Dimension Reduction for Polynomials over Gaussian Space and Applications

Badih Ghazi[1]
Google Research, 1600 Amphitheatre Parkway Mountain View, CA 94043, USA
badihghazi@gmail.com

Pritish Kamath[2]
Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA
pritish@mit.edu

Prasad Raghavendra[3]
University of California Berkeley, Berkeley, CA, USA
raghavendra@berkeley.edu

Abstract

We introduce a new technique for reducing the dimension of the ambient space of low-degree polynomials in the Gaussian space while preserving their relative correlation structure. As an application, we obtain an explicit upper bound on the dimension of an ε-optimal noise-stable Gaussian partition. In fact, we address the more general problem of upper bounding the number of samples needed to ε-approximate any joint distribution that can be *non-interactively simulated* from a correlated Gaussian source. Our results significantly improve (from Ackermann-like to "merely" exponential) the upper bounds recently proved on the above problems by De, Mossel & Neeman [CCC 2017, SODA 2018 resp.] and imply decidability of the larger alphabet case of the *gap non-interactive simulation* problem posed by Ghazi, Kamath & Sudan [FOCS 2016].

Our technique of dimension reduction for low-degree polynomials is simple and can be seen as a generalization of the Johnson-Lindenstrauss lemma and could be of independent interest.

2012 ACM Subject Classification Theory of computation → Complexity theory and logic

Keywords and phrases Dimension reduction, Low-degree Polynomials, Noise Stability, Non-Interactive Simulation

Digital Object Identifier 10.4230/LIPIcs.CCC.2018.28

Acknowledgements The authors are extremely grateful to Madhu Sudan for helpful guidance throughout all the stages of this project. The authors would also like to thank Anindya De, Elchanan Mossel and Joe Neeman for clarifying explanations of their papers and helpful discussions. PK would like to thank Irit Dinur, Mika Göös, Raghu Meka and Li-Yang Tan for helpful discussions.

1 Introduction

1.1 Gaussian Isoperimetry & Noise Stability

Isoperimetric problems over the Gaussian space have become central in various areas of theoretical computer science such as hardness of approximation and learning theory. In

[1] The work was done while the author was a student at MIT. Supported in parts by NSF CCF-1650733 and CCF-1420692.
[2] Supported in parts by NSF CCF-1420956, CCF-1420692, CCF-1218547 and CCF-1650733.
[3] Research supported by Okawa Research Grant and NSF CCF-1408643.

33rd Computational Complexity Conference (CCC 2018).
Editor: Rocco A. Servedio; Article No. 28; pp. 28:1–28:37
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

its simplest and classic form, the central question in isoperimetry is to determine what is the smallest possible surface area for a body of a given volume. Alternately, isoperimetric problems can also be formulated in terms of the notion of *Noise stability*.

Fix a real number $\rho \in [0,1]$ and let $f : \mathbb{R}^n \to \{0,1\}$ denote the indicator function of a subset (say $\mathcal{A}_f$) of the n-dimensional Gaussian space ($\mathbb{R}^n$ with the Gaussian measure γ_n given by the density function $d\gamma_n/d\boldsymbol{X} = \exp(-\|\boldsymbol{X}\|_2^2/2)/(2\pi)^{n/2}$). The noise stability $\mathsf{Stab}_\rho(f)$ is the probability that two ρ-correlated Gaussians $\boldsymbol{X}, \boldsymbol{Y}$ both fall inside or outside $\mathcal{A}_f$. More generally, the Gaussian "noise operator" U_ρ (also known as the Ornstein-Uhlenbeck operator), defined for each $\rho \in [0,1]$, acts on any $f : \mathbb{R}^n \to [0,1]$ as

$$(U_\rho f)(X) \;:=\; \mathop{\mathbb{E}}_{\boldsymbol{Z}\sim\gamma_n}\left[f\left(\rho X + \sqrt{1-\rho^2}\cdot \boldsymbol{Z}\right)\right] .$$

The noise stability is then defined as

$$\mathop{\mathsf{Stab}}_{\rho}(f) := \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\gamma_n}[f(\boldsymbol{X})\cdot U_\rho f(\boldsymbol{X}) + (1-f(\boldsymbol{X}))\cdot(1-U_\rho f(\boldsymbol{X}))]$$

Reformulated in terms of noise stability, the isoperimetric problem is to determine the largest possible value of $\mathsf{Stab}_\rho(f)$ for a function $f : \mathbb{R}^n \to [0,1]$ with a given expectation $\mathbb{E}[f] = \alpha$. The seminal isoperimetric theorem of Borell [10] shows that indicator functions of halfspaces are the most noise-stable among all functions $f : \mathbb{R}^n \to [0,1]$ with a given expectation over the Gaussian measure. Borell's theorem (along with the invariance principle [44, 42]) has had fundamental applications in theoretical computer science, e.g., in the hardness of approximation for Max-Cut under the Unique Games conjecture [39] and in voting theory [42].

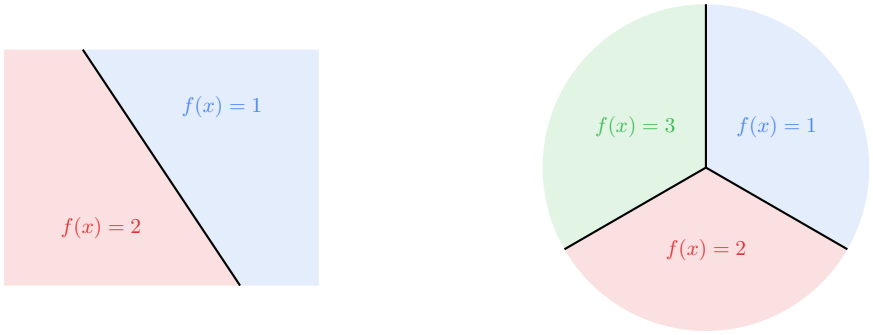
In this work, we are interested in analogues of Borell's theorem for partitions of the Gaussian space *into more than two subsets*, or equivalently noise stability of functions f taking values over $[k] := \{1,\ldots,k\}$. Towards stating these analogues, let's state Borell's theorem formally in a more general notation. Let Δ_k be the probability simplex in $\mathbb{R}^k$ (i.e., convex hull of the basis vectors $\{e_1,\ldots,e_k\}$). The Ornstein-Uhlenbeck operator naturally extends to vector valued functions $f : \mathbb{R}^n \to \mathbb{R}^k$ as $U_\rho f := (U_\rho f_1,\ldots,U_\rho f_k)$ where $f = (f_1,\ldots,f_k)$. The noise stability of functions $f : \mathbb{R}^n \to \Delta_k$ is now defined as $\mathsf{Stab}_\rho(f) := \mathbb{E}_{\boldsymbol{X}\sim\gamma_n}[\langle f(\boldsymbol{X}), U_\rho f(\boldsymbol{X})\rangle]$ where $\langle\cdot,\cdot\rangle$ denotes the inner product over $\mathbb{R}^k$. We can similarly define the noise stability of a function $f : \mathbb{R}^n \to [k]$ by embedding $[k]$ in Δ_k, i.e., identifying coordinate $i \in [k]$ with the standard basis vector $e_i \in \Delta_k$. Borell's theorem can be formally stated in this notation as follows:

Borell's Theorem [10]. *For any $f : \mathbb{R}^n \to \Delta_2$, consider the halfspace function $h = (h_1, h_2) : \mathbb{R}^n \to \Delta_2$ given by $h_1(\boldsymbol{X}) = \mathbf{1}\{\langle a, \boldsymbol{X}\rangle \geq b\}$ and $h_2(\boldsymbol{X}) = 1 - h_1(\boldsymbol{X})$, for suitable $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ such that $\mathbb{E}[f] = \mathbb{E}[h]$. Then, $\mathsf{Stab}_\rho(f) \leq \mathsf{Stab}_\rho(h)$.*

While Borell's theorem deals with the case of $k = 2$, it is natural to consider the question of maximal noise stability for $k > 2$, stated as follows.

