound depends on the second derivative of α, and to bound this we need coefficient size bounds on the h_q polynomials.

We build up to Theorem 31 in a sequence of incremental stages. In Section 5.1 we begin by describing how to decompose a single element of a given Wiener chaos. Because of our requirement that the number of polynomials produced in the decomposition must be very small relative to the eigenregularity that is achieved — see Remark 20 — even this is a non-trivial task, requiring two "layers" of decomposition and an approach that goes well beyond the decomposition techniques in previous work [DDS13a, DDS13b]. In Section 5.2 we extend this and describe how to simultaneously decompose a k-tuple of elements of the same Wiener chaos. (See the beginning of Section 5.2 for an explanation of why we need to be able to simultaneously decompose many polynomials at once.) In Section 5.3 we describe how to handle a $k(d+1)$-tuple of elements where there are k elements from each of the $d+1$ Wiener chaoses $\mathcal{W}^0,\ldots,\mathcal{W}^d$. Finally, in Section 5.4 we specialize this decomposition for a $k(d+1)$-tuple to obtain Theorem 31.

5.1 Decomposing a single multilinear element of the q-th Wiener chaos

We begin with a useful definition and fact:

Definition 32. *For $S \subseteq [n]$, we say that a tensor*

$$f = \sum_{i_1,\ldots,i_q=1}^{n} f(i_1,\ldots,i_q)\cdot e_{i_1}\otimes\cdots\otimes e_{i_q}$$

is supported on S *if $f(i_1,\ldots,i_q)=0$ whenever any of $i_1,\ldots,i_q$ do not belong to S.*

Fact 33. *Let $f\in\mathcal{H}^{\odot p}$ be supported on $S\subseteq[n]$ and $g\in\mathcal{H}^{\odot q}$ be supported on $T\subseteq[n]$ where S,T are disjoint. Then for any $0<r\le\min\{p,q\}$ we have that the contraction product $f\otimes_r g$ equals 0.*

Proof. We may write

$$f=\sum_{j_1,\ldots,j_p\in S} f(j_1,\ldots,j_p)e_{j_1}\otimes\ldots\otimes e_{j_p} \quad\text{and}\quad g=\sum_{j_1,\ldots,j_q\in T} g(j_1,\ldots,j_q)e_{j_1}\otimes\ldots\otimes e_{j_q}.$$

Note that

$$f\otimes_r g=\sum_{i_1,\ldots,i_r\in[n]}\langle f,e_{i_1}\otimes\ldots\otimes e_{i_r}\rangle\otimes\langle g,e_{i_1}\otimes\ldots\otimes e_{i_r}\rangle$$

The value $\langle f,e_{i_1}\otimes\ldots\otimes e_{i_r}\rangle$ is 0 unless all the i_j's lie in S, and likewise $\langle g,e_{i_1}\otimes\ldots\otimes e_{i_r}\rangle=0$ unless all the i_j's lie in T. Since $S\cap T=\emptyset$, the fact is proved. □

As our lowest-level decomposition tool, we show that given a symmetric tensor with a large-magnitude eigenvalue, it is possible to efficiently find two lower-dimensional symmetric tensors g_1 and g_2, supported on disjoint subsets of $[n]$, such that f is correlated with the product of g_1 and g_2:

Lemma 34. *Fix any $q\ge 2$. There is a $\text{poly}(n^q)$-time deterministic procedure with the following properties: Let $f\in\mathcal{H}^{\odot q}$ be a multilinear symmetric tensor that has $\mathbf{Var}[I_q(f)]=1$ and $\lambda_{\max}(f)\ge\eta>0$. On input f, the procedure outputs multilinear symmetric tensors $g_1\in\mathcal{H}^{\odot q_1}$ and $g_2\in\mathcal{H}^{\odot q_2}$ such that the following conditions hold:*

1. $\mathbf{Var}[I_{q_1}(g_1)] = \mathbf{Var}[I_{q_2}(g_2)] = 1$, *and* $q_1, q_2 > 0$ *with* $q_1 + q_2 = q$;

2. g_1 *and* g_2 *are supported on disjoint sets* $S, T \subset [n]$;

3. $\mathbf{E}[I_q(f) \cdot I_{q_1}(g_1) \cdot I_{q_2}(g_2)] \geq \eta/2^q$.

Proof. The procedure begins as follows:

1. For each partition of $[q]$ into S and $\overline{S} = [q] \setminus S$ where $|S|, |\overline{S}| > 0$, it computes the value $\lambda_{S,\overline{S}}(f) = \sup_{x \in \mathcal{H}^S, y \in \mathcal{H}^{\overline{S}}} \frac{\langle g, x \otimes y \rangle}{\|x\| \|y\|}$ and the associated tensors $\alpha = x/\|x\|$, $\beta = y/\|y\|$.

2. For the (α, β) pair from the iteration of step (1) which achieves the largest value for $\lambda_{S,\overline{S}}(f)$, let $\alpha' = \tilde{\alpha}/\|\tilde{\alpha}\|$ and $q_1 = |S_1|$, and let $\beta' = \tilde{\beta}/\|\tilde{\beta}\|$ and $q_2 = |S_2|$.

Recalling the variational characterization of singular values, each iteration of Step (1) (for a given choice of $S, \overline{S}$) is an SVD computation and can be performed in time $\mathrm{poly}(n^q)$. Since there are at most 2^q partitions $S, \overline{S}$ to consider, the overall running time of these first two steps is $\mathrm{poly}(n^q)$.

We proceed to analyze these first two steps of the procedure. The fact that $\lambda_{\max}(f) \geq \eta$ means that the maximizing α, β identified in Step (2) will satisfy $\|\alpha\|_2 = \|\beta\|_2 = 1$ and $\langle f, \alpha \otimes \beta \rangle \geq \eta$. Since f is a multilinear tensor these tensors α, β will also be multilinear (if, say, α had a nonzero diagonal entry then a larger value of $\langle f, \alpha \otimes \beta \rangle$ could be obtained by zeroing out that entry and rescaling). Since f is a symmetric tensor, it follows that

$$\langle f, \alpha \otimes \beta \rangle = \langle f, \tilde{\alpha} \otimes \beta \rangle = \langle f, \tilde{\alpha} \otimes \tilde{\beta} \rangle \geq \eta.$$

Since symmetrizing cannot increase the norm of a tensor, we have that $\|\tilde{\alpha}\| \leq \|\alpha\| = 1$ and likewise for β. It follows that $\langle f, \alpha' \otimes \beta' \rangle \geq \eta$, $\|\alpha'\| = \|\beta'\| = 1$, $\alpha' \in \mathcal{H}^{q_1}$, and $\beta' \in \mathcal{H}^{q_2}$.

From this point it is not difficult to achieve conditions (1) and (3) of the lemma; to achieve condition (2) we proceed as follows. Note that

$$\langle f, \alpha' \otimes \beta' \rangle = \sum_{S_1 \in [n]^{q_1}, S_2 \in [n]^{q_2}} f(S_1, S_2) \alpha'(S_1) \beta'(S_2).$$

Now consider a randomized process that divides $[n]$ into two sets $\mathcal{A}_1$ and $\mathcal{A}_2$ by independently assigning each $i \in [n]$ to $\mathcal{A}_1$ with probability $1/2$ and to $\mathcal{A}_2$ with probability $1/2$ (we will later derandomize this process below). Given an outcome of $\mathcal{A}_1$ and $\mathcal{A}_2$, we consider $\nu_1 \in \mathcal{H}^{\odot q_1}$ and $\nu_2 \in \mathcal{H}^{\odot q_2}$ defined as follows: for each $S_1 \in [n]^{q_1}, S_2 \in [n]^{q_2}$,

$$\nu_1(S_1) = \alpha'(S_1) \cdot \mathbf{1}[S_1 \subseteq \mathcal{A}_1] \quad \text{and} \quad \nu_2(S_2) = \beta'(S_2) \cdot \mathbf{1}[S_2 \subseteq \mathcal{A}_2],$$

where "$S_i \subseteq \mathcal{A}_i$" means that each coordinate of S_i lies in $\mathcal{A}_i$. We have that

$$\mathbf{E}[\langle f, \nu_1 \otimes \nu_2 \rangle] = \sum_{S_1 \in [n]^{q_1}, S_2 \in [n]^{q_2}} f(S_1, S_2) \alpha'(S_1) \beta'(S_2) \cdot (1/2)^{|S_1|} \cdot (1/2)^{|S_2|}, \tag{21}$$

where $|S_i|$ denotes the number of distinct values that are present in the coordinates of S_i. Since α' and β' are multilinear, the only nonzero contributions to the sum (21) are from (S_1, S_2) pairs with $|S_1| = q_1$ and $|S_2| = q_2 = q - q_1$. Hence we have

$$\mathbf{E}[\langle f, \nu_1 \otimes \nu_2 \rangle] = \frac{1}{2^q} \sum_{S_1 \in [n]^{q_1}, S_2 \in [n]^{q_2}} f(S_1, S_2) \alpha'(S_1) \beta'(S_2) \geq \frac{\eta}{2^q}.$$

The above analysis requires only q-wise independence, so constructing $\mathcal{A}_1$ and $\mathcal{A}_2$ (and the resulting ν_1, ν_2) using a q-wise independent distribution $\mathcal{D}$, we get that

$$\mathbf{E}_{\mathcal{D}}[\langle f, \nu_1 \otimes \nu_2 \rangle] \geq \frac{\eta}{2^q},$$

and thus some outcome in the support of $\mathcal{D}$ must achieve at least the expected value. The third step of the algorithm is to deterministically enumerate over all points in the support of the q-wise independent distribution (using well-known constructions of q-wise independent distributions [ABI85] this can be done in time $\mathrm{poly}(n^q)$) and to output a pair $g_1 = \frac{\nu_1}{\sqrt{q_1!}\|\nu_1\|}$, $g_2 = \frac{\nu_2}{\sqrt{q_2!}\|\nu_2\|}$ that has $\langle f, \nu_1 \otimes \nu_2 \rangle \geq \frac{\eta}{2^q}$.

We now verify that g_1 and g_2 satisfy all the required conditions. First, it is clear that g_1 and g_2 satisfy $g_1 \in \mathcal{H}^{\odot q_1}$, $g_2 \in \mathcal{H}^{\odot q_2}$, and it is obvious from the construction that g_1 and g_2 are supported on disjoint sets $\mathcal{A}_1$ and $\mathcal{A}_2$, so condition (2) indeed holds. Turning to condition (1), since $q_1 > 0$ we have that $\mathbf{Var}[I_{q_1}(g_1)] = \mathbf{E}[I_{q_1}(g_1)^2]$, which equals 1 by Claim 11 (and similarly we get that $\mathbf{Var}[I_{q_2}(g_2)] = 1$). For Condition (3), we first show that $I_{q_1}(g_1) \cdot I_{q_2}(g_2) = I_q(\mathrm{Sym}(g_1 \otimes g_2))$ (and hence $I_{q_1}(g_1) \cdot I_{q_2}(g_2)$ lies in the Wiener chaos of degree q). To see this, recall from the multiplication formula for Ito integrals (Theorem 13) that we have

$$I_{q_1}(g_1) \cdot I_{q_2}(g_2) = \sum_{r=0}^{\min\{q_1,q_2\}} r! \cdot \binom{q_1}{r}\binom{q_2}{r} I_{q_1+q_2-2r}\mathrm{Sym}(g_1 \otimes_r g_2)).$$

Since g_1 and g_2 are supported on disjoint sets, by Fact 33 all summands with $r \neq 0$ vanish, and we get $I_{q_1}(g_1) \cdot I_{q_2}(g_2) = I_{q_1+q_2}(\mathrm{Sym}(g_1 \otimes g_2))$ as claimed.

With this identity in hand, we have that

$$\begin{aligned}
\mathbf{E}[I_q(f)I_{q_1}(g_1)I_{q_2}(g_2)] &= \mathbf{E}[I_q(f)I_q(\mathrm{Sym}(g_1 \otimes g_2))] \\
&= q!\langle f, \mathrm{Sym}(g_1 \otimes g_2)\rangle \qquad \text{(by Claim 11)} \\
&= q!\langle f, g_1 \otimes g_2\rangle \qquad \text{(since } f \text{ is symmetric)} \\
&\geq \langle f, \nu_1 \otimes \nu_2\rangle \qquad \text{(since } \|\nu_1\| \leq \|\alpha'\| = 1 \text{ and } \|\nu_2\| \leq \|\beta'\| = 1) \\
&\geq \frac{\eta}{2^q},
\end{aligned}$$

and Lemma 34 is proved. □

We are now ready to define our first algorithmic primitive, the procedure **Split-One-Wiener**. This procedure either certifies that its input polynomial (an element $I_q(f)$ of the q-th Wiener chaos) is eigenregular, or else it "splits off" a product $P \cdot Q$ from its input polynomial. (Here and subsequently the suffix "**-One-Wiener**" indicates that the procedure applies only to one element $I_q(f)$ belonging to one level of the Wiener chaos.)

Lemma 35. *Fix any $q \geq 2$. There is a deterministic procedure* **Split-One-Wiener** *which takes as input a polynomial $I_q(f) \in \mathcal{H}_q$ that has $\mathbf{Var}[I_q(f)] = 1$ and a parameter $\eta > 0$.* **Split-One-Wiener** *runs in deterministic $\mathrm{poly}(n^q, 1/\eta)$ time and has the following guarantee:*

- *If $\lambda_{\max}(f) < \eta$, then* **Split-One-Wiener** *returns "eigenregular".*
- *Otherwise, if $\lambda_{\max}(f) \geq \eta$, then* **Split-One-Wiener** *outputs a quadruple (P, Q, R, c) with the following properties:*

  1. *$P = I_{q_1}(g_1) \in \mathcal{W}^{q_1}$ and $Q = I_{q_2}(g_2) \in \mathcal{W}^{q_2}$ where $q_1+q_2 = q$, $q_1, q_2 > 0$, and $\mathbf{Var}[I_{q_1}(g_1)] = \mathbf{Var}[I_{q_2}(g_2)] = 1$.*

2. *The tensors $g_1 \in \mathcal{H}^{\odot q_1}$ and $g_2 \in \mathcal{H}^{\odot q_2}$ are supported on disjoint sets $S, T \subset [n]$.*
3. *$P \cdot Q \in \mathcal{W}^q$ and $\mathbf{Var}[P \cdot Q] = 1$, and all of P, Q, R are multilinear.*
4. *The value $c \stackrel{def}{=} \mathbf{E}[I_q(f) \cdot P \cdot Q]$ satisfies $c \geq \eta/2^q$.*
5. *$R \in \mathcal{W}^q$ and $I_q(f) = cP \cdot Q + R$ and $\mathbf{E}[P \cdot Q \cdot R] = 0$.*
6. *$\mathbf{Var}(R) = 1 - c^2$.*

Proof. **Split-One-Wiener** runs the procedure from Lemma 34 and checks whether the largest value $\lambda_{S,\overline{S}}(f)$ achieved in Step (1) is at least η. If it is less than η then it outputs "eigenregular." Otherwise it sets $P = I_{q_1}(g_1)$, $Q = I_{q_2}(g_2)$, sets $c = \mathbf{E}[I_q(f) \cdot P \cdot Q] = q!\langle f, g_1 \otimes g_2 \rangle$, and sets $R = I_q(f) = c \cdot P \cdot Q$.

Lemma 34 directly gives properties (1),(2) and (4), and property (3) follows from the fact that $\mathbf{E}[Q] = \mathbf{E}[P] = 0$ and P and Q are independent random variables (observe that by property (2) they are polynomials over disjoint sets of variables). The first two parts of (5) are immediate; for the last part, recalling that $R = I_q(f) - cP \cdot Q$, we have that R is simply the component of $I_q(f)$ that is orthogonal to $P \cdot Q$. Since R lies in $\mathcal{W}^q$ its mean is zero, so by linear algebra we have that $\mathbf{Var}[R] = \mathbf{E}[R^2] = 1 - c^2$ as claimed. □

Building on the algorithmic primitive **Split-One-Wiener**, we now describe a procedure **Decompose-One-Wiener** which works by iteratively executing **Split-One-Wiener** on the "remainder" portion R that was "left over" from the previous call to **Split-One-Wiener**. Intuitively, the overall effect of this procedure is to break its input polynomial into a sum of products of pairs of polynomials, plus a remainder term which is either eigenregular or else has variance which is negligibly small.

(We draw the reader's attention to the quantitative bound on coefficients given by property (6) of Lemma 36. This coefficient bound will play a crucial role in the mollification procedure of Section 6.)