Question 1. [Maximum Noise Stability (MNS)] Given a positive integer $k \geq 2$ and $\alpha \in \Delta_k$, what is the maximum noise stability of a function $f : \mathbb{R}^n \to \Delta_k$ satisfying $\mathbb{E}[f] = \alpha$?

Question 1 remains open even for $k = 3$. In the particular case where $\alpha = (\frac{1}{k},\ldots,\frac{1}{k})$, the *Standard Simplex Conjecture* posits that the maximum noise stability is achieved by a "standard simplex partition" (this is equivalent to the *Plurality is Stablest* conjecture) [39, 35]. Even in the special case when $k = 3$ and $\alpha = (\frac{1}{3},\frac{1}{3},\frac{1}{3})$, the answer is still tantalizingly open.



(a) (b)

Figure 1 (a) Borell's Theorem: Halfspaces are most noise stable (b) Standard Simplex Partition for $k = 3$ conjectured to be most noise stable (also known as the "Peace Sign Conjecture").

In fact, a suprising result of [32] shows that when the α_i's are not all equal, the standard simplex partition (and any variant thereof) *does not* achieve the maximum noise stability. This indicates that the case $k \geq 3$ is fundamentally different than the case of $k = 2$. On the positive side, if we consider $0 < \rho < \rho_0(k, n)$ (for some $\rho_0(k, n)$ that goes to 0 for large n), then the Standard Simplex Conjecture has been shown to hold [31]. However, this result is not applicable in the setting where ρ is fixed and $n \to \infty$.

The fact that we do not understand optimal partitions for $k \geq 3$, led De, Mossel & Neeman [19] to ask whether the optimal partition is realized in any finite dimension. More formally:

Question 2. Given $k \geq 2$, $\rho \in (0, 1)$, and $\alpha \in \Delta_k$, let $S_n(\alpha)$ be the optimal noise stability of a function $f : \mathbb{R}^n \to \Delta_k$ subject to $\mathbb{E}[f] = \alpha$. Is there an n_0 such that $S_n(\alpha) = S_{n_0}(\alpha)$ for all $n \geq n_0$?

Even Question 2 is open as of now! In this light, De, Mossel & Neeman [19] ask whether one can obtain an *explicitly computable* $n_0 = n_0(k, \rho, \varepsilon)$ such that $S_{n_0}(\alpha) \geq S_n(\alpha) - \varepsilon$ for all $n \in \mathbb{N}$ (in other words, there exists a function $f : \mathbb{R}^{n_0} \to \Delta_k$ that comes ε-close to achieving the optimal noise stability). Note that the challenge is really about n_0 being *explicit*, since some $n_0(k, \rho, \varepsilon)$ always exists, as $S_n(\alpha)$ is a converging sequence as $n \to \infty$.

A natural approach to proving such an explicit bound is the idea of *dimension reduction*. Basically, it suffices to obtain an $n_0 = n_0(k, \rho, \varepsilon)$ such that for any n and any given function $f : \mathbb{R}^n \to \Delta_k$, there exists a function $\widetilde{f} : \mathbb{R}^{n_0} \to \Delta_k$ with $\mathbb{E}[\widetilde{f}] = \mathbb{E}[f]$ and $\mathsf{Stab}_\rho(\widetilde{f}) \geq \mathsf{Stab}_\rho(f) - \varepsilon$. Instantiating f with an optimal (or near-optimal) partition in $\mathbb{R}^n$, for arbitrarily large n, then gives an ε-optimal partition $\widetilde{f}$ in $\mathbb{R}^{n_0}$.

Indeed, De, Mossel and Neeman follow such an approach and obtain an *explicitly computable* bound on n_0. To do so, they use and build on the theory of *eigenregular polynomials* that were previously studied in [21], which in turn uses other tools such as Malliavin calculus.

In this work, we introduce fundamentally different, but more elementary techniques (elaborated on shortly), thereby significantly improving the bound in [19]. In particular, we show the following.

▶ **Theorem 1** (Dimension Bound on Approximately Optimal Noise-Stable Function). *Given parameters $k \geq 2$, $\rho \in [0, 1]$ and $\varepsilon > 0$, there exists an explicitly computable $n_0 = n_0(k, \rho, \varepsilon)$*

such that the following holds:
For any $n \in \mathbb{N}$ and $f : \mathbb{R}^n \to \Delta_k$, there exists $\widetilde{f} : \mathbb{R}^{n_0} \to \Delta_k$ such that,
1. $\| \mathbb{E}[f] - \mathbb{E}[\widetilde{f}] \|_1 \leq \varepsilon$.
2. $\mathsf{Stab}_\rho(\widetilde{f}) \geq \mathsf{Stab}_\rho(f) - \varepsilon$.
In particular, the explicit choice of n_0 can be upper bounded by $\exp\left(\mathrm{poly}\left(k, \frac{1}{1-\rho}, \frac{1}{\varepsilon}\right)\right)$.

Remarks

(i) In contrast to our theorem, the bound on n_0 in [19] has an Ackermann-type growth.
(ii) It is a slight technicality that we get $\| \mathbb{E}[f] - \mathbb{E}[\widetilde{f}] \|_1 \leq \varepsilon$ instead of $\mathbb{E}[f] = \mathbb{E}[\widetilde{f}]$ as was required. However, it is possible to slightly modify $\widetilde{f}$ to make $\mathbb{E}[f] = \mathbb{E}[\widetilde{f}]$, if we allow n_0 to depend on $\alpha = \mathbb{E}[f]$ (which is the case in Question 2).

Theorem 1 has an immediate application to showing that approximately most-stable voting schemes (among all low-influential voting schemes) can be computed efficiently. We refer the reader to [19] for the details of this application. In order to prove Theorem 1, we in fact turn to the more general setting of *non-interactive simulation*.

1.2 Non-Interactive Simulation from Correlated Gaussian Sources

Consider a more general setting where instead of a single function f, we have two players, Alice and Bob, with corresponding functions $A : \mathbb{R}^n \to \Delta_k$ and $B : \mathbb{R}^n \to \Delta_k$. They apply A and B on the sequence of random variables $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ and $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$ respectively, where $(\boldsymbol{X}, \boldsymbol{Y}) \sim \mathcal{G}_\rho^{\otimes n}$, which is the distribution of ρ-correlated Gaussians in n dimensions, i.e. each coordinate $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$ is independently sampled from $\mathcal{G}_\rho := \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right)$. The goal is to choose A and B such that $\mathbb{E}[A] = \mathbb{E}[B] = \alpha$, which is a pre-specified vector in Δ_k, while maximizing $\mathbb{E}_{(\boldsymbol{X},\boldsymbol{Y})}[\langle A(\boldsymbol{X}), B(\boldsymbol{Y})\rangle]$. Note that, this quantity is same as $\mathbb{E}_{\boldsymbol{X}\sim\gamma_n}[\langle A(\boldsymbol{X}), U_\rho B(\boldsymbol{X})\rangle]$, and hence in the restricted setting of $A = B = f$ this quantity is exactly the noise stability of f.

We can interpret the above as: Alice observes$\boldsymbol{X}$ and outputs $i \in [k]$ with probability $A_i(\boldsymbol{X})$, similarly Bob observes $\boldsymbol{Y}$ and outputs $j \in [k]$ with probability $B_j(\boldsymbol{Y})$. In this sense, Alice and Bob wish to maximize their "agreement probability", i.e., their probability of outputting the same symbol. The dimension reduction mentioned in Theorem 1 generalized to this setup would require obtaining an $n_0(k, \rho, \varepsilon)$ and a dimension reduction of A and B that approximately preserves the marginals and does not decrease the agreement probability by more than ε.

However, in this language, it is more natural to ask for a much stronger dimension reduction that preserves the entire joint distribution of symbols that Alice and Bob output, up to ε in total variation distance. We denote the joint distribution of Alice and Bob's outputs as $(A(\boldsymbol{X}), B(\boldsymbol{Y}))_{\mathcal{G}_\rho^{\otimes n}}$, which is the distribution over $(i, j) \in [k] \times [k]$ given as $\Pr[$*Alice outputs i and Bob outputs j*$] = \mathbb{E}_{(\boldsymbol{X},\boldsymbol{Y})}[A_i(\boldsymbol{X})B_j(\boldsymbol{Y})]$. In the case of $k = 2$, such a dimension reduction follows from (a more general version of) Borell's theorem with in fact $n_0 = 1$! Our main result is indeed such a dimension reduction for all $k \geq 2$.

▶ **Theorem 2** (NIS from correlated Gaussian source)**.** *Given parameters $k \geq 2$, $\rho \in (0, 1)$ and $\varepsilon > 0$, there exists an explicitly computable $n_0 = n_0(k, \rho, \varepsilon)$ such that the following holds:*
For any n and $A : \mathbb{R}^n \to \Delta_k$ and $B : \mathbb{R}^n \to \Delta_k$, there exist $\widetilde{A} : \mathbb{R}^{n_0} \to \Delta_k$ and $\widetilde{B} : \mathbb{R}^{n_0} \to \Delta_k$ such that,

$$d_{\mathrm{TV}}\left((A(\boldsymbol{X}), B(\boldsymbol{Y}))_{\mathcal{G}_\rho^{\otimes n}},\ (\widetilde{A}(\boldsymbol{a}), \widetilde{B}(\boldsymbol{b}))_{\mathcal{G}_\rho^{\otimes n_0}}\right) \leq \varepsilon\,.$$

In particular, the explicit choice of n_0 *is upper bounded as* $\exp\left(\mathrm{poly}\left(k, \frac{1}{1-\rho}, \frac{1}{\varepsilon}\right)\right)$. *The transformation satisfies a stronger property that there exists an "oblivious" randomized transformation (with a shared random seed) to go from* A *to* $\widetilde{A}$ *and from* B *to* $\widetilde{B}$*, which works with probability at least* $1-\varepsilon$*. Since the same transformation is applied on* A *and* B *with the same random seed, if* $A = B$*, then* $\widetilde{A} = \widetilde{B}$ *as well.*

Theorem 1 follows immediately from Theorem 2, by simply setting $A = B = f$ to obtain $\widetilde{f} = \widetilde{A} = \widetilde{B}$. In fact, following up on [19], De, Mossel & Neeman were able to extend their techniques to prove Theorem 2 [20] (again with Ackerman-type bounds on n_0). To do so, they build on the tools developed in [19] along with a new smoothing argument inspired by boosting procedures in learning theory and potential function arguments in complexity theory and additive combinatorics. As we shall present shortly, our approach gets directly to Theorem 2 in a much more elementary fashion.

1.3 Extension: Non-Interactive Simulation from General Discrete Sources

The *Non-Interactive Simulation of Joint Distributions* is quite well studied in Information Theory and more recently in Theoretical Computer Science. Two players, Alice and Bob, observe the sequences of random variables $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ and $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ respectively, where each pair $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ is independently drawn from a known *source* distribution μ. The fundamental question here is to understand which other *target* joint distributions ν can Alice and Bob simulate, without communicating with each other? How many samples from μ are needed to do so, or in other words, what is the *simulation rate*?

The history of this problem goes back to the classical works of Gács and Körner [24] and Wyner [56]. Specifically, consider the distribution Eq over $\{0,1\} \times \{0,1\}$ where both marginals are $\mathrm{Ber}(1/2)$ and the bits are identical with probability 1. Gács and Körner studied the special case of this problem corresponding to the target distribution $\nu = \mathsf{Eq}$. They characterized the simulation rate in this case, showing that it is equal to what is now known as the *Gács-Körner common information* of μ. On the other hand, Wyner studied the special case corresponding to the source distribution $\mu = \mathsf{Eq}$. He characterized the simulation rate in this case, showing that it is equal to what is now known as *Wyner common information* of ν. Another particularly important work was by Witsenhausen [54] who studied the case where the target distribution $\nu = \mathcal{G}_\rho$. In this case, he showed that the largest correlation ρ that can be simulated is exactly the well-known "maximal correlation coeffcient"[4] $\rho(\mu)$ which was first introduced by Hirschfeld [33] and Gebelein [25] and then studied by Rényi [52]. Witsenhausen also considered the case where the target distribution $\nu = \mathrm{DSBS}_\rho$ is a "doubly symmetric binary source", which is a pair of ρ-correlated bits (i.e., a pair of ± 1 random variables with correlation ρ), and gave an approach to simulate correlated bits by first simulating $\mathcal{G}_\rho$ starting with samples from μ, and then applying half-space functions to get outputs in $\{\pm 1\}$. Starting with μ, such a approach simulates $\mathrm{DSBS}_{\rho'}$ where $\rho' = 1 - \frac{2\arccos\rho(\mu)}{\pi}$. Indeed, this calculation is identical to one that arises in the rounding technique employed in Goemans-Williamson's approximation algorithm [30] for MaxCut 20 years later!

We will consider the modern formulation of the NIS question as defined in [37]. This formulation ignores the simulation rate, and only focuses on whether simulation is even possible or not, given infinitely many samples from μ – that is, whether the simulation rate is non-zero or not.

[4] We skip this definition as it is not central to our paper. The definition can be found in e.g. [29].

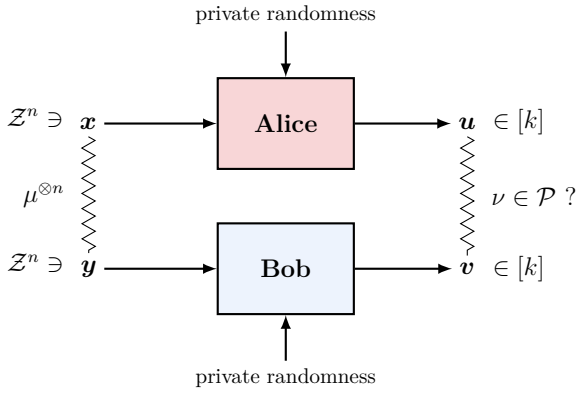


Figure 2 Non-Interactive Simulation, e.g., as studied in [37]

▶ **Definition 3** (Non-interactive Simulation of Joint Distributions [37])**.** Let $(\mathcal{Z} \times \mathcal{Z}, \mu)$ and $([k] \times [k], \nu)$ be two joint probability spaces. The distribution ν can be *non-interactively simulated* from distribution μ if there exists a sequence of functions $\left\{A^{(n)} : \mathcal{Z}^n \to \Delta_k\right\}_{n\in\mathbb{N}}$ and $\left\{B^{(n)} : \mathcal{Z}^n \to \Delta_k\right\}_{n\in\mathbb{N}}$ such that the joint distribution $\nu_n = (A^{(n)}(\boldsymbol{x}), B^{(n)}(\boldsymbol{y}))_{\mu^{\otimes n}}$ over $[k] \times [k]$ is such that $\lim_{n\to\infty} d_{\mathrm{TV}}(\nu_n, \nu) = 0$.

A central question that was left open following the work of Witsenhausen is: given distributions μ and ν, can ν be non-interactively simulated from μ? Can this be even decided algorithmically? Even when μ and ν are extremely simple, e.g., μ is uniform on the triples $\{(0,0),(0,1),(1,0)\}$ and ν is the doubly symmetric binary souce $\mathrm{DSBS}_{0.49}$, it is open if μ can simulate ν! This problem was formalized as a natural gap-problem in a work by a subset of the authors along with Sudan [29]. Here we state a slightly more general version.