Lemma 36. *Fix any $q \geq 2$. There is a deterministic procedure* **Decompose-One-Wiener** *which takes as input a polynomial $I_q(f) \in \mathcal{W}_q$ that has $\mathbf{Var}[I_q(f)] = 1$ and parameters η and ϵ.* **Decompose-One-Wiener** *runs in poly$(n^q, 1/\eta, \log(1/\epsilon))$ time and has the following guarantee:*

1. *It outputs a set L of triples $\{(c_i, P_i, Q_i)\}_{i=1}^m$ and a polynomial R such that $I_q(f) = \sum_{i=1}^m c_i P_i Q_i + R$.*
2. *For each $i = 1, \ldots, m$ we have $P_i \in \mathcal{W}^{q_{i,1}}$ and $Q_i \in \mathcal{W}^{q_{i,2}}$ with $q_{i,1}, q_{i,2} > 0$ and $q_{i,1} + q_{i,2} = q$; moreover $\mathbf{Var}[P_i] = \mathbf{Var}[Q_i] = \mathbf{Var}[P_i \cdot Q_i] = 1$ for all $i \in [m]$, $R \in \mathcal{W}^q$, and all P_i, Q_i and R are multilinear.*
3. *$m \leq O((4^q/\eta^2)\log(1/\epsilon))$.*
4. *Either R is η-eigenregular, in which case* **Decompose-One-Wiener** *returns "eigenregular remainder", or else $\mathbf{Var}[R] \leq \epsilon$, in which case* **Decompose-One-Wiener** *returns "small remainder".*
5. *$\mathbf{E}[(\sum_{j=1}^m c_j P_j \cdot Q_j) \cdot R] = 0$. As a consequence, we have $\mathbf{Var}[\sum_{j=1}^m c_j P_j \cdot Q_j] + \mathbf{Var}[R] = 1$.*
6. *$\sum_{j=1}^m c_j^2 \leq (2^q/\eta)^{4(m-1)}$.*

Proof. The procedure **Decompose-One-Wiener** is defined below. It is helpful to keep the following invariant in mind: At any stage during the execution of the algorithm, we let V_L denote the linear subspace spanned by $\{P_i \cdot Q_i\}$ where $L = \{(c_i, P_i, Q_i)\}$. The algorithm maintains the invariant that $I_q(g)$ is orthogonal to V_L (as is clear by construction).

i. Initialize L to the empty set of triples and the index m to 0.

ii. Initialize $g = f$ and hence $I_q(g) = I_q(f)$.

iii. If $\mathbf{Var}[I_q(g)] \leq \epsilon$, then output the set $L = \{(c_i, P_i, Q_i)\}_{i=1}^m$ and the polynomial $R = I_q(f) - \sum_{i=1}^m c_i P_i \cdot Q_i$, and return "small remainder."

iv. Else, choose a constant ζ so that $\mathbf{Var}[I_q(\zeta g)] = 1$.

v. Run procedure **Split-One-Wiener** (using parameter η) on $I_q(\zeta g)$. If it returns "eigenregular", then stop the procedure and output the set $L = \{c_i, P_i, Q_i\}_{i=1}^m$ and the polynomial $R = I_q(f) - \sum_{i=1}^m c_i P_i \cdot Q_i$, and return "eigenregular remainder".

vi. Else if the output of **Split-One-Wiener** is (P, Q, R, c'), then append $(c_{m+1}, P_{m+1}, Q_{m+1})$ to the list L where $c_{m+1} = c'$, $P_{m+1} = P$ and $Q_{m+1} = Q$. Now, project the polynomial $I_q(f)$ to V_L and let $I_q(g)$ denote the part of $I_q(f)$ that is orthogonal to V_L, i.e. $I_q(g) = (I_q(f))^{\perp V_L}$. Recompute the constants $c_1, \ldots, c_{m+1}$ so that with the recomputed constants we have $I_q(f) = \sum_{i=1}^{m+1} c_i P_i Q_i + I_q(g)$. Increment m and go to Step [iii].

We now establish the claimed properties. The first and fourth properties are obvious. The second property follows directly from Lemma 36. For the fifth property, note that $I_q(g)$ is orthogonal to $I_q(f) - I_q(g)$ by construction. It remains to justify the third and the sixth properties. We do this using the following claim:

Claim 37. *At each stage in the execution of* **Decompose-One-Wiener**, *when $L = \{(c_i, P_i, Q_i)\}_{i=1,\ldots,k}$, the set $\{P_i \cdot Q_i\}_{i=1}^k$ is $\eta/2^q$-far from being linearly dependent.*

Proof. The proof is by induction. The claim is trivially true for $k = 1$. For the inductive step, observe that by construction, just before $(c_{k+1}, P_{k+1}, Q_{k+1})$ is appended to the list L, we have (by property (2)) that $P_{k+1} \cdot Q_{k+1}$ is a unit vector, and that $|\langle I_q(\zeta g), P_{k+1} \cdot Q_{k+1}\rangle| \geq \eta/2^q$. Since $I_q(\zeta g)$ is orthogonal to V_L (before appending $(c_{k+1}, P_{k+1}, Q_{k+1})$), by Fact 9 we get the stated claim. □

When the **Decompose-One-Wiener** procedure terminates, note that by property (5) we have that $\mathbf{Var}[\sum_{i=1}^m c_i P_i \cdot Q_i] \leq 1$. Hence applying Claim 37 with Claim 8, we get property (6).

It remains only to establish property (3). This follows immediately from the following claim:

Claim 38. *At each stage in the execution of* **Decompose-One-Wiener**, *when $L = \{(c_i, P_i, Q_i)\}_{i=1,\ldots,k}$ and $I_q(g) = (I_q(f))^{\perp V_L}$, we have $\mathbf{Var}[I_q(g)] \leq (1 - \eta^2/4^q)^k$.*

Proof. As before the proof is by induction and the base claim (when $k = 0$) is immediate. For the inductive step, just before appending $(c_{k+1}, P_{k+1}, Q_{k+1})$ to the list L in Step (vi), note that if we define $I_q(h) = I_q(g) - c_{k+1} P_{k+1} Q_{k+1}$, then by the **Split-One-Wiener** guarantee (property (6) of Lemma 35) we have that $\mathbf{Var}[I_q(h)] \leq (1 - \eta^2/4^q) \cdot \mathbf{Var}[I_q(g)]$, which by the inductive hypothesis is at most $(1 - \eta^2/4^q)^{k+1}$. Since the vector $I_q(f) - I_q(h)$ lies in the linear span of $V_L \cup \{P_{k+1} \cdot Q_{k+1}\}$, and $\|I_q(h)\| \leq (1 - \eta^2/4^q)^{k+1}$, hence after appending $(c_{k+1}, P_{k+1}, Q_{k+1})$ to L, we have that the new polynomial $I_q(g)$ defined in step (vi) has $\|I_q(g)\| \leq (1 - \eta/2^q)^{k+1}$. This concludes the proof. □

This concludes the proof of Lemma 36. □

We note that the guarantees of the **Decompose-One-Wiener** procedure bear some resemblance to the decomposition that is used in [DDS13a] for degree-2 Gaussian polynomials. However, in our current context of working with degree-d polynomials, **Decompose-One-Wiener** is not good enough, for the following reason: Suppose that **Decompose-One-Wiener** returns "eigenregular remainder" and outputs a decomposition of $I_q(f)$ as $\sum_{i=1}^m c_i P_i Q_i + R$. While the polynomial R is η-eigenregular, it is entirely possible that the number of polynomials P_i, Q_i in the decomposition (i.e. $2m$) may be as large as $\Omega(\frac{1}{\eta^2} \log(1/\epsilon))$. We would like to apply our CLT to conclude that the joint distribution of R and the polynomials obtained from

the subsequent decomposition of $P_1, Q_1, \ldots, P_m, Q_m$ is close to a normal distribution, but since the number $2m$ of polynomials is already too large when compared to the inverse of the eigenregularity parameter, we cannot use our CLT (recall Remark 20). [3]

We surmount this difficulty by using **Decompose-One-Wiener** as a tool within an improved "two-level" decomposition procedure which we present and analyze below. This improved decomposition procedure has a stronger guarantee than **Decompose-One-Wiener** in the following sense: it breaks its input polynomial into a sum of products of pairs of polynomials plus *two* remainder terms R_{reg} (for "eigenregular") and R_{neg} (for "negligible"). The R_{neg} remainder term is guaranteed to have negligibly small variance, and the R_{reg} remainder term is guaranteed to either be zero or else to be *extremely* eigenregular – in particular, for an appropriate setting of the input parameters, its eigenregularity is much "stronger" than the number of pairs of polynomials that are produced in the decomposition. We term this improved decomposition procedure **Regularize-One-Wiener** because of this extremely strong eigenregularity guarantee.

Before giving the formal statement, we note that intuitively this procedure will be useful because it "guarantees that we make progress" for the following reason: We can always erase the small-variance R_{neg} term at the cost of a small and affordable error, and the degree-q R_{reg} remainder term is so eigenregular that it will not pose an obstacle to our ultimate goal of applying the CLT. Thus we have reduced the original polynomial to a sum of pairwise products of lower-degree polynomials, which can each be tackled inductively using similar methods (more precisely, using the generalization of procedure **Regularize-One-Wiener** to simultaneously decompose multiple polynomials which we describe in the next subsection).

Theorem 39. *Fix any $q \geq 2$. There is a procedure* **Regularize-One-Wiener** *which takes as input a polynomial $I_q(f)$ such that $\mathbf{Var}[I_q(f)] = 1$ and input parameters $\eta_0 = 1 \geq \eta_1 \geq \ldots \geq \eta_K$ and ϵ, where $K = O(1/\epsilon \cdot \log(1/\epsilon))$.* **Regularize-One-Wiener** *runs in poly$(n^q, 1/\eta_K, 1/\epsilon)$ time and has the following guarantee:*

1. *Define $M(i) = \frac{O(4^q)}{\eta_i^2}\log(1/\epsilon)$.* **Regularize-One-Wiener** *outputs a value $1 \leq \ell \leq k$, a set $L = \{(a_{i,j}, P_{i,j}, Q_{i,j})\}_{i=1,\ldots,\ell, j=1,\ldots,M(i)}$ of triples, and a pair of polynomials $R_{\mathrm{reg}}, R_{\mathrm{neg}}$ such that $I_q(f) = \sum_{i=1}^{\ell}\sum_{j=1}^{M(i)} a_{i,j}P_{i,j}\cdot Q_{i,j} + R_{\mathrm{reg}} + R_{\mathrm{neg}}$.*
2. *For each i, j we have $P_{i,j} \in \mathcal{W}^{q_{i,j,1}}$ and $Q_{i,j} \in \mathcal{W}^{q_{i,j,2}}$ with $q_{i,j,1}, q_{i,j,2} > 0$ and $q_{i,j,1} + q_{i,j,2} = q$ and $\mathbf{Var}[P_{i,j}] = \mathbf{Var}[Q_{i,j}] = \mathbf{Var}[P_{i,j}\cdot Q_{i,j}] = 1$; moreover, $P_{i,j}$ and $Q_{i,j}$ are over disjoint sets of variables. In addition, $R_{\mathrm{reg}}, R_{\mathrm{neg}} \in \mathcal{W}^q$ and all of $P_{i,j}, Q_{i,j}, R_{\mathrm{reg}}, R_{\mathrm{neg}}$ are multilinear.*
3. *The polynomial R_{neg} satisfies $\mathbf{Var}[R_{\mathrm{neg}}] \leq \epsilon$ and the polynomial R_{reg} is $\eta_{\ell+1}$-eigenregular, where we define $\eta_{K+1} = 0$.*
4. *For $1 \leq i \leq \ell$ we have $\sum_{j=1}^{M(i)}(a_{i,j})^2 \leq (2^q/\eta_i)^{4(M(i)-1)}$.*

We stress that it is crucially important that condition 3 provides $\eta_{\ell+1}$-eigenregularity rather than η_ℓ-eigenregularity.

Proof. The procedure **Regularize-One-Wiener** is given below. We note that it maintains the invariant $I_q(f) = \sum_{(a,P,Q)\in L} a\cdot P\cdot Q + I_q(g_i)$ throughout its execution (this is easily verified by inspection).

i. Initialize L to the empty set of triples.

[3] Note that the reason this problem did not arise in the degree-2 polynomial decompositions of [DDS13a] is because each polynomial P_i, Q_i obtained from **Decompose-One-Wiener** in that setting must have degree 1 (the only way to break the number 2 into a sum of non-negative integers is as 1+1). Degree-1 polynomials may be viewed as having "perfect eigenregularity" (note that any degree-1 polynomial in Gaussian variables is itself distributed precisely as a Gaussian) and so having any number of such degree-1 polynomials did not pose a problem in [DDS13a].

ii. Initialize $g_1 = f$, so $I_q(g_1) = I_q(f)$.

iii. For $i = 1$ to K do the following:

iii(a). If $\mathbf{Var}[I_q(g_i)] \leq \epsilon$ then set $R_{\text{neg}} = I_q(g)$, set $R_{\text{reg}} = 0$, output L, R_{reg} and R_{neg}, and exit.

iii(b). Otherwise, run **Decompose-One-Wiener** with parameters η_i and ϵ on the polynomial $I_q(\lambda_i g)$, where λ_i is chosen so that $\mathbf{Var}[I_q(\lambda_i g)] = 1$. Let $L_i = \{(c_{i,j}, P_{i,j}, Q_{i,j})\}$ be the set of (at most $M(i)$ many, by Lemma 36) triples and R_i be the polynomial that it outputs.

iii(c). If the call to **Decompose-One-Wiener** in step iii(b) returned "small remainder" then set L to $L \cup L'_i$ where $L'_i = \{(\frac{c_{i,j}}{\lambda_i}, P_{i,j}, Q_{i,j})\}_{(c_{i,j},P_{i,j},Q_{i,j})\in L_i}$, set R_{neg} to R_i/λ_i, set R_{reg} to 0, output L, R_{reg} and R_{neg}, and exit.

iii(d). Otherwise it must be the case that **Decompose-One-Wiener** returned "eigenregular remainder." In this case, if $\mathbf{Var}[\sum_{j=1}^{M(i)} c_{i,j}P_{i,j} \cdot Q_{i,j}] \leq \epsilon$, then set R_{neg} to $\sum_{(c_{i,j},P_{i,j},Q_{i,j})\in L_i} \frac{c_{i,j}}{\lambda_i} \cdot P_{i,j} \cdot Q_{i,j}$ and R_{reg} to R_i/λ_i, output L, R_{reg} and R_{neg}, and exit.

iii(e). Otherwise, set g_{i+1} to satisfy $I_q(g_{i+1}) = R_i/\lambda_i$, set L to $L \cup L'_i$ where $L'_i = \{(\frac{c_{i,j}}{\lambda_i}, P_{i,j}, Q_{i,j})\}_{(c_{i,j},P_{i,j},Q_{i,j})\in L_i}$, increment i, and go to the next iteration of step (iii).

For Property (1), we observe that the claimed bound on $M(i)$ follows immediately from part (3) of Lemma 36). The rest of Property (1) follows from the invariant and inspection of steps iii(c) and iii(d). Property (2) follows directly from part (2) of Lemma 36.

To establishing the remaining properties we will use the following claim:

Claim 40. *For each i we have $\mathbf{Var}[I_q(g_i)] \leq (1-\epsilon)^{i-1}$.*

Proof. The proof is by induction on i. The claim clearly holds for $i = 1$. For the inductive step, observe that the only way the procedure reaches step iii(e) and increments i is if the execution of **Decompose-One-Wiener** on $I_q(\lambda_i g)$ returned "eigenregular remainder" and the decomposition $I_q(\lambda_i g) = \sum c_{i,j}P_{i,j}Q_{i,j} + R_i$ has $\mathbf{Var}[\sum c_{i,j}P_{i,j}Q_{i,j}] > \epsilon$, and hence (by part (5) of Lemma 36) $\mathbf{Var}[R_i] \leq (1-\epsilon)\,\mathbf{Var}[I_q(\lambda_i g_i)]$. Consequently in this case we have $\mathbf{Var}[I_q(g_{i+1})] = \mathbf{Var}[R_i/\lambda_i] \leq (1-\epsilon)\,\mathbf{Var}[I_q(g_i)]$, which inductively is at most $(1-\epsilon)^i$ as desired. □

Note that this claim immediately gives that $\lambda_i \geq 1$ for all i, which together with part (6) of Lemma (36) gives Property (4).

It remains only to establish Property (3). Note that by Claim 40 it must be the case that the algorithm halts and exits at some iteration of either step iii(a), iii(c), or iii(d) — if it has not already exited by the time i reaches K, since $\mathbf{Var}[I_q(g_i)] \leq (1-\epsilon)^{i-1}$ once it reaches $i = K$ it will exit in step iii(a). We consider the three possibilities in turn. If it exits at Step iii(a) then clearly Property (3) is satisfied. If it exits at Step iii(c) then by Lemma (36) we have that $\mathbf{Var}[R_i] \leq \epsilon$; since $\lambda_i \geq 1$ this means that $\mathbf{Var}[R_{\text{neg}}] = \mathbf{Var}[R_i/\lambda_i] \leq \epsilon$ and again Property (3) holds. Finally, if it exits at Step iii(d) during the i-th iteration of the loop then observe that the value of ℓ is $i-1$ (since L_i is *not* added on to L). Lemma (36) guarantees that R_i (and hence R_{reg}) is η_i-eigenregular, i.e. $\eta_{\ell+1}$-eigenregular, and as above the fact that $\lambda_i \geq 1$ ensures that $\mathbf{Var}[R_{\text{neg}}] \leq \epsilon$, and Property (3) holds in this case as well. This concludes the proof of Theorem 39. □

5.2 Decomposing a k-tuple of multilinear elements of the q-th Wiener chaos

In this section we generalize the **Regularize-One-Wiener** procedure to simultaneously decompose multiple polynomials that all belong to $\mathcal{W}^q$. Even though our ultimate goal is to decompose a single degree-d Gaussian polynomial, we require a procedure that is capable of handling many polynomials because even

decomposing a single degree-d polynomial using **Regularize-One-Wiener** will give rise to many lower-degree polynomials which all need to be decomposed in turn. (This is also the reason why we must prove Theorem 43, which deals with k Gaussian polynomials, in order to ultimately obtain Theorem 31, which decomposes a single Gaussian polynomial.)

A natural approach to decompose r polynomials $I_q(f_1), \ldots, I_q(f_r) \in \mathcal{W}^q$ is simply to run **Regularize-One-Wiener** r separate times. However, this simpleminded approach could well result in different values $\ell_1, \ldots, \ell_r$ being obtained from the r calls, and hence in different levels of eigenregularity for the r "remainder" polynomials $R_{1,\mathrm{reg}}, \ldots, R_{r,\mathrm{reg}}$ that are constructed. This is a problem because some of the calls may yield a relatively large eigenregularity parameter, while other calls may generate very many polynomials (and a much smaller eigenregularity parameter). Since the CLT can only take advantage of the largest eigenregularity parameter, the key advantage of **Regularize-One-Wiener** — that the number of polynomials it produces compares favorably with the eigenregularity of these polynomials — is lost.