▶ **Problem 4** (Gap-NIS$((\mathcal{Z} \times \mathcal{Z}, \mu), \mathcal{P}, k, \varepsilon)$, cf. [29])**.** *Given a joint probability space $(\mathcal{Z} \times \mathcal{Z}, \mu)$, a family of joint probability spaces $\mathcal{P}$ supported over $[k] \times [k]$, and an error parameter $\varepsilon > 0$, distinguish between the following cases:*

- **(i)** *there exists n and $A : \mathcal{Z}^n \to \Delta_k$ and $B : \mathcal{Z}^n \to \Delta_k$, s.t. the distribution $\nu' = (A(\boldsymbol{x}), B(\boldsymbol{y}))_{\mu^{\otimes n}}$ satisfies $d_{\mathrm{TV}}(\nu', \nu) \le \varepsilon$ for some $\nu \in \mathcal{P}$.*
- **(ii)** *for all n and all $A : \mathcal{Z}^n \to \Delta_k$ and $B : \mathcal{Z}^n \to \Delta_k$, the distribution $\nu' = (A(\boldsymbol{x}), B(\boldsymbol{y}))_{\mu^{\otimes n}}$ satisfies $d_{\mathrm{TV}}(\nu', \nu) > 2\varepsilon$ for all $\nu \in \mathcal{P}$.*[5]

In prior work [29], it was shown that Gap-NIS for discrete distributions μ and ν is decidable, in the special case where $k = 2$. This was done by introducing a framework, which reduced the problem to understanding Gap-NIS for the special case where $\mu = \mathcal{G}_\rho$. Indeed, the reason why the case of $k = 2$ was easier was precisely because Borell's theorem [10] gives an exact characterization of the distributions over $[2] \times [2]$ that can be simulated from $\mathcal{G}_\rho$. The lack of understanding of the distributions over $[k] \times [k]$ that can be simulated from $\mathcal{G}_\rho$ was suggested in [29] as a barrier for extending their result to $k > 2$. With Theorem 2 in hand, it is possible to extend the framework in [29] of using a Regularity Lemma and Invariance principle, to yield the following theorem (as also done in [20], but with Ackerman-type bounds).

[5] the choice of constant 2 is arbitrary. Indeed, we could replace it by any constant greater than 1.

▶ **Theorem 5** (NIS from Discrete Sources). *Let $(\mathcal{Z} \times \mathcal{Z}, \mu)$ be a joint probability space. Given parameters $k \geq 2$ and $\varepsilon > 0$, there exists an explicitly computable $n_0 = n_0(k, \mu, \varepsilon)$ such that the following holds:*
For any n and $A : \mathcal{Z}^n \to \Delta_k$ and $B : \mathcal{Z}^n \to \Delta_k$, there exist $\widetilde{A} : \mathcal{Z}^{n_0} \to \Delta_k$ and $\widetilde{B} : \mathcal{Z}^{n_0} \to \Delta_k$ such that,

$$d_{\mathrm{TV}}\left((A(\boldsymbol{x}), B(\boldsymbol{y}))_{\mu^{\otimes n}},\ (\widetilde{A}(\boldsymbol{a}), \widetilde{B}(\boldsymbol{b}))_{\mu^{\otimes n_0}}\right) \leq \varepsilon\ .$$

In particular, the explicit choice of n_0 is upper bounded as $\exp\left(\mathrm{poly}\left(k, \frac{1}{\varepsilon}, \frac{1}{1-\rho}, \log\left(\frac{1}{\alpha}\right)\right)\right)$, where $\alpha = \alpha(\mu)$ is the smallest atom in μ and $\rho = \rho(\mu)$ is the maximal correlation coefficient of μ.

The above theorem immediately suggests a brute force algorithm to decide Gap-NIS$((\mathcal{Z} \times \mathcal{Z}, \mu), \mathcal{P}, k, \varepsilon)$. We do not provide details of the proof of the above theorem in this extended abstract. The interested reader is referred to the full version of this paper [27] (available online) for details.

1.4 Dimension Reduction for Polynomials over Gaussian Space

We now describe the main technique of "*dimension reduction for low-degree polynomials*" that we introduce in this work, which could be of independent interest. We highlight that this technique is the main contribution of this paper.

Let's start with Theorem 2, and explain the main ideas behind its proof. We are given two vector-valued functions $A : \mathbb{R}^n \to \Delta_k$ and $B : \mathbb{R}^n \to \Delta_k$. We wish to reduce the dimension n of the Gaussian space on which A and B act while preserving the joint distribution $(A(\boldsymbol{X}), B(\boldsymbol{Y}))_{\mathcal{G}_\rho^{\otimes n}}$ over $[k] \times [k]$. Recall that $\mathbb{E}_{(\boldsymbol{X},\boldsymbol{Y}) \sim \mathcal{G}_\rho^{\otimes n}}[A_i(\boldsymbol{X}) \cdot B_j(\boldsymbol{Y})]$ is the probability of the event [*Alice outputs i and Bob outputs j*]. For succinctness, we write this expectation as $\langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes n}}$. In order to approximately preserve the joint distribution $(A(\boldsymbol{X}), B(\boldsymbol{Y}))_{\mathcal{G}_\rho^{\otimes n}}$, it suffices to approximately preserve $\langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes n}}$ for each $(i, j) \in [k] \times [k]$ upto an additive ε/k^2. Thus, to prove Theorem 2, we wish to find an explicit constant $n_0 = n_0(\rho, k, \varepsilon)$, along with functions $\widetilde{A} : \mathbb{R}^{n_0} \to \Delta_k$ and $\widetilde{B} : \mathbb{R}^{n_0} \to \Delta_k$ such that

$$\left| \left\langle \widetilde{A}_i, \widetilde{B}_j \right\rangle_{\mathcal{G}_\rho^{\otimes n_0}} - \langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes n}} \right| \ \leq\ \frac{\varepsilon}{k^2}\ .$$

Achieving this directly is highly unclear, since a priori, we have no structural information about A and B! To get around this, we show that it is possible to first apply a structural transformation on A and B to convert them to low-degree and multilinear polynomials (see subsection 2.2 for formal definitions). Such transformations are described in section 4. This however creates a new problem that the transformed A and B no longer map to Δ_k. Nevertheless, we will show that after the said transformation, we still have that the outputs of A and B are close to Δ_k in expected ℓ_2^2 distance, that is, $\mathrm{dist}(A, \Delta_k) := (\mathbb{E}_{\boldsymbol{X}} \|\mathcal{R}(A(\boldsymbol{X})) - A(\boldsymbol{X})\|_2^2)^{1/2}$ is small (where $\mathcal{R} : \mathbb{R}^k \to \Delta_k$ denotes the *rounding operator* that maps any $v \in \mathbb{R}^k$ to its closest point in Δ_k). This will ensure that rounding the outputs of A and B to Δ_k will approximately preserve the correlations $\langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes n}}$.

We are now able to revise our objective as follows: Given two (vector-valued) degree-d polynomials $A : \mathbb{R}^n \to \mathbb{R}^k$ and $B : \mathbb{R}^n \to \mathbb{R}^k$, does there exist an explicit function $n_0 = n_0(k, d, \delta)$, along with polynomials $\widetilde{A} : \mathbb{R}^{n_0} \to \mathbb{R}^k$ and $\widetilde{B} : \mathbb{R}^{n_0} \to \mathbb{R}^k$ that δ-approximately preserve (i) the correlation $\langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes n}}$ for all $(i, j) \in [k] \times [k]$ and (ii) closeness of the outputs of A and B to Δ_k, that is, $\mathrm{dist}(A, \Delta_k)$ and $\mathrm{dist}(B, \Delta_k)$?