We get around this with the **MultiRegularize-One-Wiener** procedure that is presented and analyzed below. It augments the **Regularize-One-Wiener** procedure with ideas from the decomposition procedure for k-tuples of degree-2 polynomials that was presented and analyzed in [DDS13b] (and which in turn built on ideas from the decomposition of [GOWZ10] for simultaneously dealing with multiple degree-1 polynomials, using a different notion of "eigenregularity"). Crucially, it guarantees that the *overall* number of polynomials that are produced from all the r decompositions compares favorably with the overall eigenregularity parameter that is obtained.

Theorem 41. *Fix any $q \geq 2$. There is a procedure* **MultiRegularize-One-Wiener** *which takes as input an r-tuple of polynomials $(I_q(f_1), \ldots, I_q(f_r))$ such that $\mathbf{Var}[I_q(f_i)] = 1$ for all i, and input parameters $\eta_0 = 1 \geq \eta_1 \geq \ldots \geq \eta_K$ and ϵ, where $K = O(r/\epsilon \cdot \log(1/\epsilon))$.* **MultiRegularize-One-Wiener** *runs in poly$(n^q, 1/\eta_K, r/\epsilon)$ time and has the following guarantee:*

1. *Define $M(i) = O(\frac{4^q}{\eta_i^2}\log(1/\epsilon))$.* **MultiRegularize-One-Wiener** *outputs an index t with $0 \leq t \leq K$ and for each $s \in [r]$ a set L_s of triples $\{(a_{s,i,j}, P_{s,i,j}, Q_{s,i,j})\}_{i=1,\ldots,t, j=1,\ldots,M(i)}$ and a pair of polynomials $R_{s,\mathrm{reg}}, R_{s,\mathrm{neg}}$, such that*

$$I_q(f_s) = \sum_{i=1}^{t}\sum_{j=1}^{M(i)} a_{s,i,j}P_{s,i,j}\cdot Q_{s,i,j} + a_{s,\mathrm{reg}}\cdot R_{s,\mathrm{reg}} + R_{s,\mathrm{neg}}. \tag{22}$$

2. *For each s, i, j we have $P_{s,i,j} \in \mathcal{W}^{q_{s,i,j,1}}$ and $Q_{s,i,j} \in \mathcal{W}^{q_{s,i,j,2}}$ with $q_{s,i,j,1}, q_{s,i,j,2} > 0$ and $q_{s,i,j,1} + q_{s,i,j,2} = q$ and $\mathbf{Var}[P_{s,i,j}] = \mathbf{Var}[Q_{s,i,j}] = \mathbf{Var}[P_{s,i,j}\cdot Q_{s,i,j}] = 1$. Similarly we have $R_{s,\mathrm{reg}}, R_{s,\mathrm{neg}} \in \mathcal{W}^q$, and $\mathbf{Var}[R_{s,\mathrm{reg}}] = 1$. Moreover $P_{s,i,j}$ and $Q_{s,i,j}$ are over disjoint sets of variables, and all of $P_{s,i,j}, Q_{s,i,j}, R_{s,\mathrm{reg}}$ and $R_{s,\mathrm{neg}}$ are multilinear.*

3. *For each s we have that $\mathbf{Var}[R_{s,\mathrm{neg}}] \leq \epsilon$ and that $a_{s,\mathrm{reg}}\cdot R_{s,\mathrm{reg}}$ is η_{t+1}-eigenregular, where we define $\eta_{K+1} = 0$.*

4. *For $1 \leq s \leq r$ and $1 \leq i \leq t$ we have $\sum_{j=1}^{M(i)}(a_{s,i,j})^2 \leq (2^q/\eta_i)^{4(M(i)-1)}$.*

Proof. Similar to **Regularize-One-Wiener**, the procedure **MultiRegularize-One-Wiener** maintains the invariant that for each $s \in [r]$, we have $I_q(f_s) = \sum_{(a,P,Q)\in L_s} a\cdot P\cdot Q + I_q(g_{s,i})$ throughout its execution.

Before giving the detailed description we provide some useful points to keep in mind. The set $[r] \setminus \mathbf{live}$ contains the indices of those polynomials for which the desired decomposition has already been achieved, while $\mathbf{live}$ contains those polynomials that are still being decomposed. The variable $\mathbf{hit}_s$ maintains the number of times that the decomposition procedure **Decompose-One-Wiener** has been applied to $I_q(g_{s,i})$ for some i.

Here is the procedure **MultiRegularize-One-Wiener**:

i. For all $s \in [r]$ initialize $\mathbf{hit}_s$ to be 0, initialize L_s to be the empty set of triples, and initialize $g_{s,1} = f_s$, so $I_q(g_{s,1}) = I_q(f_s)$. Initialize the set $\mathbf{live}$ to be $[s]$.

ii. For $i = 1$ to K do the following:

ii(a). For each $s \in \mathbf{live}$, if $\mathbf{Var}[I_q(g_{s,i})] \leq \epsilon$ then set $R_{s,\mathrm{neg}} = I_q(g_{s,i})$, set $a_{s,\mathrm{reg}} = 0$ and set $R_{s,\mathrm{reg}}$ to be any unit variance element of $\mathcal{W}^q$ (the choice of $R_{s,\mathrm{reg}}$ is immaterial), and remove s from $\mathbf{live}$.

ii(b). If $\mathbf{live}$ is empty then for each $s \in [r]$ output the set L_s and the pair $R_{s,\mathrm{reg}}, R_{s,\mathrm{neg}}$, and exit. Otherwise, for each $s \in \mathbf{live}$, run **Decompose-One-Wiener** with parameters η_i and ϵ on the polynomial $I_q(\lambda_{s,i} g_s)$, where $\lambda_{s,i}$ is chosen so that $\mathbf{Var}[I_q(\lambda_{s,i} g_s)] = 1$. Let $L_{s,i} = \{(c_{s,i,j}, P_{s,i,j}, Q_{s,i,j})\}$ be the set of (at most $M(i)$ many, by Lemma 36) triples and $R_{s,i}$ be the polynomial that it outputs.

ii(c). If the call to **Decompose-One-Wiener** returned "small remainder" for any polynomial $I_q(\lambda_{s,i} g_s)$, then for each such s set L_s to $L_s \cup L'_{s,i}$ where $L'_{s,i} = \{(\frac{c_{s,i,j}}{\lambda_{s,i}}, P_{s,i,j}, Q_{s,i,j})\}_{(c_{s,i,j},P_{s,i,j},Q_{s,i,j}) \in L_{s,i}}$, set $R_{s,\mathrm{neg}}$ to $R_{s,i}/\lambda_{s,i}$, set $a_{s,\mathrm{reg}} = 0$ and $R_{s,\mathrm{reg}}$ to be any unit variance element of $\mathcal{W}^q$ (as before, the choice of $R_{s,\mathrm{reg}}$ is immaterial), and remove s from $\mathbf{live}$.

ii(d). If for every $s \in \mathbf{live}$ it is the case that $\mathbf{Var}[\sum_{(c_{s,i,j},P_{s,i,j},Q_{s,i,j}) \in L_{s,i}} c_{s,i,j} P_{s,i,j} \cdot Q_{s,i,j}] \leq \epsilon$, then set $R_{s,\mathrm{neg}}$ to $\sum_{(c_{s,i,j},P_{s,i,j},Q_{s,i,j}) \in L_{s,i}} \frac{c_{s,i,j}}{\lambda_{s,i}} P_{s,i,j} \cdot Q_{s,i,j}$. Also, set $a_{s,\mathrm{reg}} = \sqrt{\mathbf{Var}(R_{s,i}/\lambda_{s,i})}$ and and $R_{s,\mathrm{reg}} = R_{s,i}/(\lambda_{s,i} \cdot a_{s,\mathrm{reg}})$. For each $s \in [r]$ output the set L_s, and the triple $a_{s,\mathrm{reg}}, R_{s,\mathrm{reg}}, R_{s,\mathrm{neg}}$, and exit.

ii(e). Otherwise, for each $s \in \mathbf{live}$ such that $\mathbf{Var}[\sum_{(c_{s,i,j},P_{s,i,j},Q_{s,i,j}) \in L_{s,i}} c_{s,i,j} P_{s,i,j} \cdot Q_{s,i,j}] > \epsilon$, increase $\mathbf{hit}_s$ by 1, set L_s to $L_s \cup L'_{s,i}$ where $L'_{s,i} = \{(\frac{c_{s,i,j}}{\lambda_{s,i}}, P_{s,i,j}, Q_{s,i,j})\}_{(c_{s,i,j},P_{s,i,j},Q_{s,i,j}) \in L_{s,i}}$, and set $g_{s,i+1}$ to satisfy $I_q(g_{s,i+1}) = R_{s,i}/\lambda_{s,i}$. Increment i and go to the next iteration of step (ii).

Property (1) follows from the discussion preceding the algorithm description and inspection of step ii(d). Property (2) follows directly from part (2) of Lemma 36 (Note that the algorithm ensures that $R_{s,\mathrm{reg}}$ has unit variance).

We have the following analogue of Claim 40:

Claim 42. *At each stage in the execution of the algorithm, for each $s \in \mathbf{live}$ we have $\mathbf{Var}[I_q(g_{s,i})] \leq (1-\epsilon)^{\mathbf{hit}_s}$.*

Proof. The proof is an easy adaptation of the proof of Claim 40, using the criterion for incrementing $\mathbf{hit}_s$ that is employed in step ii(e). □

Claim 42 implies that for each $s \in \mathbf{live}$ we have $\lambda_{s,i} \geq 1$, so as in the proof of Claim 40 we get that Property (4) holds.

Observe that if an index s is removed from $\mathbf{live}$ (either in Step ii(a) or Step ii(c)), then the polynomial $R_{s,\mathrm{reg}}$ is 0-eigenregular, and since $\lambda_{s,i} \geq 1$, the polynomial $R_{s,\mathrm{neg}}$ has $\mathbf{Var}[R_{s,\mathrm{neg}}] \leq \epsilon$. Hence as a consequent of the above-mentioned invariant, it is easily verified that each $s \in [r] \setminus \mathbf{live}$ satisfies (22).

The last step is to establish Property (3). The key observation is that each time the algorithm increments i in step ii(e) and returns to step ii(a), at least one $s \in [r]$ must have had $\mathbf{hit}_s$ incremented. Once a given value of s has $\mathbf{hit}_s$ reach $O(1/\epsilon \cdot \log(1/\epsilon))$, by Claim 42 it will be the case that s is removed from $\mathbf{live}$ in Step ii(a). Since $K = O(r/\epsilon \cdot \log(1/\epsilon))$, it follows that the algorithm must halt and exit in some iteration of step ii(b) or ii(d). If the algorithm exits in step ii(b) then it is clear from the above discussion that Property (3) holds. Finally, if the algorithm exits in step ii(d), then similar to the final paragraph of the proof of

Theorem 39, the value of t is $i-1$ (since for the elements $s \in \mathbf{live}$ at the start of that execution of step ii(d), the elements of $L_{s,i}$ are not added on to L_s). Similar to before we get that Lemma (36) guarantees that $R_{s,i}$ (and hence $R_{s,\mathrm{reg}}$) is η_i-eigenregular, i.e. η_{t+1}-eigenregular, the fact that $\lambda_{s,i} \geq 1$ ensures that $\mathbf{Var}[R_{s,\mathrm{neg}}] \leq \epsilon$, and hence Property (3) holds. The proof is complete. □

5.3 Beyond the homogeneous case: handling multiple levels of Wiener chaos

In this subsection we describe and analyze our most involved decomposition procedure, **MultiRegularize-Many-Wieners**, for decomposing a $k(d+1)$-tuple consisting of k elements from the j-th Wiener chaos for each $j = 0, \ldots, d$. We will obtain Theorem 31 in the following subsection using **MultiRegularize-Many-Wieners**.

An informal "top-down" description. We begin with an informal description of how the decomposition procedure works. Let $p_1, \ldots, p_k$ be k degree-d multilinear Gaussian polynomials. Each p_i has a unique expansion in terms of symmetric q-tensors $f_{i,q} \in \mathcal{H}^{\odot q}$ as $p_i = \sum_{q=0}^{d} p_{i,q}$, where $p_{i,q} = I_q(f_{i,q})$. For $2 \leq q \leq d$ let OLD_q denote the set of polynomials $\{I_q(f_{i,q})\}_{i=1,\ldots,k}$.

The high-level idea of the decomposition is to "work downward" from higher to lower levels of the Wiener chaos in successive stages, at each stage using **MultiRegularize-One-Wiener** to simultaneously decompose all of the polynomials at the current level. By carefully choosing the eigenregularity parameters at each stage we can ensure that at the end of the decomposition we are left with a collection of "not too many" polynomials (the $A_{i,j,\ell}$'s of Theorem 31) all of which are highly eigenregular.

In a bit more detail, in the first stage we simultaneously decompose the k degree-d polynomials $I_d(f_{1,d}), \ldots, I_d(f_{k,d})$ using the **MultiRegularize-One-Wiener** algorithm with parameters $1 = \eta_0^{(d)} \gg \cdots \gg \eta_{K^{(d)}}^{(d)}$ and $\epsilon^{(d)}$. This generates:

- k polynomials in $\mathcal{W}^d$ that are each $\eta_{t^{(d)}+1}^{(d)}$-eigenregular, for some $1 \leq t^{(d)} \leq K^{(d)} - 1$ where $K^{(d)} \leq O(k/\epsilon^{(d)} \cdot \log(1/\epsilon^{(d)}))$ (intuitively, these should be thought of as "extremely eigenregular" polynomials); these are the $R_{s,\mathrm{reg}}$ polynomials. Let REG_d denote this set of polynomials. It also generates
- For each $1 \leq q \leq d-1$, "not too many" (at most $k \cdot O((k/\epsilon^{(d)}) \log(1/\epsilon^{(d)})) \cdot (1/(\eta_{t^{(d)}}^{(d)})^2) \log(1/\epsilon^{(d)}))$ new polynomials in $\mathcal{W}^q$; these are the $P_{s,i,j}$ and $Q_{s,i,j}$ polynomials that lie in $\mathcal{W}^q$. Let NEW_q denote this set of polynomials.

Let ALL_{d-1} denote the union of OLD_{d-1} and NEW_{d-1}. Note that every element of ALL_{d-1} belongs to $\mathcal{W}^{d-1}$, and that the size of $|ALL_{d-1}|$ is upper bounded by

$$k \cdot O((k/\epsilon^{(d)}) \log(1/\epsilon^{(d)})) \cdot (1/(\eta_{t^{(d)}}^{(d)})^2) \log(1/\epsilon^{(d)}).$$

The key qualitative point is that after this first stage of decomposition, the number of eigenregular polynomials that have been produced is $|REG_d| \leq k$, and each such polynomial is $\eta_{t^{(d)}+1}^{(d)}$-eigenregular, while the number of polynomials of lower degree that remain to be dealt with (the pieces that came from the original polynomial plus the elements of $NEW_2 \cup \cdots \cup NEW_{d-1}$) is upper bounded in terms of $k, \epsilon^{(d)}$ and $\eta_{t^{(d)}}^{(d)}$.

In the second stage we simultaneously decompose all the elements of ALL_{d-1} by applying the **MultiRegularize-One-Wiener** algorithm to those $|ALL_{d-1}|$ polynomials, using input parameters $1 = \eta_0^{(d-1)} \geq \eta_1^{(d-1)} \geq \cdots \geq \eta_{K^{(d-1)}}^{(d-1)}$ where $K^{(d-1)} = O(|ALL_{d-1}|/\epsilon^{(d-1)} \cdot \log(1/\epsilon^{(d-1)}))$ and $\epsilon^{(d-1)}$. This generates

- at most $|ALL_{d-1}|$ polynomials in $\mathcal{W}^{d-1}$ that are each $\eta^{(d-1)}_{t^{(d-1)}+1}$-eigenregular, for some $1 \leq t^{(d-1)} \leq K^{(d-1)} - 1$ (intuitively, these should be thought of as "extremely eigenregular" polynomials); these are the $R_{s,\mathrm{reg}}$ polynomials. Let REG_{d-1} denote this set of polynomials. It also generates

- For each $1 \leq q \leq d-2$, "not too many" (at most $|ALL_{d-1}| \cdot K^{(d-1)}(1/(\eta^{(d-1)}_{t^{(d-1)}})^2)\log(1/\epsilon^{(d-1)})$) new polynomials in $\mathcal{W}^q$; these are the $P_{s,i,j}$ and $Q_{s,i,j}$ polynomials that lie in $\mathcal{W}^q$. Add these polynomials to the set NEW_q.

Similar to above, the key qualitative point to observe is that the number of eigenregular polynomials that have been produced is $|REG_d| + |REG_{d-1}|$, which is a number depending only on $k, \epsilon^{(d)}, \epsilon^{(d-1)}$, and $1/\eta^{(d)}_{t^{(d)}}$, while the eigenregularity of each such polynomial is at most $\max\{\eta^{(d)}_{t^{(d)}+1}, \eta^{(d-1)}_{t^{(d-1)}+1}\}$, where the first expression inside the max comes from the polynomials in $|REG_d|$ and the second from the polynomials in $|REG_{d-1}|$. By setting the η-parameters so that $\eta^{(d)}_{t^{(d)}+1}$ and $\eta^{(d-1)}_{t^{(d-1)}+1}$ are both much smaller than $\eta^{(d)}_{t^{(d)}}$, we can ensure that the number of polynomials that are produced compares favorable with their eigenregularity.