We introduce a very simple and natural dimension-reduction procedure for low-degree multilinear polynomials over Gaussian space. Specifically, for $\boldsymbol{M}$ that is a randomly sampled $n \times n_0$ matrix with i.i.d. standard Gaussian entries, we set

$$\widetilde{A}(a) := A\left(\frac{\boldsymbol{M}a}{\|a\|_2}\right) \qquad \text{and} \qquad \widetilde{B}(b) := B\left(\frac{\boldsymbol{M}b}{\|b\|_2}\right) \qquad \text{for } a, b \in \mathbb{R}^{n_0} \setminus \{0\} \ . \tag{1}$$

We leave $\widetilde{A}$ and $\widetilde{B}$ undefined on $0 \in \mathbb{R}^{n_0}$. This is inconsequential as $\{0\}$ is a measure zero set under γ_n. Our main dimension-reduction theorem for polynomials is stated as follows.

▶ **Theorem 6** (Dimension Reduction Over Gaussian Space)**.** *Given parameters $k \geq 2$, $d \in \mathbb{Z}_{\geq 0}$, $\rho \in (0,1)$ and $\delta > 0$, there exists an explicitly computable $n_0 = n_0(d, k, \delta)$ such that the following holds:*

Let $A : \mathbb{R}^n \to \mathbb{R}^k$ and $B : \mathbb{R}^n \to \mathbb{R}^k$ be degree-d multilinear polynomials. Additionally, suppose that $\operatorname{dist}(A, \Delta_k), \operatorname{dist}(B, \Delta_k) \leq \delta$. Consider the functions $\widetilde{A} : \mathbb{R}^{n_0} \to \mathbb{R}^k$ and $\widetilde{B} : \mathbb{R}^{n_0} \to \mathbb{R}^k$ as defined in Equation 1. With probability at least $1 - O(\delta)$ over the choice of $\boldsymbol{M} \sim \gamma_1^{\otimes(n \times n_0)}$, the following holds:

1. *For every $i, j \in [k]$: $\left| \langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes n}} - \left\langle \widetilde{A}_i, \widetilde{B}_j \right\rangle_{\mathcal{G}_\rho^{\otimes n_0}} \right| \leq \delta$.*
2. *$\operatorname{dist}(\widetilde{A}, \Delta_k) \leq \sqrt{\delta}$ and $\operatorname{dist}(\widetilde{B}, \Delta_k) \leq \sqrt{\delta}$.*

In particular, the explicit choice of n_0 is upper bounded as $\exp\left(\operatorname{poly}\left(d, \log k, \log(\frac{1}{\delta})\right)\right)$.

It is clear from the construction of $\widetilde{A}$ and $\widetilde{B}$ that this theorem is giving us an "oblivious" randomized transformation, as also remarked in Theorem 2. The proof of Theorem 6 is obtained by combining Theorem 8 and Proposition 9 in section 3.

Proof outline and analogy with the Johnson-Lindenstrauss lemma.

We will now highlight a few parallels between our proof of Theorem 6 and the proof of the Johnson-Lindenstrauss (JL) lemma (cf. [36, 18]), which has been extremely influential in computer science with numerous applications including compressed sensing, manifold learning, unsupervised learning and graph embedding.

Suppose that we have two unit vectors $u, v \in \mathbb{R}^n$. We wish to obtain a randomized transformation $\Psi_{\boldsymbol{s}} : \mathbb{R}^n \to \mathbb{R}^{n_0}$ (for some random seed $\boldsymbol{s}$) that approximately preserves the inner product, that is, $\langle \Psi_{\boldsymbol{s}}(u), \Psi_{\boldsymbol{s}}(v) \rangle \approx_\delta \langle u, v \rangle$ holds with probability $1 - \delta$, over the randomness of seed $\boldsymbol{s}$; note that here $\langle \cdot, \cdot \rangle$ denotes the inner product over $\mathbb{R}^n$ and $\mathbb{R}^{n_0}$ respectively. Indeed, there is such a transformation, namely, $\Psi_{\boldsymbol{M}}(u) = \frac{\boldsymbol{M} \cdot u}{\sqrt{n_0}}$ where $\boldsymbol{M} \sim \gamma_1^{\otimes n_0 \times n}$. Let $F(\boldsymbol{M}) = \langle \Psi_{\boldsymbol{M}}(u), \Psi_{\boldsymbol{M}}(v) \rangle$. Such a transformation satisfies,

$$\mathop{\mathbb{E}}_{\boldsymbol{M}}[F(\boldsymbol{M})] = \langle u, v \rangle \qquad \text{and} \qquad \mathop{\mathsf{Var}}_{\boldsymbol{M}}\left(F(\boldsymbol{M})\right) \ = \ \frac{\langle u, v \rangle^2 + \|u\|_2^2 \|v\|_2^2}{n_0} \ \leq \ \frac{2}{n_0},$$

where we use that u and v are unit vectors. Thus, if we choose $n_0 = 2/\delta^3$, then we can make the variance smaller than δ^3. Thereby, using Chebyshev's inequality, we get that with probability at least $1 - \delta$, the inner product $\langle u, v \rangle$ is preserved, that is, $|\langle \Psi_{\boldsymbol{M}}(u), \Psi_{\boldsymbol{M}}(v) \rangle - \langle u, v \rangle| \leq \delta$. Thus, we have a *oblivious* randomized dimension reduction that reduces the dimension of any pair of unit vectors to $O(1/\delta^3)$, independent of n. Note that, instead of using Chebyshev's inequality, we could use a much sharper concentration bound to show that $n_0 = O(1/\varepsilon^2 \log(1/\delta))$ suffices to preserve the inner product up to an additive ε, with probability $1 - \delta$. However, we described the Chebyshev's inequality version as this is similar to our proof of Theorem 6.

The problem we are facing, although morally similar, is technically entirely different. For simplicity, let's first consider the task of reducing the dimension of the domain of a *single* pair of polynomials $A : \mathbb{R}^n \to \mathbb{R}$ and $B : \mathbb{R}^n \to \mathbb{R}$. And for the moment, consider the transformation such that $\Psi_{\boldsymbol{M}} A : \mathbb{R}^{n_0} \to \mathbb{R}$ is given by $A(\boldsymbol{M}a/\sqrt{n_0})$. Similarly, $\Psi_{\boldsymbol{M}} B(b) = B(\boldsymbol{M}b/\sqrt{n_0})$. Our proof of Theorem 6 proceeds along similar lines as the above proof of the JL Lemma, that is, by considering $F(\boldsymbol{M}) = \langle \Psi_{\boldsymbol{M}} A, \Psi_{\boldsymbol{M}} B \rangle_{\mathcal{G}_\rho^{\otimes n_0}}$, and proving bounds on $\mathbb{E}_{\boldsymbol{M}}[F(\boldsymbol{M})]$ and $\mathsf{Var}(F(\boldsymbol{M}))$. This turns out to be quite delicate! Unlike the JL case, we don't even have $\mathbb{E}_{\boldsymbol{M}}[F(\boldsymbol{M})] = \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes n}}$. What we do show however is that,

$$\left| \mathbb{E}_{\boldsymbol{M}}[F(\boldsymbol{M})] \; - \; \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes n}} \right| \; \leq \; o_{n_0}(1) \quad \text{and} \quad \operatorname*{\mathsf{Var}}_{\boldsymbol{M}}(F(\boldsymbol{M})) \; \leq \; o_{n_0}(1) \, ,$$

that is, both are converging to 0 at an explict rate determined by n_0 (with some dependence on the degree d of A and B). Interestingly however, in the case of $d = 1$, it turns out that $F(\boldsymbol{M})$ is in fact an unbiased estimator of $\langle A, B \rangle_{\mathcal{G}_\rho^{\otimes n}}$. Indeed, this is not a coincidence! We leave it to the interested reader to figure out that in the case of $d = 1$, our transformation is in fact identical to the above described JL transformation on the n-dimensional space of Hermite coefficients of A and B.

Our actual transformation is slightly different, namely $\Psi_{\boldsymbol{M}} A(a) = A(\boldsymbol{M}a/\|a\|_2)$. This is to ensure item 2, about preserving the closeness of the output of A to Δ_k. The proof gets a little more technical due to this change, but is intuitively similar to the above transformation since $\|\boldsymbol{a}\|_2$ is tightly concentrated around $\sqrt{n_0}$. It is important to note that item 2 is quite critical to the entire approach. If it were not for this restriction, item 1 is very easy to satisfy on its own by other more direct dimension reduction operations on the Hermite coefficients.

The mean and variance bounds on $F(\boldsymbol{M})$ are presented as Lemma 10. This is the most technical part of this work, but we stress that the main ideas are conceptually simple and elementary (for the most part). We provide a brief sketch of the proof in subsection 3.1 that illustrates all the main ideas in under a page, and defer all details to Appendix A. To prove these mean and variance bounds, we first analyze the case when A and B are multi-linear monomials (subsection A.2) and then combine these monomial calculations to obtain bounds for general multilinear polynomials (subsection A.3).

1.5 Comparisons with recent works of De, Mossel & Neeman

Our main theorems (Theorems 1, 2, 5) significantly improve the bounds in the versions proved by De, Mossel & Neeman [19, 20]. Our work was inspired by [19, 20] through several high-level ideas, such as the use of the transformation to low-degree and multilinear polynomials (although these tranformations are technically different in our case). However, it seems that the key insight into "*why dimension reduction is possible*" provided by the works of De Mossel & Neeman and the current work are fundamentally different.

The key insight for dimension reduction in the work of De, Mossel & Neeman is (quoting [19]): *"the fact that a collection of homogeneous polynomials can be replaced by polynomials in bounded dimensions is a tensor analogue of the fact that for any k vectors in $\mathbb{R}^n$, there exist k vectors in $\mathbb{R}^k$ with the same matrix of inner products"*. By contrast, the main intuition in our work is an "oblivious" dimension reduction technique, very similar to the Johnson-Lindenstrauss Lemma, as described in subsection 1.4.

Also, we point out a minor difference in our versions of Theorem 1. In [19] the function $\widetilde{f}$ maps to $[k]$, while in our theorem $\widetilde{f}$ maps to Δ_k. Interestingly however, this is not a major difference and it follows from a thresholding lemma in [19, Lemma 15 & 16] that any such $\widetilde{f}$ can be modified to have range $[k]$, while preserving $\mathbb{E}[\widetilde{f}]$ without decreasing the noise stability.

1.6 Other Related Work and Future Directions

Information Theory

Several previous works in information theory and theoretical computer science study "non-interactive simulation" type of questions. For instance, the non-interactive simulation of joint distributions question studied in this work is a generalization of the "non-interactive correlation distillation" problem[6] which was studied by [43, 45]. Moreover, recent works in the information theory community [37, 8] derive analytical tools (based on hypercontractivity and the so-called *strong data processing constant*) to prove impossibility results for NIS. While these results provide stronger bounds for some sources, they are not tight in general. Finally, the "non-interactive agreement distillation" problem studied by [9] can also be viewed as a particular case of the NIS setup.

Randomness in Computation

As discussed in [29], one motivation for studying NIS problems stems from the study of the role of randomness in distributed computing. Specifically, recent works in cryptography [2, 3, 11, 17, 41, 51], quantum computing [47, 16, 23] and communication complexity [6, 15, 28, 26] study how the ability to solve various computational tasks gets affected by weakening the source of shared randomness. In this context, it is very natural to ask how well can a source of randomness be transformed into another (more structured) one, which is precisely the setup of non-interactive simulation.

The classic Newman's theorem [46] tells us that any communication protocol with n-bit inputs and 0-1 outputs can be simulated with only $O(\log n)$ bits of randomness. On the other hand, if we consider the setting where Alice and Bob run a communication protocol with correlated randomness, such as those defined in [6, 15], then reducing the randomness requirement of such protocols is not clear. Theorem 5 implies randomness reduction for zero-communication or even simultaneous message protocols, and hence can be seen as a first step towards understanding the randomness requirements of arbitrary (one or two way) communication protocols with access to correlated randomness.

Tensor Power problems

Another motivation comes from the fact that NIS belongs to the class of *tensor power* problems, which have been very challenging to analyze. In such questions, the goal is to understand how some combinatorial quantity behaves in terms of the dimensionality of the problem as the dimension tends to infinity. A famous instance of such problems is the *Shannon capacity of a graph* [53, 40] where the aim is to understand how the independence number of the power of a graph behaves in terms of the exponent. The question of showing the computability of the Shannon capacity remains open to this day [4]. Other examples of such open problems (which are more closely related to NIS) arise in the problems of *local state transformation of quantum entanglement* [7, 22], the problem of computing the *entangled value of a 2-prover 1-round game* (see for, e.g., [38] and also the open problems [1]). Another example is the problem of computing the *amortized value of parallel repetitions of a 2-prover 1-round game* [49, 34, 48, 50, 5]. While we don't have computability results for the amortized value, there has been a recent work that tries to characterize it in terms of an information theoretic quantity [12]. Yet another example of a tensor-power problem is the

[6] which considered the problem of maximizing agreement on a single bit, in various multi-party settings.

task of computing the *amortized communication complexity of a communication problem*. Braverman-Rao [13] showed that this equals the information complexity of the communication problem, however the computability of information complexity was shown only recently [14].

We hope that the recent progress on the Non-Interactive Simulation problem would stimulate progress on these other notable tensor-power problems. A concrete question is whether the techniques used for NIS (regularity lemma, invariance principle, etc.) can be translated to any of the above mentioned setups.

Deterministic Approximate Counting

We also point out that the notions of eigenregularity used in [19, 20] were originally introduced and used in [21] to give the only known fixed-polynomial time *deterministic* approximate counting algorithm for polynomial threshold functions (PTFs). Our randomized techniques don't seem directly applicable to the PTF counting problem, as the emphasis there is on being *deterministic*. However, it will be interesting if our techniques could yield some further insights into approximate counting problems and pseudorandomness in general.

1.7 Organization of the Paper

In section 2, we summarize some useful definitions and provide a simple lemma that will be useful later. In section 3, we state our main technique of dimension reduction for polynomials (Theorem 6) and provide a brief sketch of the proof, with most details deferred to Appendix A. In section 4, we describe the transformations to make functions low-degree and multilinear, with proofs deferred to Appendix B. Finally, in section 5, we prove Theorem 2 (which implies Theorem 1 as a corollary).

2 Preliminaries

2.1 Gaussian Probability Spaces

Throughout this paper, we deal with the n-dimensional Gaussian space, i.e. $\mathbb{R}^n$ equipped with Gaussian measure γ_n given by the density function

$$\frac{d\gamma_n}{d\boldsymbol{X}} := \frac{1}{(2\pi)^{n/2}} \cdot \exp\left(-\frac{1}{2} \cdot \|\boldsymbol{X}\|_{\mathbb{R}^n}^2\right).$$

where $\|\cdot\|_{\mathbb{R}^n}$ denotes the ℓ_2 norm of a vector. We use letters such as X, Y to denote points in $\mathbb{R}^n$, bold symbols such as $\boldsymbol{X}$, $\boldsymbol{Y}$ to denote random variables, subscripts such as X_i or $\boldsymbol{X}_i$ denote the i-th coordinate.