Continuing in this fashion, the crux of the analysis is to argue that this can be done "all the way down," so that the total number of polynomials that are ever produced in the analysis is far smaller than $1/\eta$, where η is the largest eigenregularity of any of these polynomials. However, it is somewhat awkward to argue this using the "top-down" view on the decomposition procedure that we have adopted so far. Instead, in the formal proof which we give below, we take a "bottom-up" view of the decomposition procedure: we first show that it can be successfully carried out for low-degree polynomials, and use this fact to show that it can be successfully carried out for higher-degree polynomials.

5.3.1 The MultiRegularize-Many-Wieners procedure and its analysis

Now we present and analyze the actual **MultiRegularize-Many-Wieners** procedure. Theorem 43 gives a performance bound on the procedure. Its proof will be by induction on the degree: we first establish the result for degree 2 and then use the fact that the theorem holds for degrees $2, \ldots, d-1$ to prove the result for degree d.

Theorem 43. *Fix $d \geq 2$ and fix any non-increasing computable function $\beta : [1, \infty) \to (0, 1)$ that satisfies $\beta(x) \leq 1/x$. There is a procedure* **MultiRegularize-Many-Wieners**$_{d,\beta}$ *with the following properties. The procedure takes as input the following:*

- *It is given k lists of $d+1$ multilinear Gaussian polynomials; the s-th list is $p_{s,0}, \ldots, p_{s,d}$ where $p_{s,q} \in \mathcal{W}^q$ and $\mathbf{Var}[p_{s,q}] = 1$ for $1 \leq q \leq d$.*

- *It also takes as input a parameter $\tau > 0$.*

The procedure runs in $\mathrm{poly}(n^d) \cdot O_{k,d,\tau}(1)$ time and outputs, for each input polynomial $p_{s,q}$, a polynomial $\mathrm{Out}(p_{s,q})$ and a collection of polynomials that we denote $\{\mathrm{In}(p_{s,q})_\ell\}_{\ell=1,\ldots,\mathrm{num}(p_{s,q})}$; here $\mathrm{num}(p_{s,q})$ is the number of arguments of the polynomial $\mathrm{Out}(p_{s,q})$. ("Out" stands for "outer" and "In" stands for "inner".)

For $s = 1, \ldots, k$, $0 \leq q \leq d$ and $x \in \mathbb{R}^n$, let

$$\tilde{p}_{s,q}(x) = \mathrm{Out}(p_{s,q})\left(\mathrm{In}(p_{s,q})_1(x), \ldots, \mathrm{In}(p_{s,q})_{\mathrm{num}(p_{s,q})}(x)\right) \tag{23}$$

(Intuitively, each $\tilde{p}_{s,q}$ is a polynomial that has been decomposed into constituent sub-polynomials $\mathrm{In}(p_{s,q})_1, \ldots, \mathrm{In}(p_{s,q})_{\mathrm{num}(p_s)_q}$; $\tilde{p}_{s,q}$ is meant to be a good approximator for $p_{s,q}$. The following conditions make this precise.)

The following conditions hold:

1. *For each $s \in [k], 0 \le q \le d$ the polynomial $\tilde{p}_{s,q}(x)$ belongs to the q-th Wiener chaos $\mathcal{W}^q$. Additionally, each polynomial $\mathrm{In}(p_{s,q})_\ell$ with $q \ge 1$ lies in $\mathcal{W}^j$ for some $1 \le j \le d$ and has $\mathbf{Var}[\mathrm{In}(p_{s,q})_\ell] = 1$.*

2. *For each $s \in [k], 0 \le q \le d$, we have $\mathbf{Var}[p_{s,q} - \tilde{p}_{s,q}] \le \tau$.*

3. *Each polynomial $\mathrm{Out}(p_{s,q})$ is a multilinear polynomial in its $\mathrm{num}(p_{s,q})$ arguments. Moreover,there exists $N = N_\beta(k,d,\tau)$ and $M = M_\beta(k,d,\tau)$ such that if $\mathrm{Coeff}(p_{s,q})$ denotes the sum of the absolute values of the coefficients of $\mathrm{Out}(p_{s,q})$, then $\sum_{s,q} \mathrm{Coeff}(p_{s,q}) \le M$ and $\sum_{s,q} \mathrm{num}(p_{s,q}) \le N$.*

4. *Further, let $\mathrm{Num} = \sum_{s=1,\dots,k,q=0,\dots,d} \mathrm{num}(p_{s,q})$ and $\mathrm{Coeff} = \sum_{s=1,\dots,k,q=0,\dots,d} \mathrm{Coeff}(p_{s,q})$. Then, each polynomial $\mathrm{In}(p_{s,q})_\ell$ is $\beta(\mathrm{Num} + \mathrm{Coeff})$-eigenregular.*

After proving Theorem 43, in the next subsection we will obtain Theorem 31 from it as a special case, by writing the degree-d polynomial p as $\sum_{q=0}^d p_q$ where $q \in \mathcal{W}^q$ and applying Theorem 43 to $(p_0, \dots, p_d)$.

Base case: Proof of Theorem 43 in the case $d = 2$. Fix any non-increasing function $\beta : [1,\infty) \to (0,1)$ that satisfies $\beta(x) \le 1/x$. The main step is to use the **MultiRegularize-One-Wiener** procedure on the vector of (at most k) polynomials $\{p_{s,2}\}_{s\in[k]}$; we now specify precisely how to set the parameters of this procedure. Fix $\epsilon = \tau$ and let $K = O(k/\epsilon \cdot \log(1/\epsilon))$ as in the statement of Theorem 41. We define the parameters $1 = \eta_0 \ge \eta_1 \cdots \ge \eta_K > n_{K+1} = 0$ as follows: for $t = 0, \dots, K-1$, we have

$$\eta_{t+1} := \beta\left(C\frac{k^2}{\epsilon} \cdot \frac{1}{\eta_t^2} \cdot (\log 1/\epsilon)^2 + C' \cdot k^3 \cdot \left(\frac{4}{\eta_t}\right)^{C' \cdot \frac{1}{\eta_t^2} \cdot \log(1/\epsilon)} \right) \tag{24}$$

where $C, C' \ge 1$ are absolute constants defined below (see (27) and (28); intuitively, the first term correspond to an upper bound on Num whereas the second term corresponds to an upper bound on Coeff). (Note the assumption that $\beta(x) \le 1/x$ implies that indeed $\eta_{t+1} \le \eta_t$.) When called with these parameters on the vector of polynomials $\{p_{s,2}\}_{s\in[k]}$, by Theorem 41 **MultiRegularize-One-Wiener** outputs an index t with $0 \le t \le K$ and a decomposition of each $p_{s,2}$ as

$$p_{s,2} = \sum_{i=1}^{t} \sum_{j=1}^{M(i)} a_{s,i,j} P_{s,i,j} \cdot Q_{s,i,j} + a_{s,\mathrm{reg}} \cdot R_{s,\mathrm{reg}} + R_{s,\mathrm{neg}}, \tag{25}$$

(recall that $M(i) = \frac{1}{\eta_i^2}\log(1/\epsilon)$), where for each $s \in [k]$,

1. $\mathbf{Var}[R_{s,\mathrm{neg}}] \le \epsilon = \tau$,

2. each $a_{s,\mathrm{reg}} \cdot R_{s,\mathrm{reg}}$ is η_{t+1}-eigenregular, $R_{s,\mathrm{reg}}$ has variance 1, and

3. For each s, i, j, the polynomials $P_{s,i,j}$ and $Q_{s,i,j}$ are both in $\mathcal{W}^1$ and are defined over disjoint sets of variables and have $\mathbf{Var}[P_{s,i,j}] = \mathbf{Var}[Q_{s,i,j}] = 1$. Moreover $R_{s,\mathrm{neg}}$ and $R_{s,\mathrm{reg}}$ both lie in $\mathcal{W}^2$.

We now describe the polynomials $\mathrm{In}(p_{s,2})$ and $\mathrm{Out}(p_{s,2})$, whose existence is claimed in Theorem 43, for each $s \in [k]$. For each $s \in [k]$, the polynomials $\mathrm{In}(p_{s,2})_\ell$ include all of the polynomials $P_{s,i,j}, Q_{s,i,j}$ from (25). Furthermore, $R_{s,\mathrm{reg}}$ belongs to $\mathrm{In}(p_{s,2})_\ell$ if and only if $a_{s,\mathrm{reg}} \ne 0$. The polynomial $\tilde{p}_{s,2}(x)$ is $p_{s,2} - R_{s,\mathrm{neg}}$, and we have

$$\tilde{p}_{s,2}(x) = \mathrm{Out}(p_{s,2})(\{\mathrm{In}(p_{s,2})_\ell\}) = \sum_{i=1}^{t} \sum_{j=1}^{M(i)} a_{s,i,j} \cdot P_{s,i,j} \cdot Q_{s,i,j} + a_{s,\mathrm{reg}} \cdot R_{s,\mathrm{reg}}. \tag{26}$$

Further, by the guarantee from Theorem 41, we get that for any s, i, j, $P_{s,i,j}$ and $Q_{s,i,j}$ are on disjoint sets of variables. The degree-1 and degree-0 portions are even simpler: for each $s \in [k]$ we have

$$\tilde{p}_{s,1}(x) = \mathrm{Out}(p_{s,1})(p_{s,1}) = p_{s,1} \quad \text{and} \quad \tilde{p}_{s,0}(x) = \mathrm{Out}(p_{s,0})(p_{s,0}) = p_{s,0}.$$

It is clear from this description that property (1) holds (for each s, q, the polynomial $\tilde{p}_{s,q}$ indeed belongs to $\mathcal{W}^q$).

Next, we show that Condition 2 holds. This is immediate for $q = 0, 1$ since the polynomials $p_{s,0}$ and $p_{s,1}$ are identical to $\tilde{p}_{s,0}$ and $\tilde{p}_{s,1}$ respectively. For $q = 2$ given any $s \in [k]$, we have that $p_{s,2} - \tilde{p}_{s,2} = R_{s,\mathrm{neg}}$, and hence by the upper bound on $\mathbf{Var}[R_{s,\mathrm{neg}}]$ (see Item (1) above) we have that $\mathbf{Var}[p_{s,2} - \tilde{p}_{s,2}] = \mathbf{Var}[R_{s,\mathrm{neg}}] \leq \tau$ as required by Condition (2).

For Condition (3), the multilinearity of each $\mathrm{Out}(p_{s,q})$ is easily seen to hold as a consequence of Theorem 41. To check the remaining part of Condition (3), note that for $q = 0, 1$ the polynomial $\tilde{p}_{s,q}$ is simply the identity polynomial $x \mapsto x$. For $q = 2$, Equation (26) immediately gives that

$$\mathrm{num}(p_{s,2}) \leq 2t \cdot M(t) + 1 \leq 2K \cdot M(t) + 1$$

Observe that since $\mathrm{Out}(p_{s,0})(\cdot)$ and $\mathrm{Out}(p_{s,1})(\cdot)$ is simply the identity map $x \mapsto x$. Thus,

$$\sum_{q=0}^{2} \mathrm{num}(p_{s,q}) \leq 2K \cdot M(t) + 3.$$

As $s \in [k]$,

$$\mathrm{Num} = \sum_{s \in [k]} \sum_{q=0}^{2} \mathrm{num}(p_{s,q}) \leq 2K \cdot k \cdot M(t) + 3k.$$

Note that we can choose C to be a sufficiently large constant (independent of t) so that

$$\mathrm{Num} \leq 2K \cdot k \cdot M(t) + 3k \leq C \cdot \frac{k^2}{\epsilon} \log^2 \frac{1}{\epsilon} \cdot \frac{1}{\eta_t^2}. \tag{27}$$

We next upper bound the sum of the absolute values of the coefficients appearing in $\mathrm{Out}(p_{s,q})$. We begin by observing that $\mathrm{Out}(p_{s,1})$ and $\mathrm{Out}(p_{s,0})$ are just the identity function and hence the absolute values of the coefficients is just 1. For $\mathrm{Out}(p_{s,2})$, note that Item (4) of Theorem 41, gives that

$$\sum_{i=1}^{t} \sum_{j=1}^{M(i)} a_{s,i,j}^2 \leq \sum_{i=1}^{t} \left(\frac{4}{\eta_i}\right)^{4(M(i)-1)} \leq t \cdot \left(\frac{4}{\eta_t}\right)^{4(M(t)-1)}.$$

Thus, summing over all $s \in [k]$, we get

$$\sum_{s \in [k]} \sum_{i=1}^{t} \sum_{j=1}^{M(i)} a_{s,i,j}^2 \leq k \cdot t \cdot \left(\frac{4}{\eta_t}\right)^{4(M(t)-1)}.$$

Recalling that $\sum_{s \in [k]} \sum_{i=1}^{t} M(i) \leq 2K \cdot k \cdot M(t) + 3k$ and applying Cauchy-Schwarz, we get

$$\sum_{s \in [k]} \sum_{i=1}^{t} |a_{s,i,j}| \leq \left(t \cdot \left(\frac{4}{\eta_t}\right)^{4(M(t)-1)}\right) \cdot (2K \cdot k \cdot M(t) + 3k)$$

Thus, we can choose a sufficiently large constant C' (independent of t) such that

$$\text{Coeff} = \sum_{s\in[k]} \sum_{i=1}^{t} |a_{s,i,j}| \leq C' \cdot k^3 \cdot \left(\frac{4}{\eta_t}\right)^{C' \cdot \frac{1}{\eta_t^2} \cdot \log(1/\epsilon)} \tag{28}$$

Using (28) and (27), the recursive definition of η_t from (24) and $t \leq K$, there exists $N = N_\beta(k, d, \tau)$ and $M = M_\beta(k, d, \tau)$ such that $\text{Num} \leq N$ and $\text{Coeff} \leq M$.

It remains only to bound the eigenregularity of the $\text{In}(p_{s,q})_\ell$ polynomials. Each such polynomial is either a degree-0 polynomial $p_{s,0}$, a degree-1 polynomial $p_{s,1}$ or $P_{s,i,j}$ or $Q_{s,i,j}$ (from (26)), or a degree-2 polynomial $R_{s,\text{reg}}$ for which $a_{s,\text{reg}} \neq 0$. Since degree-0 and degree-1 polynomials are 0-eigenregular, it remains only to bound the eigenregularity of each $R_{s,\text{reg}}$. Since each $R_{s,\text{reg}}$ is η_{t+1}-eigenregular, our choice of the sequence $1 = \eta_0 \geq \cdots \geq \eta_K$ (see (24)) implies that each $R_{s,\text{reg}}$ is indeed $\beta(\text{Num} + \text{Coeff})$-eigenregular as required by Condition (4). This concludes the proof of the degree-2 base case of the induction for Theorem 43. □

Inductive step: Proof of 43. Fix a value $d \geq 3$; with the base case in hand from above, we may assume that Theorem 43 holds for degrees $2, \ldots, d-1$. Similar to the degree-2 base case, the first main step of the algorithm is to use the **MultiRegularize-One-Wiener** procedure on the vector of polynomials $(p_{s,d})_{s\in[k]}$; we now specify how to set the parameters of **MultiRegularize-One-Wiener**. Fix $\epsilon = \tau/8$ and let $K = O(k/\epsilon \cdot \log(1/\epsilon))$ as in the statement of Theorem 41. We define the parameters $1 = \eta_0 \geq \eta_1 \cdots \geq \eta_K > n_{K+1} = 0$ as follows: for $t = 0, \ldots, K-1$, we have

$$\eta_{t+1} := \beta\left(N_{\beta_t^*}\left(L_t, d-1, \frac{\tau}{16 \cdot L_t \cdot L_t^{L_t}}\right)^2 + M_{\beta_t^*}\left(L_t, d-1, \frac{\tau}{16 \cdot L_t \cdot L_t^{L_t}}\right)^2 \cdot L_t^{L_t} + 2k + L_t^{L_t}\right), \tag{29}$$

$$\text{where} \quad L_t \stackrel{\text{def}}{=} C' \cdot \frac{k^2 d}{\epsilon} \cdot \frac{1}{\eta_t^2} \cdot (\log 1/\epsilon)^2,$$

and $C' > 0$ is an absolute constant defined below and β_t^* is defined in (31). The reader can verify that the sequence $\{\eta_t\}$ is defined consistently. As before, intuitively, the first term in the argument to β corresponds to an upper bound on Num whereas the second term corresponds to an upper bound on Coeff. (Note that from the recursive definition (29), for all $t = 1, \ldots, K$ we have that $\eta_t = \kappa_t(k, \tau, d)$ for some function κ_t; this will be useful later.) When called with these parameters on the vector of polynomials $(p_{s,d})_{s\in[k]}$, by Theorem 41 **MultiRegularize-One-Wiener** outputs an index t with $0 \leq t \leq K$ and a decomposition of each $p_{s,d}$ as

$$p_{s,d} = \sum_{i=1}^{t} \sum_{j=1}^{M(i)} a_{s,i,j} P_{s,i,j} \cdot Q_{s,i,j} + a_{s,\text{reg}} \cdot R_{s,\text{reg}} + R_{s,\text{neg}}, \tag{30}$$

(recall that $M(i) = \frac{1}{\eta_i^2}\log(1/\epsilon)$), where for each $s \in [k]$,

- $\mathbf{Var}[R_{s,\text{neg}}] \leq \epsilon = \tau/8$,
- each $a_{s,\text{reg}} \cdot R_{s,\text{reg}}$ is η_{t+1}-eigenregular, $R_{s,\text{reg}}$ has variance 1, and
- For each s, i, j, the polynomial $P_{s,i,j}$ belongs to $\mathcal{W}^{q_{s,i,j,1}}$ and $Q_{s,i,j}$ belongs to $\mathcal{W}^{q_{s,i,j,2}}$ where $0 < q_{s,i,j,1}, q_{s,i,j,2}$, $q_{s,i,j,1} + q_{s,i,j,2} = d$, and $P_{s,i,j}, Q_{s,i,j}$ are defined over disjoint sets of variables and have $\mathbf{Var}[P_{s,i,j}] = \mathbf{Var}[Q_{s,i,j}] = 1$. Moreover $R_{s,\text{neg}}$ and $R_{s,\text{reg}}$ both lie in $\mathcal{W}^d$.