The ℓ_2-norm of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as $\|f\|_2 := \left[\mathop{\mathbb{E}}_{\boldsymbol{X}\sim\gamma_n} f(\boldsymbol{X})^2\right]^{1/2}$. We use $L^2(\mathbb{R}^n, \gamma_n)$ to denote the space of all ℓ_2-integrable functions $f : \mathbb{R}^n \to \mathbb{R}$, i.e. $\|f\|_2 < \infty$. All functions we consider will be ℓ_2-integrable. The inner product of $f, g \in L^2(\mathbb{R}^n, \gamma_n)$ is defined as $\langle f, g\rangle_{\gamma_n} := \mathbb{E}_{\boldsymbol{X}\sim\gamma_n}[f(\boldsymbol{X})g(\boldsymbol{X})]$.

The joint distribution of ρ-correlated Gaussians is denoted as $\mathcal{G}_\rho$, which is a 2-dimensional Gaussian distribution $(\boldsymbol{X}, \boldsymbol{Y})$, where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are marginally distributed according to γ_1, with $\mathbb{E}[\boldsymbol{X}\boldsymbol{Y}] = \rho$. For $A, B \in L^2(\mathbb{R}^n, \gamma_n)$, the *noisy correlation between A and B over $\mathcal{G}_\rho^{\otimes n}$* is defined as,

$$\langle A, B\rangle_{\mathcal{G}_\rho^{\otimes n}} := \mathop{\mathbb{E}}_{(\boldsymbol{X},\boldsymbol{Y})\sim\mathcal{G}_\rho^{\otimes n}} [A(\boldsymbol{X}) \cdot B(\boldsymbol{Y})]$$

Finally, the total variation distance between two distributions μ and ν over domain Ω is defined as,

$$d_{\mathrm{TV}}(\mu,\nu) \;:=\; \sup_{S\subseteq\Omega} \left|\mu(S)-\nu(S)\right|.$$

2.2 Hermite Analysis

The set of Hermite polynomials $\{H_r : \mathbb{R}\to\mathbb{R} : r\in\mathbb{Z}_{\geq 0}\}$ form an orthonormal basis for functions in $L^2(\mathbb{R},\gamma_1)$ with respect to the inner product $\langle\cdot,\cdot\rangle_{\gamma_1}$. The r-th Hermite polynomial $H_r:\mathbb{R}\to\mathbb{R}$ (for $r\in\mathbb{Z}_{\geq 0}$) is defined as,

$$H_0(x)=1;\quad H_1(x)=x;\quad H_r(x)=\frac{(-1)^r}{\sqrt{r!}}e^{x^2/2}\cdot\frac{d^r}{dx^r}e^{-x^2/2}\,.$$

Hermite polynomials can also be obtained via the generating function, $e^{xt-\frac{t^2}{2}}=\sum_{r=0}^{\infty}\frac{H_r(x)}{\sqrt{r!}}\cdot t^r$.

For any $\sigma=(\sigma_1,\ldots,\sigma_n)\in\mathbb{Z}^n_{\geq 0}$, define $H_\sigma:\mathbb{R}^n\to\mathbb{R}$ as $H_\sigma(\boldsymbol{X}) = \prod_{i=1}^n H_{\sigma_i}(\boldsymbol{X}_i)$. It easily follows that the set $\left\{H_\sigma : \sigma\in\mathbb{Z}^n_{\geq 0}\right\}$ forms an orthonormal basis for $L^2(\mathbb{R}^n,\gamma_n)$. Thus, every $A\in L^2(\mathbb{R}^n,\gamma_n)$ has a *Hermite expansion* given by $A(\boldsymbol{X})=\sum_{\sigma\in\mathbb{Z}^n_{\geq 0}}\widehat{A}(\sigma)\cdot H_\sigma(\boldsymbol{X})$, where the $\widehat{A}(\sigma)$'s are the *Hermite coefficients* of A obtained as $\widehat{A}(\sigma)=\langle A,H_\sigma\rangle_{\gamma_n}$. The degree of σ is defined as $|\sigma|:=\sum_{i\in[n]}\sigma_i$, and the degree of A is the largest $|\sigma|$ for which $\widehat{A}(\sigma)\neq 0$. We say that $A\in L^2(\mathbb{R}^n,\gamma_n)$ is *multilinear* if $\widehat{A}(\sigma)$ is non-zero only if $\sigma_i\in\{0,1\}$ for all $i\in[n]$.

We list several useful facts about Hermite coefficients:

(1) Parseval's identity: $\left\|A\right\|_2^2=\sum_{\sigma\in\mathbb{Z}^n_{\geq 0}}\widehat{A}(\sigma)^2$ and $\mathsf{Var}(A)=\sum_{\mathbf{0}\neq\sigma\in\mathbb{Z}^n_{\geq 0}}\widehat{A}(\sigma)^2$.

(2) Plancherel's identity: $\langle A,A'\rangle_{\gamma_n}=\sum_{\sigma\in\mathbb{Z}^n_{\geq 0}}\widehat{A}(\sigma)\widehat{A'}(\sigma)$.

(3) Ornstein-Uhlenbeck operator: $U_\rho A(X)=\sum_{\sigma\in\mathbb{Z}^n_{\geq 0}}\rho^{|\sigma|}\cdot\widehat{A}(\sigma)\cdot H_\sigma(X)$.

(4) Noisy Correlation: $\langle A,B\rangle_{\mathcal{G}_\rho^{\otimes n}}=\langle A,U_\rho B\rangle_{\gamma_n}=\sum_{\sigma\in\mathbb{Z}^n_{\geq 0}}\rho^{|\sigma|}\widehat{A}(\sigma)\widehat{B}(\sigma)$

For convenience, $U_\rho(X)$ denotes the distribution $(\rho X+\sqrt{1-\rho^2}\boldsymbol{Z})$ where $\boldsymbol{Z}\sim\gamma_n$, for any $X\in\mathbb{R}^n$.

2.3 Vector-valued functions

For any function $A:\mathbb{R}^n\to\mathbb{R}^k$, we will interpret A as a vector of functions $(A_1,\cdots,A_k)$, where $A_i:\mathbb{R}^n\to\mathbb{R}$ is the i-th coordinate of the output of A. The definitions of Hermite analysis extend naturally to vector-valued functions as follows. For $A:\mathbb{R}^n\to\mathbb{R}^k$, the Hermite coefficient $\widehat{A}(\sigma)$ is $\left(\widehat{A}_1(\sigma),\ldots,\widehat{A}_k(\sigma)\right)\in\mathbb{R}^k$. We can extend the definition of ℓ_2-norm as $\|A\|^2:=\mathbb{E}_{\boldsymbol{X}\sim\gamma_n}\|A(\boldsymbol{X})\|^2$ or equivalently $\|A_1\|^2+\cdots+\|A_k\|^2=\sum_{\sigma\in\mathbb{Z}^n_{\geq 0}}\|\widehat{A}(\sigma)\|^2$. Also, $\deg(A):=\max_{i\in[k]}\deg(A_i)$. Again, all the vector-valued functions with domain $\mathbb{R}^n$ that we consider will be such that the function in each coordinate is in $L^2(\mathbb{R}^n,\gamma_n)$.

For $k\in\mathbb{N}$ and $i\in[k]$, let e_i be the unit vector along coordinate i in $\mathbb{R}^k$. The simplex Δ_k is defined as the convex hull formed by $\{e_i : i\in[k]\}$. Equivalently, $\Delta_k=\left\{v\in\mathbb{R}^k : \|v\|_1=1\right\}$ is the set of probability distributions over $[k]$. While we will consider vector-valued functions mapping to $\mathbb{R}^k$, we will be primarily interested in functions which map to Δ_k. The rounding operator $\mathcal{R}^{(k)}:\mathbb{R}^k\to\Delta_k$ maps any $v\in\mathbb{R}^k$ to its closest point in Δ_k. In particular, it is the identity map on Δ_k. We will drop the superscript on $\mathcal{R}$, as k is fixed throughout this paper. Similar to our notation for vector-valued functions, $\mathcal{R}_i$ denotes the i-th coordinate of $\mathcal{R}$. Thus, while the i-th coordinate of A is denoted by A_i, the i-th coordinate of $\mathcal{R}(A)$ is denoted by $\mathcal{R}_i(A)$.