Define the function $\beta_t^* : [1, \infty) \to (0, 1)$, $\beta_t^*(x) \leq 1/x$, as

$$\beta_t^*(x) := \beta(x^2 \cdot k \cdot L_t^{L_t} + k + L_t^{L_t}). \tag{31}$$

The second main step of the algorithm is to run the procedure **MultiRegularize-Many-Wieners**$_{d-1,\beta_t^*}$ with its inputs set in the following way:

- There are k lists of d multilinear polynomials $p_{s,0}, \dots, p_{s,d-1}$ as s ranges from 1 to k. Additionally, for each of the $P_{s,i,j}$ and $Q_{s,i,j}$ polynomials that are obtained from (30), there is a list of d multilinear polynomials in $\mathcal{W}^0, \dots, \mathcal{W}^{d-1}$. In each such list all the polynomials are 0 except for the $P_{s,i,j}$ or $Q_{s,i,j}$ polynomial. (Note that each of these polynomials indeed belongs to a single specific layer $\mathcal{W}^q$ of the Wiener chaos for some $1 \leq q \leq d-1$, as required by **MultiRegularize-Many-Wieners**$_{d-1,\beta_t^*}$; note further that since the variance of each $P_{s,i,j}$ and $Q_{s,i,j}$ polynomial is 1, the variance condition in the first bullet of Theorem 43 is satisfied. Finally, we emphasize that the $R_{s,\mathrm{reg}}$ polynomial, which lies in $\mathcal{W}^d$, is not used here.) Thus, note that we can choose C' to be a sufficiently large constant so that the total number of input polynomials that are given to the procedure is at most

  $$L_t \stackrel{\text{def}}{=} C' \cdot \frac{k^2 d}{\epsilon} \cdot \frac{1}{\eta_t^2} \cdot (\log 1/\epsilon)^2. \tag{32}$$

  Note that $L_t = \kappa_1(k, \tau, d)$ for some function κ_1; this will be useful later. Another thing that will be useful is the following. Define

  $$A_d = (\sum_{s \in [k]} \sum_{i=1}^{t} \sum_{j=1}^{M(i)} |a_{s,i,j}|)^2. \tag{33}$$

  Note that mimicking the calculations preceding (28), it can be shown that C' can be chosen so that $A_d \leq L_t^{L_t}$.

- The "τ" parameter of the procedure is set to $\frac{\tau}{16 \cdot L_t \cdot L_t^{L_t}}$.

By the inductive hypothesis, when invoked this way, the procedure **MultiRegularize-Many-Wieners**$_{d-1,\beta_t^*}$ returns the following:

- For each $P_{s,i,j}$ and $Q_{s,i,j}$ from (30), outer polynomials $\mathrm{Out}(P_{s,i,j})$ and $\mathrm{Out}(Q_{s,i,j})$, and likewise for each $p_{s,q}$ ($1 \leq s \leq k, 0 \leq q \leq d-1$) an outer polynomial $\mathrm{Out}(p_{s,q})^{(d-1)}$; the "$(d-1)$" superscript here is to emphasize that this polynomial is obtained from the call to **MultiRegularize-Many-Wieners**$_{d-1,\beta_t^*}$.

- For each $P_{s,i,j}$ and $Q_{s,i,j}$ from (30), a collection of "inner polynomials"

  $$\{\mathrm{In}(P_{s,i,j})_\ell\}_{\ell=1,\dots,\mathrm{num}(P_{s,i,j})} \quad \text{and} \quad \{\mathrm{In}(Q_{s,i,j})_\ell\}_{\ell=1,\dots,\mathrm{num}(Q_{s,i,j})} \tag{34}$$

  and likewise for each $p_{s,q}$ ($1 \leq s \leq k, 0 \leq q \leq d-1$) a collection of inner polynomials

  $$\{\mathrm{In}(p_{s,q})_\ell^{(d-1)}\}_{\ell=1,\dots,\mathrm{num}(p_{s,q})}; \tag{35}$$

  similar to above, the "$(d-1)$" superscript here is to emphasize that these polynomials are obtained from a call to **MultiRegularize-Many-Wieners**$_{d-1,\beta_t^*}$.

For each $p_{s,q}$ with $1 \leq s \leq k, 0 \leq q \leq d-1$, let us write $\tilde{p}_{s,q}^{(d-1)}$ to denote $\mathrm{Out}(p_{s,q})^{(d-1)}(\{\mathrm{In}(p_{s,q})_\ell^{(d-1)}\})$.

The pieces are now in place for us to describe the polynomials $\mathrm{Out}(p_{s,q})$ and $\{\mathrm{In}(p_{s,q})_\ell\}_{\ell=1,\ldots,\mathrm{num}(p_{s,q})}$ whose existence is asserted by Theorem 43.

Fix any $s \in [k]$. We begin with the easy case of $q < d$; so fix any $0 \leq q < d$. The polynomial $\mathrm{Out}(p_{s,q})$ is simply $\mathrm{Out}(p_{s,q})^{(d-1)}$, and the polynomials $\{\mathrm{In}(p_{s,q})_\ell\}$ are simply the polynomials $\{\mathrm{In}(p_{s,q})_\ell^{(d-1)}\}$.

Now we turn to the more involved case of $q = d$. Fix any $s \in [k]$. Recall the decomposition of $p_{s,d}$ given by (30). Fix any $i \in [t], j \in [M(i)]$ and consider first the polynomial $P_{s,i,j}$ from (30). By the inductive hypothesis, the call to **MultiRegularize-Many-Wieners**$_{d-1,\beta_t^*}$ yields a polynomial $\tilde{P}_{s,i,j}$ which is defined via (23) in terms of the $\mathrm{Out}(P_{s,i,j})$ and $\mathrm{In}(P_{s,i,j})_\ell$ polynomials, namely

$$\tilde{P}_{s,i,j} = \mathrm{Out}(P_{s,i,j})(\mathrm{In}(P_{s,i,j})_1, \ldots, \mathrm{In}(P_{s,i,j})_{\mathrm{num}(P_{s,i,j})}). \tag{36}$$

Similarly, considering the polynomial $Q_{s,i,j}$ from (30), the call to **MultiRegularize-Many-Wieners**$_{d-1,\beta_t^*}$ also yields a polynomial $\tilde{Q}_{s,i,j}$ which is defined via (23) in terms of the $\mathrm{Out}(Q_{s,i,j})$ and $\mathrm{In}(Q_{s,i,j})_\ell$ polynomials, namely

$$\tilde{Q}_{s,i,j} = \mathrm{Out}(Q_{s,i,j})(\mathrm{In}(Q_{s,i,j})_1, \ldots, \mathrm{In}(Q_{s,i,j})_{\mathrm{num}(Q_{s,i,j})}). \tag{37}$$

The polynomials $\{\mathrm{In}(p_{s,d})_\ell\}$ are the elements of

$$\left(\bigcup_{i=1}^{t} \bigcup_{j=1}^{M(i)} \left(\{\mathrm{In}(P_{s,i,j})_{\ell'}\}_{\ell'=1,\ldots,\mathrm{num}(P_{s,i,j})} \cup \{\mathrm{In}(Q_{s,i,j})_{\ell''}\}_{\ell''=1,\ldots,\mathrm{num}(Q_{s,i,j})} \right) \right) \cup \{R_{s,\mathrm{reg}} : a_{s,\mathrm{reg}} \neq 0\} \tag{38}$$

and the polynomial $\mathrm{Out}(p_{s,d})$ is given by

$$\mathrm{Out}(p_{s,d})(\{\mathrm{In}(p_{s,d})_\ell\}) = \sum_{i=1}^{t} \sum_{j=1}^{M(i)} a_{s,i,j} \tilde{P}_{s,i,j} \cdot \tilde{Q}_{s,i,j} + a_{s,\mathrm{reg}} \cdot R_{s,\mathrm{reg}}. \tag{39}$$

Recalling (23), (36) and (37), we may write this more explicitly as

$$\tilde{p}_{s,d} = \sum_{i=1}^{t} \sum_{j=1}^{M(i)} a_{s,i,j} \cdot (\mathrm{Out}(P_{s,i,j})(\{\mathrm{In}(P_{s,i,j})_\ell\})) \cdot (\mathrm{Out}(Q_{s,i,j})(\{\mathrm{In}(Q_{s,i,j})_{\ell'}\})) + a_{s,\mathrm{reg}} \cdot R_{s,\mathrm{reg}}. \tag{40}$$

This concludes the specification of the $\mathrm{Out}(p_{s,q})$, $\{\mathrm{In}(p_{s,q})_\ell\}$ and $\tilde{p}_{s,q}$ polynomials; it remains to show that these polynomials satisfy all of the claimed Conditions (1)-(4).

First we consider Condition (1), that for each s, q the polynomial $\tilde{p}_{s,q}$ belongs to the q-th Wiener chaos $\mathcal{W}^q$ and each polynomial $\mathrm{In}(p_{s,q})_\ell$ has $\mathbf{Var}[\mathrm{In}(p_{s,q})_\ell] = 1$. For $q < d$ this follows from the inductive hypothesis. For $q = d$, consider (30) and (39). Part (2) of Theorem 41 ensures that $R_{s,\mathrm{reg}}$ and each $P_{s,i,j} \cdot Q_{s,i,j}$ all lie in $\mathcal{W}^d$ and that $P_{s,i,j}$ and $Q_{s,i,j}$ are on disjoint sets of variables. Using the inductive hypothesis applied to each $P_{s,i,j}$ and $Q_{s,i,j}$ and the fact $P_{s,i,j}$ and $Q_{s,i,j}$ are on disjoint sets of variables, each product $\tilde{P}_{s,i,j} \cdot \tilde{Q}_{s,i,j}$ also must lie in $\mathcal{W}^d$. Since $R_{s,\mathrm{reg}}$ lies in $\mathcal{W}^d$ and $\mathcal{W}^d$ is a subspace, we get that $\tilde{p}_{s,d} \in \mathcal{W}^d$ for all $s \in [k]$. To see that each polynomial in $\{\mathrm{In}(P_{s,i,j})_\ell\}$ and $\{\mathrm{In}(Q_{s,i,j})_{\ell'}\}$ has variance 1, we use the induction hypothesis. Also, note that by Theorem 41, $\mathbf{Var}(R_{s,\mathrm{reg}}) = 1$. This shows that all elements of $\{\mathrm{In}(p_{s,q})_\ell\}$ have variance 1 as claimed.

Next let us consider Condition (2), i.e. we must upper bound each $\mathbf{Var}[\tilde{p}_{s,q} - p_{s,q}]$. Fix any $s \in [k]$ and first consider some $1 \leq q \leq d-1$. In this case we have that

$$\begin{aligned} \tilde{p}_{s,q} &= \mathrm{Out}(p_{s,q}) \left(\mathrm{In}(p_{s,q})_1(x), \ldots, \mathrm{In}(p_{s,q})_{\mathrm{num}(p_{s,q})}(x) \right) \\ &= \mathrm{Out}(p_{s,q})^{(d-1)} \left(\mathrm{In}(p_{s,q})_1^{(d-1)}(x), \ldots, \mathrm{In}(p_{s,q})_{\mathrm{num}(p_{s,q})}^{(d-1)}(x) \right) \\ &= \tilde{p}_{s,q}^{(d-1)} \end{aligned}$$

so we get the desired bound on $\mathbf{Var}[\tilde{p}_{s,q} - p_{s,q}]$ from the inductive hypothesis, recalling that the "τ" parameter that was provided to **MultiRegularize-Many-Wieners**$_{d-1,\beta_t^*}$ was set to $\tau/(16 \cdot L_t \cdot L_t^{L_t}) \leq \tau$.

Next, consider the case $q = d$. In this case,

$$p_{s,d} - \tilde{p}_{s,d} = R_{s,\text{neg}} + \sum_{i=1}^{t} \sum_{j=1}^{M(i)} a_{s,i,j}(\tilde{P}_{s,i,j} \cdot \tilde{Q}_{s,i,j} - P_{s,i,j} \cdot Q_{s,i,j})$$

We will record the following fact which will be used a couple of times.

Fact 44. $\mathbf{Var}[X_1 + \ldots + X_m] \leq \sum_{i=1}^{m} m \cdot \mathbf{Var}[X_i]$.

First, applying Fact 44, we can say that

$$\mathbf{Var}[p_{s,d} - \tilde{p}_{s,d}] \leq 2\,\mathbf{Var}[R_{s,\text{neg}}] + 2\,\mathbf{Var}\left[\sum_{i=1}^{t} \sum_{j=1}^{M(i)} a_{s,i,j}(\tilde{P}_{s,i,j} \cdot \tilde{Q}_{s,i,j} - P_{s,i,j} \cdot Q_{s,i,j})\right].$$

We recall that $\mathbf{Var}[R_{s,\text{neg}}] \leq \tau/8$. To bound the second term, define

$$\kappa_{\max} = \max_{i,j} \mathbf{Var}(\tilde{P}_{s,i,j} \cdot \tilde{Q}_{s,i,j} - P_{s,i,j} \cdot Q_{s,i,j}).$$

Then, we have that

$$\begin{aligned}
&\mathbf{Var}\left[\sum_{i=1}^{t} \sum_{j=1}^{M(i)} a_{s,i,j}(\tilde{P}_{s,i,j} \cdot \tilde{Q}_{s,i,j} - P_{s,i,j} \cdot Q_{s,i,j})\right] \\
\leq\ & \left(\sum_{i=1}^{t} M(i)\right) \cdot \left(\sum_{i=1}^{t} \sum_{j=1}^{M(i)} a_{s,i,j}^2\, \mathbf{Var}(\tilde{P}_{s,i,j} \cdot \tilde{Q}_{s,i,j} - P_{s,i,j} \cdot Q_{s,i,j})\right) \\
\leq\ & \left(\sum_{i=1}^{t} M(i)\right) \cdot \left(\sum_{i=1}^{t} \sum_{j=1}^{M(i)} a_{s,i,j}^2\right) \cdot \kappa_{\max} \\
\leq\ & \left(\sum_{i=1}^{t} M(i)\right) \cdot \left(\sum_{i=1}^{t} \sum_{j=1}^{M(i)} |a_{s,i,j}|\right)^2 \cdot \kappa_{\max} \\
\leq\ & L_t \cdot A_d \cdot \kappa_{\max} \leq L_t \cdot L_t^{L_t} \cdot \kappa_{\max}
\end{aligned}$$

In the above, the first inequality uses Fact 44, the fourth inequality uses (32) and (33) and the fifth inequality uses the bound on A_d. It remains to bound $\kappa_{\max}$. However, note that for any $1 \leq i \leq t$ and $1 \leq j \leq M(i)$, we have that

$$\begin{aligned}
\mathbf{Var}(\tilde{P}_{s,i,j} \cdot \tilde{Q}_{s,i,j} - P_{s,i,j} \cdot Q_{s,i,j}) &\leq 2 \cdot \mathbf{Var}(Q_{s,i,j} \cdot (\tilde{P}_{s,i,j} - P_{s,i,j})) + 2 \cdot \mathbf{Var}(\tilde{P}_{s,i,j} \cdot (\tilde{Q}_{s,i,j} - Q_{s,i,j})) \\
&\leq 2\,\mathbf{Var}(Q_{s,i,j}) \cdot \mathbf{Var}(\tilde{P}_{s,i,j} - P_{s,i,j}) + 2 \cdot \mathbf{Var}(\tilde{P}_{s,i,j}) \cdot \mathbf{Var}(\tilde{Q}_{s,i,j} - Q_{s,i,j})) \\
&\leq 2\,\mathbf{Var}(\tilde{P}_{s,i,j} - P_{s,i,j}) + 4\,\mathbf{Var}(\tilde{Q}_{s,i,j} - Q_{s,i,j})) \leq \frac{3 \cdot \tau}{8 \cdot L_t \cdot L_t^{L_t}}.
\end{aligned}$$

In the above, the first inequality uses Fact 44 and the second inequality uses that $P_{s,i,j}$ and $Q_{s,i,j}$ are on disjoint sets of variables. The third inequality uses that $\mathbf{Var}(Q_{s,i,j}) = 1$ and that $\mathbf{Var}(\tilde{P}_{s,i,j}) \leq 2$ and the fourth one follows from the choice of the "τ" parameter in the recursive procedure. Hence, we get that

$$\mathbf{Var}\left[\sum_{i=1}^{t} \sum_{j=1}^{M(i)} a_{s,i,j}(\tilde{P}_{s,i,j} \cdot \tilde{Q}_{s,i,j} - P_{s,i,j} \cdot Q_{s,i,j})\right] \leq \frac{3\tau}{8}.$$

As a consequence, we get

$$\mathbf{Var}(p_{s,d} - \tilde{p}_{s,d}) \leq 2\,\mathbf{Var}(R_{s,\mathrm{neg}}) + 2\,\mathbf{Var}\left(\sum_{i=1}^{t}\sum_{j=1}^{M(i)} a_{s,i,j}(\tilde{P}_{s,i,j}\cdot\tilde{Q}_{s,i,j} - P_{s,i,j}\cdot Q_{s,i,j})\right) \leq \tau$$

.

For Condition (3), the multilinearity of each $\mathrm{Out}(p_{s,q})$ follows easily from the inductive hypothesis and from (39), using the fact that each $R_{s,\mathrm{reg}}$ is multilinear and that for each (s,i,j) triple the two multilinear polynomials $P_{s,i,j}$ and $Q_{s,i,j}$ are defined over disjoint sets of variables.

To finish establishing condition (3), we now bound the overall number of inner polynomials produced in these decompositions. We start by bounding the total number of inner polynomials that **MultiRegularize-Many-Wieners**$_{d,\beta}$ produces. Recall first that the number of polynomials that are input to the call to **MultiRegularize-Many-Wieners**$_{d-1,\beta_t^*}$ is $L_t = \kappa_t(k,\tau,d)$ (see 32). From the specification of the $\{\mathrm{In}(p_{s,q})_\ell\}$ polynomials given earlier, all but precisely k of these inner polynomials returned by **MultiRegularize-Many-Wieners**$_{d,\beta}$ are simply the inner polynomials that are returned from the call to **MultiRegularize-Many-Wieners**$_{d-1,\beta_t^*}$, and the other k polynomials are $R_{1,\mathrm{reg}},\ldots,R_{k,\mathrm{reg}}$. Since the **MultiRegularize-Many-Wieners**$_{d-1,\beta_t^*}$, procedure is called on L_t many polynomials, by the inductive hypothesis the total number of inner polynomials returned by this procedure call is some value

$$N_{\beta_t^*}\left(L_t, d-1, \frac{\tau}{16\cdot L_t\cdot L_t^{L_t}}\right) = O_{k,\tau,d}(1). \tag{41}$$

Consequently the total number of all inner polynomials $\{\mathrm{In}(p_{s,q})_\ell\}$ returned by **MultiRegularize-Many-Wieners**$_{d,\beta}$ is bounded by $N_{\beta_t^*}\left(L_t, d-1, \frac{\tau}{16\cdot L_t\cdot L_t^{L_t}}\right) + k$. Noting that $L_t \leq L_K$, we have that the total number of inner polynomials returned by **MultiRegularize-Many-Wieners**$_{d,\beta}$ is bounded by

$$N_{\beta_K^*}\left(L_K, d-1, \frac{\tau}{16\cdot L_K\cdot L_K^{L_K}}\right) + k,$$

which we define to be $N_\beta(k,d,\tau)$.

The next task is to upper bound $\sum_{s,q}\mathrm{Coeff}(p_{s,q})$. The main step is to bound the contribution from $q=d$. Recalling (40), we have

$$\tilde{p}_{s,d} = \sum_{i=1}^{t}\sum_{j=1}^{M(i)} a_{s,i,j}\cdot(\mathrm{Out}(P_{s,i,j})(\{\mathrm{In}(P_{s,i,j})_\ell\}))\cdot(\mathrm{Out}(Q_{s,i,j})(\{\mathrm{In}(Q_{s,i,j})_{\ell'}\})) + a_{s,\mathrm{reg}}\cdot R_{s,\mathrm{reg}}$$

and thus,

$$\mathrm{Coeff}(p_{s,d}) \leq 1 + \sum_{i=1}^{t}\sum_{j=1}^{M(i)} |a_{s,i,j}|\cdot\mathrm{Coeff}(P_{s,i,j})\cdot\mathrm{Coeff}(Q_{s,i,j}).$$

Note that $\sum_{s\in[k]}\sum_{i=1}^{t}\sum_{j=1}^{M(i)}|a_{s,i,j}| \leq L_t^{L_t}$ and $|a_{s,\mathrm{reg}}| \leq 1$. Further, by induction hypothesis,

$$\sum_{s\in[k]}\sum_{i=1}^{t}\sum_{j=1}^{M(i)}(\mathrm{Coeff}(P_{s,i,j}) + \mathrm{Coeff}(Q_{s,i,j})) + \sum_{s\in[k]}\sum_{q=1}^{d-1}\mathrm{Coeff}(p_{s,q}) \leq M_{\beta_t^*}\left(L_t, d-1, \frac{\tau}{16\cdot L_t\cdot L_t^{L_t}}\right).$$

Hence,

$$\sum_{s\in[k]}\sum_{q=1}^{d}\mathrm{Coeff}(\mathrm{Out}(p_{s,q})) \le k + L_t^{L_t}\cdot\left(M_{\beta_t^*}\left(L_t, d-1, \frac{\tau}{16\cdot L_t\cdot L_t^{L_t}}\right)\right)^2.$$

By definition,

$$L_t^{L_t}\cdot\left(M_{\beta_t^*}\left(L_t, d-1, \frac{\tau}{16\cdot L_t\cdot L_t^{L_t}}\right)\right)^2 \le L_K^{L_K}\cdot\left(M_{\beta_K^*}\left(L_K, d-1, \frac{\tau}{16\cdot L_K\cdot L_K^{L_K}}\right)\right)^2.$$

As $K = O(k/\epsilon)\log(1/\epsilon)$, we can define the quantity on the right hand side to be $M_\beta(k,d,\tau)$ which verifies condition (3).

Our last task is to bound the eigenregularity of the $\mathrm{In}(p_{s,q})_\ell$ polynomials. As noted above these polynomials are of two types: the k polynomials $R_{1,\mathrm{reg}},\ldots,R_{k,\mathrm{reg}}$ and the inner polynomials $\{\mathrm{In}(p_{s,q})_\ell^{(d-1)}\}$ that were returned from the call to **MultiRegularize-Many-Wieners**$_{d-1,\beta_t^*}$ on its L_t input polynomials.

We first tackle the $\{\mathrm{In}(p_{s,q})_\ell^{(d-1)}\}$ polynomials. Define Coeff_{rec} as

$$\mathrm{Coeff}_{rec} = \sum_{s\in[k]}\sum_{i=1}^{t}\sum_{j=1}^{M(i)}\left(\mathrm{Coeff}(P_{s,i,j}) + \mathrm{Coeff}(Q_{s,i,j})\right) + \sum_{s\in[k]}\sum_{q=1}^{d-1}\mathrm{Coeff}(p_{s,q}),$$

the sum of the absolute values of the coefficients of the outer polynomials returned by the recursive call. Likewise, define Num_{rec} as

$$\mathrm{Num}_{rec} = \left|\left\{\bigcup_{s\in[k],q\in[1,\ldots,d-1]}\mathrm{In}(p_{s,q}) \cup \bigcup_{i=1}^{t}\bigcup_{j=1}^{M(i)}\left(\{\mathrm{In}(P_{s,i,j})_{\ell'}\}_{\ell'=1,\ldots,\mathrm{num}(P_{s,i,j})} \cup \{\mathrm{In}(Q_{s,i,j})_{\ell''}\}_{\ell''=1,\ldots,\mathrm{num}(Q_{s,i,j})}\right)\right\}\right|,$$

the total number of inner polynomials returned by the recursive call. Then, note that $\mathrm{Coeff} \le \mathrm{Coeff}_{rec}^2\cdot L_t^{L_t} + k$ and $\mathrm{Num} = \mathrm{Num}_{rec} + k$. By inductive hypothesis, the polynomials $\{\mathrm{In}(p_{s,q})_\ell^{(d-1)}\}$ are $\beta_t^*(\mathrm{Num}_{rec} + \mathrm{Coeff}_{rec})$-eigenregular. However, note that

$$\begin{aligned}\beta_t^*(\mathrm{Num}_{rec}+\mathrm{Coeff}_{rec}) &= \beta((\mathrm{Num}_{rec}+\mathrm{Coeff}_{rec})^2\cdot k\cdot L_t^{L_t} + k + L_t^{L_t})\\ &\le \beta(\mathrm{Coeff}_{rec}^2\cdot L_t^{L_t} + k + \mathrm{Num}_{rec} + k) \le \beta(\mathrm{Num}+\mathrm{Coeff}).\end{aligned}$$

Here the inequality uses the fact that β_t^* is a non-increasing function. This shows that the polynomials $\{\mathrm{In}(p_{s,q})_\ell^{(d-1)}\}$ polynomials are $\beta(\mathrm{Num}+\mathrm{Coeff})$-eigenregular.

Next, note that $R_{s,\mathrm{reg}}$ is η_{t+1}-eigenregular. We have

$$\begin{aligned}\eta_{t+1} &= \beta\left(N_{\beta_t^*}\left(L_t, d-1, \frac{\tau}{16\cdot L_t\cdot L_t^{L_t}}\right)^2 + M_{\beta_t^*}\left(L_t, d-1, \frac{\tau}{16\cdot L_t\cdot L_t^{L_t}}\right)^2\cdot L_t^{L_t} + 2k + L_t^{L_t}\right)\\ &\le \beta\left(\mathrm{Num}_{rec}^2 + \mathrm{Coeff}_{rec}^2\cdot L_t^{L_t} + 2k + L_t^{L_t}\right) \le \beta\left(\mathrm{Num}+\mathrm{Coeff}\right).\end{aligned}$$

This implies that $R_{s,\mathrm{reg}}$ is $\beta(\mathrm{Num}+\mathrm{Coeff})$-eigenregular for all $s\in[k]$ verifying Condition (4). This concludes the proof of the inductive degree-d case, and thus concludes the proof of Theorem 43. □

5.4 Proof of Theorem 31

With Theorem 31 in hand it is straightforward to prove Theorem 43. As in the statement of Theorem 43, let us write $p(x)$ as $\sum_{q=0}^{d} c_q p_q(x)$ where $p_q \in \mathcal{W}^q$ for all q and $\mathbf{Var}[p_q] = 1$ for $1 \leq q \leq d$. Since $\mathbf{Var}[p] = 1$, by Fact 12 we have that $\sum_{q=1}^{d} c_q^2 = 1$. The procedure **Regularize-Poly**$_\beta$ calls **MultiRegularize-Many-Wieners**$_{d,\beta}$ on the $d+1$ polynomials $p_0, p_1, \ldots, p_d$ with the "τ" parameter of **MultiRegularize-Many-Wieners**$_{d,\beta}$ set to be $\frac{1}{d} \cdot (\tau/d)^{3d}$. This call to **MultiRegularize-Many-Wieners**$_{d,\beta}$ returns polynomials $\mathrm{Out}(p_0), \ldots, \mathrm{Out}(p_d)$ (these are the polynomials $h_0, \ldots, h_d$) and $\{\mathrm{In}(p_0)_\ell\}_{\ell=1,\ldots,m_0}, \ldots, \{\mathrm{In}(p_d)_\ell\}_{\ell=1,\ldots,m_d}$ (these are the polynomials $\{A_{q,\ell}\}_{q=0,\ldots,d,\ell=1,\ldots,m_q}$).

Condition (1) of Theorem 31 follows directly from Condition (1) of Theorem 43. Condition (2) of Theorem 43 implies that each $q \in \{0, \ldots, d\}$ has $\mathbf{Var}[p_q - \mathrm{Out}(p_q)(\{\mathrm{In}(p_q)\}_\ell)] \leq (1/d) \cdot (\tau/d)^{3d}$. Observing that $p_q - \mathrm{Out}(p_q)(\{\mathrm{In}(p_q)\}_\ell)$ lies in $\mathcal{W}_q$, using Fact 12, we have $\mathbf{Var}[p - \tilde{p}] \leq (\tau/d)^{3d}$. Using Lemma 5, we get that

$$\left|\mathbf{Pr}_{x \sim N(0,1)^n}[p(x) \geq 0] - \mathbf{Pr}_{x \sim N(0,1)^n}[\tilde{p}(x) \geq 0]\right| \leq O(\tau).$$

This concludes the verification of Condition (2). Conditions (3) and (4) in Theorem 31 follow from the respective conditions in Theorem 43.

6 Proof of Theorem 2: Deterministic approximate counting for degree-d PTFs over $N(0,1)^n$

In this section we combine the tools developed in the previous sections to prove Theorem 2. We do this in two main steps. First we use the CLT from Section 4 and the decomposition procedure from Section 5 to reduce the original problem (of ϵ-approximately counting satisfying assignments of a degree-d PTF under $N(0,1)^n$) to the problem of ϵ-approximating an expectation $\mathbf{E}_{G \sim N(0^r, \Sigma)}[\tilde{g}_c(G)]$, where $N(0^r, \Sigma)$ is a mean-0 r-dimensional Gaussian with covariance matrix Σ, and $\tilde{g}_c : \mathbb{R}^r \to [0,1]$ is a particular explicitly specified function. The key points here are that the value of r, the description length (bit complexity) of $\tilde{g}_c$, and the bit complexity of each entry of the covariance matrix Σ are all $O_{d,\epsilon}(1)$ (completely independent of n). Next, we describe how an $O_{d,\epsilon}(1)$-time deterministic algorithm can ϵ-approximate the desired expectation $\mathbf{E}_{G \sim N(0^r, \Sigma)}[\tilde{g}_c(G)]$.

Theorem 2 follows directly from Theorems 45 and 50, the main results of Sections 6.1 and 6.2 respectively.

Our approach is based on mollification (so that we can apply our CLT); we will need the following definitions from [DKN10]. We first define the bump function $b : \mathbb{R}^r \to \mathbb{R}$ as follows:

$$b(x) = \begin{cases} \sqrt{C_r}(1 - \|x\|_2^2) & \text{if } \|x\|_2 \leq 1 \\ 0 & \text{if } \|x\|_2 > 1, \end{cases}$$

where the constant C_r is chosen so that $\int_{x \in \mathbb{R}^r} b^2(x)dx = 1$. We let $\widehat{b}$ denote the Fourier transform of b, so

$$\widehat{b}(x) = \frac{1}{(2\pi)^{r/2}} \cdot \int_{y \in \mathbb{R}^r} b(y) \cdot \exp(-i\langle x, y\rangle)dy.$$

For $c > 0$ we define $B_c : \mathbb{R}^r \to \mathbb{R}$ as

$$B_c(x) = c^r \cdot \widehat{b}(c \cdot x)^2.$$

Using Parseval's identity, it is easy to see that $B_c(x)$ is non-negative and $\int_x B_c(x) = 1$, so B_c is a density function. We observe that $\|\widehat{b}\|_\infty$ is upper bounded by $O_r(1)$ and hence $\|B_c\|_\infty = O_{c,r}(1)$. Finally, for

$g : \mathbb{R}^r \to [0,1]$ and $c > 0$, $\tilde{g}_c : \mathbb{R}^r \to [0,1]$ is defined as

$$\tilde{g}_c(x) = \int_{y \in \mathbb{R}^r} g(x-y) \cdot B_c(y) dy. \tag{42}$$

6.1 Reducing to an $O_{d,\epsilon}(1)$-dimensional problem

In this subsection we prove the following. (The function $\mathrm{sign}_{0,1}(z)$ below outputs 1 if $z \geq 0$ and outputs 0 otherwise.)

Theorem 45. *There is an $O_{d,\epsilon}(1) \cdot \mathrm{poly}(n^d)$-time deterministic algorithm with the following performance guarantee: Given as input a degree-d real polynomial $p(x_1, \ldots, x_n)$ and a parameter $\epsilon > 0$, it outputs an integer r, a matrix of covariances $\Sigma \in \mathbb{R}^{r \times r}$ (whose diagonal entries are all 1), and a description of a function $\tilde{g}_c : \mathbb{R}^r \to [0,1]$, such that*

$$\left| \mathbf{Pr}_{x \sim N(0,1)^n}[p(x) \geq 0] - \mathbf{E}_{G \sim N(0^r, \Sigma)}[\tilde{g}_c(G)] \right| \leq O(\epsilon). \tag{43}$$

Moreover, $\tilde{g}_c$ is of the form given in (42), where $g = \mathrm{sign}_{0,1}(\phi)$ and $\phi : \mathbb{R}^r \to \mathbb{R}$ is a degree-d polynomial whose coefficients are rational numbers with numerator and denominator that are each integers of magnitude $O_{d,\epsilon}(1)$. Also, r is $O_{d,\epsilon}(1)$, and hence the description length of $\tilde{g}_c$ is $O_{d,\epsilon}(1)$ bits. Finally, each entry of Σ is a rational number whose numerator and denominator are both integers of magnitude $O_{d,\epsilon}(1)$.

The following lemma will be useful for us.

Lemma 46. *Let $a : \mathbb{R}^r \to [0,1]$ be a c-Lipschitz function. Let $\Sigma, \Sigma' \in \mathbb{R}^{r \times r}$ be two psd matrices such that $\|\Sigma - \Sigma'\|_2 \leq \delta$. Then $\left| \mathbf{E}_{G \sim N(0^r, \Sigma)}[a(G)] - \mathbf{E}_{G' \sim N(0^r, \Sigma')}[a(G')] \right| \leq cr(\delta + 3\sqrt{\delta \|\Sigma\|_2})$.*

Proof. Let $Z \sim N(0^r, \Sigma)$ and $Z' \sim N(0^r, \Sigma')$. It is shown in [DL82] that we have

$$d_{W,2}(Z, Z')^2 = \mathrm{Tr}(\Sigma + \Sigma' - 2(\Sigma^{1/2}\Sigma'\Sigma^{1/2})^{1/2}),$$

where $d_{W,2}(\cdot,\cdot)$ denotes the Wasserstein distance between two distributions in the ℓ_2 metric. Further, it is known [Bha00] that if A and B are psd matrices, then

$$\|A^{1/2} - B^{1/2}\|_2 \leq \sqrt{\|A - B\|_2}.$$

Observe that both $\Sigma + \Sigma'$ and $4\Sigma^{1/2}\Sigma'\Sigma^{1/2}$ are psd. As a consequence,

$$\|\Sigma + \Sigma' - 2(\Sigma^{1/2}\Sigma'\Sigma^{1/2})^{1/2}\|_2 \leq \sqrt{\|(\Sigma + \Sigma')^2 - 4\Sigma^{1/2}\Sigma'\Sigma^{1/2}\|_2}.$$

However, note that if we let $\Delta = \Sigma' - \Sigma$ so that $\|\Delta\|_2 \leq \delta$, then it is easy to see that

$$\sqrt{\|(\Sigma + \Sigma')^2 - 4\Sigma^{1/2}\Sigma'\Sigma^{1/2}\|_2} \leq \delta + 3\sqrt{\delta \|\Sigma\|_2}.$$

Hence we have $\mathrm{Tr}(\Sigma + \Sigma' - 2(\Sigma^{1/2}\Sigma'\Sigma^{1/2})^{1/2}) \leq r(\delta + 3\sqrt{\delta\|\Sigma\|_2})^2$, so $d_{W,2}(Z, Z') \leq \sqrt{r}(\delta + 3\sqrt{\delta\|\Sigma\|_2})$. Using Cauchy-Schwarz, we get

$$d_{W,1}(Z, Z') \leq r(\delta + 3\sqrt{\delta\|\Sigma\|_2}).$$

Using the fact that a is c-Lipschitz and recalling the coupling interpretation of the Wasserstein distance, we get the stated result. □

Now we turn to the proof of Theorem 45. We first recall that by Theorem 14 we may assume without loss of generality that the polynomial $p(x_1, \ldots, x_n)$ is multilinear. By rescaling we may further assume that $\mathbf{Var}[p] = 1$.

The algorithm begins by running the procedure **Decompose-Poly**$_\beta$ on p with its "τ" parameter set to ϵ and the β function set to be

$$\beta(x) = \Big(\frac{\epsilon}{C \cdot d \cdot x}\Big)^{Cd^2} \tag{44}$$

where C is a sufficiently large absolute constant (we discuss the choice of C below; see (53)). By Theorem 31, **Decompose-Poly** outputs a polynomial

$$\tilde{p}(x) = \sum_{j=0}^{d} h_j(A_{j,1}, \ldots, A_{j,m_j}(x))$$

that satisfies $|\mathbf{Pr}_{x\sim N(0,1)^n}[p(x) \geq 0] - \mathbf{Pr}_{x\sim N(0,1)^n}[\tilde{p}(x) \geq 0]| \leq \epsilon$. Furthermore, for each $j \in [1, \ldots, d]$ and $k \in [1, \ldots, m_j]$ we have that $\mathbf{Var}(A_{j,k}) = 1$, $A_{j,k} \in \mathcal{W}^q$ for some $q \in [1, j]$, $\sum_{j=1}^{d} m_j \leq r = r(d, \tau)$, and the sum of squared coefficients of h_j is bounded by $S = S(d, \tau)$. We moreover have that $\mathbf{Var}[\tilde{p}] \in [1/2, 3/2]$, that each h_j is a multilinear polynomial of degree at most d, and that each $h_j(A_{j,1}, \ldots, A_{j,m_j}(x))$ lies in the j-th Wiener chaos (and hence has expectation zero). Furthermore, if A_{j,i_1} and A_{j,i_2} appear in a monomial of h_j together, then they contain disjoint sets of variables.

Since each h_j has at most r^d coefficients, if we round each coefficient of each h_j to the nearest integer multiple of $\sqrt{(\epsilon/d)^{3d}/(dr^d)}$ and call the resulting polynomial $h_{new,j}$, and subsequently define

$$\tilde{p}_{new}(x) \stackrel{\text{def}}{=} \mathbf{E}[p] + \sum_{j=1}^{d} h_{new,j}(A_{j,1}(x), \ldots, A_{j,m_j}(x)),$$

then using Fact 44 we have that

$$\mathbf{Var}[\tilde{p}_{new}(x) - \tilde{p}(x)] \leq (dr^d) \cdot ((\epsilon/d)^{3d}/(dr^d)) = (\epsilon/d)^{3d}.$$

Since $\tilde{p}$ and $\tilde{p}_{new}$ have the same mean we may apply Lemma 5 and we get that $|\mathbf{Pr}_{x\sim N(0,1)^n}[\tilde{p}_{new}(x) \geq 0] - \mathbf{Pr}_{x\sim N(0,1)^n}[\tilde{p}(x) \geq 0]| \leq O(\epsilon)$. From now on, we will work with the polynomial $\tilde{p}_{new}$.

At a high level, the plan is as follows: $\mathbf{Pr}_{x\sim N(0,1)^n}[\tilde{p}_{new}(x) \geq 0]$ is equal to

$$\mathbf{E}_{x\sim N(0,1)^n}\left[\mathrm{sign}_{0,1}\left(\mathbf{E}[\tilde{p}] + \sum_{j=1}^{d} h_{new,j}(A_{j,1}(x), \ldots, A_{j,m_j}(x))\right)\right]. \tag{45}$$

Since the $\mathrm{sign}_{0,1}$ function is discontinuous, we cannot apply our CLT from Section 4 to (45) directly (recall that the CLT only applies to functions with bounded second derivatives). To get around this we define a "mollified" version (a version that is continuous and has bounded second derivative) of the relevant function and show that the expectation (45) is close to the corresponding expectation of the mollified function. This allows us to apply the CLT to the mollified function. A final step is that we will round the covariance matrix Σ' obtained from the CLT to Σ so that all its entries have bounded bit complexity; Lemma 46 will ensure that we can do this and only incur an acceptable increase in the error.

We now enter into the details. Let $\phi : \mathbb{R}^r \to \mathbb{R}$ be defined as

$$\phi(x_{1,1}, \ldots, x_{1,m_1}, \ldots, x_{d,1}, \ldots, x_{d,m_d}) = \mathbf{E}[p] + \sum_{j=1}^{d} h_{new,j}(x_{j,1}, \ldots, x_{j,m_j}),$$

and let $g : \mathbb{R}^r \to \{0,1\}$ be defined as $g(x) = \mathrm{sign}_{0,1}(\phi(x))$. In the subsequent discussion, for any function $F : \mathbb{R}^r \to \mathbb{R}$, we write $F^{(k)} : \mathbb{R}^r \to \mathbb{R}^m$ (where $m = \binom{r+k-1}{k}$) to denote the function whose coordinates are all k-th order partial derivatives of F, so $\|F^{(k)}\|_\infty$ denotes the supremum of all values achieved by any k-th order partial derivative at any input in $\mathbb{R}^r$. Intuitively, our goal is to construct a mollification $\tilde{g} : \mathbb{R}^r \to \mathbb{R}$ for g such that $\|\tilde{g}^{(2)}\|_\infty < \infty$ and $\tilde{g}$ is a "good approximation" to g at most points in $\mathbb{R}^r$. There are many different mollification constructions that could potentially be used; we shall use the following theorem from [DKN10]:

Theorem 47. *[DKN10] For any region $R \subseteq \mathbb{R}^r$ and any $c > 0$, the mollification $\widetilde{I}_{R,c} : \mathbb{R}^r \to [0,1]$ of the $\{0,1\}$-valued function $I_R(x) \stackrel{def}{=} \mathbf{1}_{x \in R}$ has the property that for every point $x \in \mathbb{R}^r$,*

$$\left| I_R(x) - \widetilde{I}_{R,c}(x) \right| \leq \min\left\{1, O\left(\frac{r^2}{c^2 \cdot \mathrm{dist}(x, \partial R)^2}\right)\right\},$$

where $\mathrm{dist}(x, \partial R)$ denotes the Euclidean distance between x and the boundary of R. Moreover, this mollification satisfies $\|\widetilde{I}^{(1)}_{R,c}\|_\infty \leq 2c$ and $\|\widetilde{I}^{(2)}_{R,c}\|_\infty \leq 4c^2$.

Applying Theorem 47 to the region $R \stackrel{\text{def}}{=} \{x \in \mathbb{R}^r : g(x) = 1\}$, we get that for any $c > 0$ there is a mollified function $\tilde{g}_c : \mathbb{R}^r \to [0,1]$ such that $\|\tilde{g}_c^{(2)}\|_\infty \leq 4c^2$, $\|\tilde{g}_c^{(1)}\|_\infty \leq 2c$, and

$$|g(x) - \tilde{g}_c(x)| \leq \min\left\{1, O\left(\frac{r^2}{c^2 \cdot \mathrm{dist}(x, \partial R)^2}\right)\right\}. \tag{46}$$

The following lemma ensures that for a suitable choice of c, the mollification $\tilde{g}_c$ is indeed a useful proxy for g for our purposes:

Lemma 48. *For c as specified in (51) below, we have that the function $\tilde{g}_c : \mathbb{R}^r \to [0,1]$ described above satisfies*

$$\mathbf{E}_{x \sim N(0,1)^n}\left[|g(A_{1,1}(x), \ldots, A_{d,m_d}(x)) - \tilde{g}_c(A_{1,1}(x), \ldots, A_{d,m_d}(x))|\right] \leq O(\epsilon). \tag{47}$$

Proof. We will use the following claim:

Claim 49. *Let $x, y \in \mathbb{R}^r$ and $\|x\|_\infty \leq B$. If $\|x - y\|_2 \leq \delta \leq B$, then $|\phi(x) - \phi(y)| \leq d(2B)^d r^{d/2} \cdot \sqrt{S} \cdot \delta$.*

Proof. Recall that $\phi(x)$ is a multilinear degree-d polynomial in r variables for which the sum of squares of coefficients is at most S. Let us write ϕ as

$$\phi(x) = \sum_{\mathcal{A} \in \binom{[r]}{\leq d}} c_{\mathcal{A}} x_{\mathcal{A}}$$

where $x_{\mathcal{A}}$ represents the monomial $\prod_{i \in \mathcal{A}} x_i$ corresponding to the set $\mathcal{A}$, so we have $\sum_{\mathcal{A}} (c_{\mathcal{A}})^2 \leq S$. For any fixed $\mathcal{A} \in \binom{[r]}{\leq d}$, a simple application of the triangle inequality across the (at most) d elements of $\mathcal{A}$ gives that

$$|x_{\mathcal{A}} - y_{\mathcal{A}}| \leq d \cdot (2B)^d \cdot \delta.$$

Since the number of monomials in $\phi(x)$ is at most r^d, using Cauchy-Schwarz we get that

$$|\phi(x) - \phi(y)| = \left|\sum_{\mathcal{A}} c_{\mathcal{A}}(x_{\mathcal{A}} - y_{\mathcal{A}})\right| \leq \sqrt{\sum_{\mathcal{A}} (c_{\mathcal{A}})^2} \cdot \sqrt{\sum_{\mathcal{A}} (x_{\mathcal{A}} - y_{\mathcal{A}})^2} \leq \sqrt{S} \cdot r^{d/2} \cdot d \cdot (2B)^d \cdot \delta,$$

the claimed bound. □

By Claim 49, we have that if $x \in \mathbb{R}^r$ has $\|x\|_\infty \le B$ and $|\phi(x)| > d(2B)^d r^{d/2} \cdot \sqrt{S} \cdot \delta$, where $\delta \le B$, then $\|\mathrm{dist}(x, \partial R)\|_2 > \delta$. By (46), if $\mathrm{dist}(x, \partial R) > \delta$ then $|g(x) - \tilde{g}_c(x)| \le O(\frac{r^2}{c^2\delta^2})$. Hence provided that we take $\delta \le B$, we may upper bound $\mathbf{E}_{x\sim N(0,1)^n}\left[|g(A_{1,1}(x), \ldots, A_{d,m_d}(x)) - \tilde{g}_c(A_{1,1}(x), \ldots, A_{d,m_d}(x))|\right]$ (the LHS of (47) by

$$\mathbf{Pr}_{x\sim N(0,1)^n}\left[\max_{i\in[1,\ldots,d], j\in[1,\ldots,m(i)]} |A_{i,j}(x)| > B\right]$$
$$+ \mathbf{Pr}_{x\sim N(0,1)^n}[|\phi(A_{1,1}(x), \ldots, A_{d,m_d}(x))| \le d(2B)^d r^{d/2} \cdot \sqrt{S} \cdot \delta] + O\left(\frac{r^2}{c^2\delta^2}\right). \quad (48)$$

To bound the second summand above, we recall that $\phi(A_{1,1}(x), \ldots, A_{d,m_d}(x))$ is a multilinear degree-d polynomial whose variance is at least $1/2$. By the anti-concentration bound Theorem 4, we get that

$$\mathbf{Pr}_{x\sim N(0,1)^n}[|\phi(A_{1,1}(x), \ldots, A_{d,m_d}(x))| \le d \cdot (2B)^d r^{d/2} \cdot \sqrt{S} \cdot \delta] \le O\left(d \cdot B \cdot \sqrt{r} \cdot S^{1/2d} \cdot \delta^{1/d}\right).$$

To bound the first summand, we observe that since each $A_{i,j}$ is a mean-0 variance-1 degree-d polynomial, by the degree-d Chernoff bound (Theorem 3) and a union bound over the r polynomials $A_{i,j}$, for any $B > e^d$ we have that

$$\mathbf{Pr}_{x\sim N(0,1)^n}\left[\max_{i\in[1,\ldots,d], j\in[1,\ldots,m(i)]} |A_{i,j}(x)| > B\right] \le r \cdot d \cdot e^{-\Omega(B^{2/d})}. \quad (49)$$

Thus we get that (48) is at most

$$r \cdot d \cdot e^{-\Omega(B^{2/d})} + O\left(d \cdot B \cdot \sqrt{r} \cdot S^{1/2d} \cdot \delta^{1/d}\right) + O\left(\frac{r^2}{c^2\delta^2}\right). \quad (50)$$

Choosing

$$B = \left(\Omega(1) \cdot \ln \frac{rd}{\epsilon}\right)^{d/2}, \quad \delta = \left(\frac{\epsilon}{d \cdot B \cdot \sqrt{r} \cdot S^{1/2d}}\right)^d, \quad \text{and} \quad c = \frac{r}{\delta\sqrt{\epsilon}} \quad (51)$$

(note that these choices satisfy the requirements that $B \ge e^d$ and $\delta \le B$), we get that each of the three summands constituting (50) is $O(\epsilon)$, and Lemma 48 is proved. □

Using Condition (4) of Theorem 31, we have that each $A_{j,k}$ is η-eigenregular where $\eta \le \beta(r+S)$. Now since $\|\tilde{g}_c^{(2)}\|_\infty \le 4c^2$ and each $A_{j,k}(x)$ is a mean-0, degree-d Gaussian polynomial with $\mathbf{Var}[A_{j,k}(x)] = 1$, we may apply our CLT, Theorem 19, and we get that

$$\left|\mathbf{E}_{x\sim N(0,1)^n}[\tilde{g}_c(A_{1,1}(x), \ldots, A_{d,m_d}(x))] - \mathbf{E}_{G'\sim N(0^r,\Sigma')}[\tilde{g}_c(G')]\right| \le 2^{O(d\log d)} \cdot r^2 \cdot \sqrt{\beta(r+s)} \cdot 4c^2, \quad (52)$$

where $\Sigma' \in \mathbb{R}^{r\times r}$ is the covariance matrix corresponding to the $A_{j,k}$'s (note that the variance bound on each $A_{j,k}$ ensures that the diagonal entries are indeed all 1 as claimed). It is easy to see that there exists a choice of C in our definition of the function β (see (44)) which has the property

$$\beta(r+s) \le \frac{\epsilon^2}{2^{O(d\log d)} r^4 c^4}. \quad (53)$$

As a result, the triangle inequality applied to (52) and Lemma 48 gives that

$$\left|\mathbf{Pr}_{x\sim N(0,1)^n}[\tilde{p}(x) \ge 0] - \mathbf{E}_{G'\sim N(0^r,\Sigma')}[\tilde{g}_c(G')]\right| \le O(\epsilon).$$

We are almost done; it remains only to pass from Σ' to Σ, which we do using Lemma 46. By Theorem 47 we have that $\|\tilde{g}_c^{(1)}\|_\infty \le 2c$ and hence $\tilde{g}_c$ is $2c$-Lipschitz. Observing that each entry of Σ' is in $[-1, 1]$, we have that $\|\Sigma'\|_2 = O_{d,\epsilon}(1)$. Hence by taking $\Sigma \in \mathbb{R}^{r\times r}$ to be a psd matrix that is sufficiently close to Σ' with respect to $\|\cdot\|_2$, we get that Σ has all its coefficients rational numbers with numerator and denominator of magnitude $O_{d,\epsilon}(1)$, and from Lemma 46 we get that $|\mathbf{E}_{G\sim N(0^r,\Sigma)}[\tilde{g}_c(G)] - \mathbf{E}_{G'\sim N(0^r,\Sigma')}[\tilde{g}_c(G')] \le O(\epsilon)$. Thus Theorem 45 is proved. □

6.2 Solving the $O_{d,\epsilon}(1)$-dimensional problem in $O_{d,\epsilon}(1)$ time

The last step is to establish the following:

Theorem 50. *There is a deterministic $O_{d,\epsilon}(1)$-time algorithm which, given as input the output $r, \Sigma, \tilde{g}_c$ of Theorem 45 and the value of $\epsilon > 0$, outputs a value ν such that*

$$\left|\nu - \mathbf{E}_{(G_1,\ldots,G_r)\sim N(0^r,\Sigma)}[\tilde{g}_c(G_1,\ldots,G_r)]\right| \leq \epsilon.$$

We only sketch the proof since the details of the argument are tedious and we are content with an $O_{d,\epsilon}(1)$ time bound. The first observation is that since each entry of the covariance matrix Σ is at most 1 in magnitude and $\tilde{g}_c$ is everywhere bounded in $[0,1]$, it suffices to estimate the expected value conditioned on $(G_1,\ldots,G_r) \sim N(0^r,\Sigma)$ lying in an origin-centered cube $[-Z,Z]^r$ for some $Z = O_{d,\epsilon}(1)$. Given this, it is possible to simulate this conditional normal distribution with a discrete probability distribution X supported on a finite set of points in $[-Z,Z]^r$. Recall that Σ has entries as specified in Theorem 45 (rational numbers of magnitude $O_{d,\epsilon}(1)$). Since $\tilde{g}_c$ is $2c$-Lipschitz as noted earlier, by using a sufficiently fine grid of $O_{d,\epsilon}(1)$ points in $[-Z,Z]^r$ and (deterministically) estimating the probability that $N(0^r,\Sigma)$ assigns to each grid point to a sufficiently small $1/O_{d,\epsilon}(1)$ additive error, Theorem 50 reduces to the following claim:

Claim 51. *Given any point $x \in [-Z,Z]^r$ and any accuracy parameter $\xi > 0$, the function $\tilde{g}_c(x)$ can be computed to within additive accuracy $\pm\xi$ in time $O_{d,\epsilon,\xi}(1)$.*

Proof. We only sketch the proof of Claim 51 here as the argument is routine and the details are tedious. We first note that (as can be easily verified from the description of B_c given earlier) there is an (easily computed) value $W = O_{d,\epsilon}(1)$ such that if $S = \{x \in \mathbb{R}^r : \|x\|_\infty > W\}$, then we have $\int_{z\in S} B_c(z) \leq \xi/2$. As a consequence, since g is everywhere bounded in $[0,1]$, we get that

$$\left|\int_{y\in\mathbb{R}^r} g(x-y)\cdot B_c(y)dy - \int_{y\in(\mathbb{R}^r\setminus S)} g(x-y)\cdot B_c(y)dy\right| \leq \xi/2, \tag{54}$$

and hence it suffices to estimate

$$\int_{y\in[-W,W]^r} g(x-y)\cdot B_c(y)dy \tag{55}$$

to within an additive $\pm\xi/2$. Now observe that $\|B_c\|_\infty = O_{d,\epsilon}(1)$ and $\|B_c^{(1)}\|_\infty = O_{d,\epsilon}(1)$. It follows that given any $y \in [-W,W]^r$ and any accuracy parameter $\rho > 0$, we can compute $B_c(y)$ to additive accuracy $\pm\rho$ in time $O_{d,\epsilon,\rho}(1)$. Recalling that $g(x) = \mathrm{sign}_{0,1}(\phi(x))$ where $\phi(x)$ is a degree-d polynomial whose coefficients are $O_{d,\epsilon}(1)$-size rational numbers, it follows that by taking a sufficiently fine grid of $O_{d,\epsilon,\xi,W}(1)$ points in $[-W,W]^r$, we can use such a grid to estimate (55) to an additive $\pm\xi/2$ in $O_{d,\epsilon,\xi,W}(1)$ time as desired. □

7 Deterministic approximate counting for degree-d polynomials over $\{-1,1\}^n$

In this section we use Theorem 2 to prove Theorem 1. Since the arguments here are identical to those used in [DDS13a] (where an algorithm for deterministic approximate counting of degree-2 PTF satisfying assignments over $N(0,1)^n$ is used to obtain an algorithm for satisfying assignments over $\{-1,1\}^n$), we only sketch the argument here.

We recall the "regularity lemma for PTFs" of [DSTW10]. This lemma says that every degree-d PTF $\mathrm{sign}(p(x))$ over $\{-1,1\}^n$ can be expressed as a shallow decision tree with variables at the internal nodes and degree-d PTFs at the leaves, such that a random path in the decision tree is quite likely to reach a leaf that has a "close-to-regular" PTF. As explained in [DDS13a], the [DSTW10] proof actually provides an

efficient deterministic procedure for constructing such a decision tree given p as input, and thus we have the following lemma (see Theorem 36 of [DDS13a] for a detailed explanation of how Theorem 52 follows from the results of [DSTW10]):

Theorem 52. *Let $p(x_1, \ldots, x_n)$ be a multilinear degree-d PTF. Fix any $\tau > 0$. There is an algorithm $A_{\mathrm{Construct-Tree}}$ which, on input p and a parameter $\tau > 0$, runs in $\mathrm{poly}(n, 2^{\mathrm{depth}(d,\tau)})$ time and outputs a decision tree $\mathcal{T}$ of depth*

$$\mathrm{depth}(d, \tau) := \frac{1}{\tau} \cdot \left(d \log \frac{1}{\tau} \right)^{O(d)},$$

where each internal node of the tree is labeled with a variable and each leaf ρ of the tree is labeled with a pair $(p_\rho, \mathrm{label}(\rho))$ where $\mathrm{label}(\rho) \in \{+1, -1, \text{"fail"}, \text{"regular"}\}$. The tree $\mathcal{T}$ has the following properties:

1. *Every input $x \in \{-1,1\}^n$ to the tree reaches a leaf ρ such that $p(x) = p_\rho(x)$;*
2. *If leaf ρ has $\mathrm{label}(\rho) \in \{+1, -1\}$ then $\mathbf{Pr}_{x \in \{-1,1\}^n}[\mathrm{sign}(p_\rho(x)) \neq \mathrm{label}(\rho)] \leq \tau$;*
3. *If leaf ρ has $\mathrm{label}(\rho) =$ "regular" then p_ρ is τ-regular; and*
4. *With probability at most τ, a random path from the root reaches a leaf ρ such that $\mathrm{label}(\rho) =$ "fail".*

Proof of Theorem 1: The algorithm for approximating $\mathbf{Pr}_{x \in \{-1,1\}^n}[p(x) \geq 0]$ to $\pm\epsilon$ works as follows. It first runs $A_{\mathrm{Construct-Tree}}$ with its "τ" parameter set to $\Theta((\epsilon/d)^{4d+1})$ to construct the decision tree $\mathcal{T}$. It then iterates over all leaves ρ of the tree. For each leaf ρ at depth d_ρ that has $\mathrm{label}(\rho) = +1$ it adds 2^{-d_ρ} to v (which is initially zero), and for each leaf ρ at depth d_ρ that has $\mathrm{label}(\rho) =$ "regular" it runs the algorithm of Theorem 2 on p_ρ (with its "ϵ" parameter set to $\Theta((\epsilon/d)^{4d+1})$) to obtain a value $v_\rho \in [0,1]$ and adds $v_\rho \cdot 2^{-d_\rho}$ to v. It outputs the value $v \in [0,1]$ thus obtained.

Theorems 52 and 2 imply that the running time is as claimed. To establish correctness of the algorithm we will use the "invariance principle" of [MOO10] (see Theorem 2.1):

Theorem 53 ([MOO10])**.** *Let $p(x) = \sum_{S \subseteq [n], |S| \leq d} p_S x_S$ be a degree-d multilinear polynomial over $\{-1,1\}^n$ with $\mathbf{Var}[p] = 1$. Suppose each coordinate $i \in [n]$ has $\mathbf{Inf}_i(p) \leq \tau$. Then*

$$\sup_{t \in \mathbb{R}} |\, \mathbf{Pr}_x[p(x) \leq t] - \mathbf{Pr}_{\mathcal{G} \sim N(0,1)^n}[p(\mathcal{G}) \leq t]| \leq O(d\tau^{1/(4d+1)}).$$

By Theorem 52, the leaves of $\mathcal{T}$ that are marked $+1$, -1 or "fail" collectively contribute at most $\Theta((\epsilon/d)^{4d+1}) \leq \epsilon/2$ to the error of the output value v. Theorem 53 implies that each leaf ρ at depth d_ρ that is marked "regular" contributes at most $2^{-d_\rho} \cdot \epsilon/2$ to the error, so the total contribution from all such leaves is at most $\epsilon/2$. This concludes the proof of Theorem 1. □

8 Application of Theorem 1: A fixed-parameter deterministic multiplicative approximation algorithm for absolute moments

Consider the following computational problem, which we call ABSOLUTE-MOMENT: Given a degree-d polynomial $p(x_1, \ldots, x_n)$ and an integer parameter $k \geq 1$, compute the value $\mathbf{E}_{x \in \{-1,1\}^n}[|p(x)|^k]$ of the *k-th absolute moment* of p. It is clear that the *raw moment* $\mathbf{E}[p(x)^k]$ can be computed in roughly n^k time by expanding out the polynomial $p(x)^k$, performing multilinear reduction, and outputting the constant term. Since the k-th raw moment equals the k-th absolute moment for even k, this gives an n^k time algorithm for ABSOLUTE-MOMENT for even k. However, as shown in [DDS13a], even for $d = 2$ the ABSOLUTE-MOMENT problem is #P-hard for any odd $k \geq 1$, and thus it is natural to seek approximation algorithms.

Using the hypercontractive inequality [Bon70, Bec75] it is not difficult to show that the obvious randomized algorithm (draw uniform points from $\{-1,1\}^n$ and use them to empirically estimate $\mathbf{E}_{x\in\{-1,1\}^n}[|p(x)|^k]$) with high probability gives a $(1\pm\epsilon)$-accurate estimate of the k-th absolute moment of p in in $\mathrm{poly}(n^d, 2^{dk\log k}, 1/\epsilon)$ time. In this section we observe that Theorem 1 yields a *deterministic* fixed-parameter-tractable $(1\pm\epsilon)$-multiplicative approximation algorithm for ABSOLUTE-MOMENT:

Theorem 54. *There is a deterministic algorithm which, given any degree-d polynomial $p(x_1,\dots,x_n)$ over $\{-1,1\}^n$, any integer $k\geq 1$, and any $\epsilon>0$, runs in $O_{d,k,\epsilon}(1)\cdot\mathrm{poly}(n^d)$ time and outputs a value v that multiplicatively $(1\pm\epsilon)$-approximates the k-th absolute moment:*

$$v\in\left[(1-\epsilon)\,\mathbf{E}_{x\in\{-1,1\}^n}[|p(x)|^k],(1+\epsilon)\,\mathbf{E}_{x\in\{-1,1\}^n}[|p(x)|^k]\right].$$

Theorem 54 is a generaliation of the $d=2$ special case which was proved in [DDS13a] (using the deterministic approximate counting result for degree-2 PTFs which is the main result of that paper). The proof is closely analogous to the degree-2 case so we only sketch it below; see [DDS13a] for a detailed argument.

The first step is the following easy observation:

Observation 55. *Let $p(x)$ be a degree-d polynomial over $\{-1,1\}^n$ that has $\mathbf{E}_{x\in\{-1,1\}^n}[p(x)^2]=1$. Then for all $k\geq 1$ we have that the k-th absolute moment $\mathbf{E}_{x\in\{-1,1\}^n}[|p(x)|^k]$ is at least c_d where $c_d>0$ is some universal constant (depending only on d).*

Given an input degree-d polynomial $p(x_1,\dots,x_n)$, we may divide by $\|p\|_2$ to obtain a scaled version $q=p/\|p\|_2$ which has $\|q\|_2=1$. Observation 55 implies that an additive $\pm\epsilon$-approximation to $\mathbf{E}[|q(x)|^k]$ is also a multiplicative $(1\pm O_d(\epsilon))$-approximation to $\mathbf{E}[|q(x)|^k]$. Multiplying the approximation by $\|p\|_2^k$ we obtain a multiplicative $(1\pm O_d(\epsilon))$-approximation to $\mathbf{E}[|p(x)|^k]$. Thus to prove Theorem 54 it suffices to give a deterministic algorithm which finds an additive $\pm\epsilon$-approximation to $\mathbf{E}[|q(x)|^k]$ for degree-d polynomials with $\|q\|_2=1$. This follows from Theorem 56 below:

Theorem 56. *Let $q(x)$ be an input degree-d polynomial over $\{-1,1\}^n$ with $\mathbf{E}[q(x)^2]=1$. There is an algorithm A_{moment} that, on input $k\in\mathbb{Z}^+$, q, and $\epsilon>0$, runs in time $O_{k,d,\epsilon}(1)\cdot\mathrm{poly}(n^d)$ and outputs a value $\tilde{\mu}_k$ such that*

$$\left|\tilde{\mu}_k-\mathbf{E}_{x\in\{-1,1\}}[|q(x)|^k]\right|\leq\epsilon.$$

The idea behind the proof of Theorem 56 is simple. By Theorem 1, we can estimate $\mathbf{Pr}_{x\sim\{-1,1\}^n}[q(x)\geq t]$ to high accuracy for any t of our choosing. Doing this repeatedly for different choices of t, we can get a detailed picture of where the probability mass of the random variable $q(x)$ lies (for x uniform over $\{-1,1\}^n$), and with this detailed picture it is straightforward to estimate the k-th moment.

In a bit more detail, let $\gamma_q(t)$ denote the probability mass function of $q(x)$ when x is distributed uniformly over $\{-1,1\}^n$. We may write the k-th absolute moment of q as

$$\mathbf{E}_{x\in\{-1,1\}^n}[|q(x)|^k]=\int_{-\infty}^{\infty}|t|^k\gamma_q(t)dt. \tag{56}$$

Using standard tail bounds on polynomials over $\{-1,1\}^n$, for a suitable choice of $M=O_{k,d,\epsilon}(1)$ we have that

$$\mathbf{E}_{x\in\{-1,1\}^n}[|q(x)|^k]\in\left[\int_{-M}^{M}|t|^k\gamma_q(t)dt,\int_{-M}^{M}|t|^k\gamma_q(t)dt+\epsilon/4\right],$$

and hence to approximate $\mathbf{E}_{x\sim N(0,1)^n}[|q(x)|^k]$ to an additive $\pm\epsilon$, it suffices to approximate $\int_{-M}^{M}|t|^k\gamma_q(t)dt$ to an additive $\pm 3\epsilon/4$.

We may write $\int_{-M}^{M} |t|^k \gamma_q(t)dt$ as

$$\sum_{j=1-M/\Delta}^{M/\Delta} \int_{(j-1)\Delta}^{j\Delta} |t|^k \gamma_q(t)dt$$

(here Δ should be viewed as a small positive value). When $|j|$ is small the summand $\int_{(j-1)\Delta}^{j\Delta} |t|^k \gamma_q(t)dt$ may be well-approximated by zero, and when $|j|$ is not small the summand $\int_{(j-1)\Delta}^{j\Delta} |t|^k \gamma_q(t)dt$ may be well-approximated by $|j\Delta|^k q_{j,\Delta}$, where

$$q_{j,\Delta} \stackrel{\text{def}}{=} \mathbf{Pr}_{x\in\{-1,1\}^n}[q(x) \in [(j-1)\Delta, j\Delta).$$

Using Theorem 1 twice we may approximate $q_{j,\Delta}$ to high accuracy. With a suitable choice of $\Delta = O_{k,d,\epsilon}(1)$ and the cutoff for j being "small," it is possible to approximate $\int_{-M}^{M} |t|^k \gamma_q(t)dt$ to an additive $\pm 3\epsilon/4$, and thus obtain Theorem 56, following this approach. We leave the detailed setting of parameters to the interested reader.

Acknowledgement. We thank Ilias Diakonikolas for his contributions in the early stages of this project. We also thank Rafal Latala, Michel Ledoux, Elchanan Mossel, Ivan Nourdin and Krzysztof Oleszkiewicz for answering questions about the CLT. Part of this work was done when A.D. was hosted by Oded Regev and the Simons Institute. A.D. would like to thank them for their kind hospitality and support.

References

[ABI85] N. Alon, L. Babai, and A. Itai. A fast and simple randomized algorithm for the maximal independent set problem. *J. of Algorithms*, 7:567–583, 1985.

[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.

[Bec75] W. Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, 102:159–182, 1975.

[Bha00] Rajendra Bhatia. *Matrix Analysis*. Springer Verlag, Basel, 2000.

[Bon70] A. Bonami. Etude des coefficients Fourier des fonctiones de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)*, 20(2):335–402, 1970.

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

[CS13] D. Cartwright and B. Sturmfels. The number of eigenvalues of a tensor. *Linear algebra and its applications*, 432(2):942–952, 2013.

[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.

[DDS13a] A. De, I. Diakonikolas, and R. Servedio. Deterministic approximate counting for degree-2 polynomial threshold functions. manuscript, 2013.

[DDS13b] A. De, I. Diakonikolas, and R. Servedio. Deterministic approximate counting for juntas of degree-2 polynomial threshold functions. manuscript, 2013.

[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.

[DL82] D.C Dowson and B.V Landau. The frechet distance between multivariate normal distributions. *Journal of Multivariate Analysis*, 12(3):450 – 455, 1982.

[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.

[FW95] J. Friedman and A. Wigderson. On the Second Eigenvalue of Hypergraphs. *Combinatorica*, 15(1):43–65, 1995.

[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.

[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.

[Kan11a] Daniel M. Kane. k-independent gaussians fool polynomial threshold functions. In *IEEE Conference on Computational Complexity*, pages 252–261, 2011.

[Kan11b] Daniel M. Kane. A small prg for polynomial threshold functions of gaussians. In *FOCS*, pages 257–266, 2011.

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

[Kan12b] Daniel M. Kane. A pseudorandom generator for polynomial threshold functions of gaussian with subpolynomial seed length. *CoRR*, abs/1210.1280, 2012.

[Kan12c] 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.

[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.

[Lat06] R. Latala. Estimates of moments and tails of gaussian chaoses. *Annals of Probability*, 34(6):2315–2331, 2006.

[Lat13] R. Latala. Personal communication, 2013.

[Led13] M. Ledoux. Personal communication, 2013.

[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.

[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.

[MZ09] R. Meka and D. Zuckerman. Pseudorandom Generators for Polynomial Threshold Functions. Available at http://arxiv.org/abs/0910.4122, 2009.

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

[NN93] J. Naor and M. Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM J. on Comput.*, 22(4):838–856, 1993. Earlier version in STOC'90.

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

[Nou13] I. Nourdin. Personal communication, 2013.

[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.

[NPR10] I. Nourdin, G. Peccati, and A. Réveillac. Multivariate normal approximation using Stein's method and Malliavin calculus. *Ann. Inst. H. Poincaré Probab. Statist.*, 46(1):45–58, 2010.

[Ole13] K. Oleszkiewicz. Personal communication, 2013.

[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.

[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