As mentioned already, an important relaxation in our work is to consider functions that do not map to Δ_k, but instead map to $\mathbb{R}^k$. For such functions to be meaningful, we will require that the outputs are *usually close* to Δ_k, in which case, we will be rounding them to the simplex Δ_k. Towards this end, the following simple proposition will be very useful, which says that if we modify the strategies of Alice and Bob slightly (in ℓ_2-distance), then the correlation between the strategies does not change significantly. The proof follows by a simple triangle inequality and the Cauchy-Schwarz inequality.

▶ **Proposition 7** (Close strategies, have similar correlations). *Let $A, \widetilde{A}, B, \widetilde{B} \in L^2(\mathbb{R}^n, \gamma_n)$ such that $\|A\|_2$, $\|\widetilde{A}\|_2$, $\|B\|_2$, $\|\widetilde{B}\|_2 \le 1$. If $\|A - \widetilde{A}\|_2 \le \varepsilon$ and $\|B - \widetilde{B}\|_2 \le \varepsilon$, then it holds that,*

$$\left| \left\langle \widetilde{A}, \widetilde{B} \right\rangle_{\mathcal{G}_\rho^{\otimes n}} - \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes n}} \right| \;\le\; 2\varepsilon \;.$$

3 Dimension Reduction for Low-Degree Multilinear Polynomials

In this section, we present our main technique of dimension reduction for low-degree multilinear polynomials over Gaussian space. Theorem 6 is obtained immediately as a combination of Theorem 8 and Proposition 9 stated below.

▶ **Theorem 8.** *Given $d \in \mathbb{Z}_{>0}$, $\rho \in [0,1]$ and $\delta > 0$, there exists an explicitly computable $n_0 = n_0(d, \delta)$, such that the following holds:*
Let $A : \mathbb{R}^n \to \mathbb{R}$ and $B : \mathbb{R}^n \to \mathbb{R}$ be degree-d multilinear polynomials, s.t. $\|A\|_2, \|B\|_2 \le 1$. For $\boldsymbol{M} \in \mathbb{R}^{n \times n_0}$ with entries i.i.d. sampled from γ_1, define the functions[7] $A_{\boldsymbol{M}} : \mathbb{R}^{n_0} \to \mathbb{R}$ and $B_{\boldsymbol{M}} : \mathbb{R}^{n_0} \to \mathbb{R}$ as

$$A_{\boldsymbol{M}}(a) \;=\; A\left(\frac{\boldsymbol{M}a}{\|a\|_2}\right) \quad \text{and} \quad B_{\boldsymbol{M}}(b) = B\left(\frac{\boldsymbol{M}b}{\|b\|_2}\right) \qquad \text{for } a, b \in \mathbb{R}^{n_0} \setminus \{0\} \;.$$

Then, with probability at least $1 - \delta$ (over the choice of $\boldsymbol{M}$), it holds that,

$$\left| \langle A_{\boldsymbol{M}}, B_{\boldsymbol{M}} \rangle_{\mathcal{G}_\rho^{\otimes n_0}} \;-\; \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes n}} \right| \;<\; \delta \;.$$

In particular, the explicit choice of n_0 is upper bounded as $\frac{d^{O(d)}}{\delta^4}$.

In other words, for a typical choice of $\boldsymbol{M} \sim \gamma_1^{\otimes(n \times n_0)}$, the correlation between A and B is approximately preserved if we replace $(\boldsymbol{X}, \boldsymbol{Y}) \sim \mathcal{G}_\rho^{\otimes n}$ by $(\boldsymbol{M}\boldsymbol{a}/\|\boldsymbol{a}\|_2, \boldsymbol{M}\boldsymbol{b}/\|\boldsymbol{b}\|_2)$, where $(\boldsymbol{a}, \boldsymbol{b}) \sim \mathcal{G}_\rho^{\otimes n_0}$. Intuitively, $\boldsymbol{M}$ can be thought of as a means to "stretch" n_0 coordinates of $\mathcal{G}_\rho$ into effectively n coordinates of $\mathcal{G}_\rho$, while "fooling" correlations between degree-d multilinear polynomials.

Before we prove the above theorem, we prove a simple proposition that completely handles item 2 of Theorem 6 by showing that if this dimension reduction were applied to vector-valued functions whose outputs lie close to the simplex Δ_k, then with high probability, even the dimension-reduced functions will have outputs close to the simplex. More formally,

▶ **Proposition 9.** *Let $A : \mathbb{R}^n \to \mathbb{R}^k$ and $B : \mathbb{R}^n \to \mathbb{R}^k$, such that $\|\mathcal{R}(A) - A\|_2$, $\|\mathcal{R}(B) - B\|_2 \le \delta$. Then, with probability at least $1 - 2\delta$ (over choice of $\boldsymbol{M}$), it holds that,*

$$\|\mathcal{R}(A_{\boldsymbol{M}}) - A_{\boldsymbol{M}}\|_2 \le \sqrt{\delta} \quad \text{and} \quad \|\mathcal{R}(B_{\boldsymbol{M}}) - B_{\boldsymbol{M}}\|_2 \le \sqrt{\delta} \;.$$

[7] $A_{\boldsymbol{M}}$ and $B_{\boldsymbol{M}}$ can be defined arbitrarily on $0 \in \mathbb{R}^{n_0}$. This is inconsequential as $\{0\}$ is a measure zero set under γ_n.

Proof. Observe that even for a fixed non-zero $a \in \mathbb{R}^{n_0}$, the distribution of $\frac{\boldsymbol{M}a}{\|a\|_2}$ is identical to that of a standard n-variate Gaussian distribution γ_n. Thus, we immediately have that,

$$\mathop{\mathbb{E}}_{\boldsymbol{M}} \mathop{\mathbb{E}}_{\boldsymbol{a}} \left\| \mathcal{R}\left(A\left(\frac{\boldsymbol{M}\boldsymbol{a}}{\|\boldsymbol{a}\|_2}\right)\right) - A\left(\frac{\boldsymbol{M}\boldsymbol{a}}{\|\boldsymbol{a}\|_2}\right)\right\|_2^2 = \mathop{\mathbb{E}}_{\boldsymbol{X}} \left\| \mathcal{R}\left(A\left(\boldsymbol{X}\right)\right) - A\left(\boldsymbol{X}\right)\right\|_2^2$$

$$\text{Alternately,} \qquad \mathop{\mathbb{E}}_{\boldsymbol{M}} \left\| \mathcal{R}(A_{\boldsymbol{M}}) - A_{\boldsymbol{M}}\right\|_2^2 = \left\| \mathcal{R}(A) - A\right\|_2^2 \leq \delta^2$$

Thus, by Markov's inequality, $\left\| \mathcal{R}(A_{\boldsymbol{M}}) - A_{\boldsymbol{M}}\right\|_2 \leq \sqrt{\delta}$ holds with probability at least $1-\delta$. We can similarly argue for $B_{\boldsymbol{M}}$, and a union bound completes the proof. ◀

To prove Theorem 8, we primarily use the second moment method (i.e., Chebyshev's inequality). In particular, let $F(\boldsymbol{M})$ be defined as,

$$F(\boldsymbol{M}) \stackrel{\text{def}}{=} \langle A_{\boldsymbol{M}}, B_{\boldsymbol{M}} \rangle_{\mathcal{G}_\rho^{\otimes n_0}}$$

The most technical part of this work is to show sufficently good bounds on the mean and variance of $F(\boldsymbol{M})$ for a random choice of $\boldsymbol{M} \sim \gamma_1^{\otimes(n \times n_0)}$, given by the following lemma.

▶ **Lemma 10.** (Mean & Variance Bound). *Given d and δ, there exists an explicitly computable $n_0 := n_0(d, \delta)$ such that for $\boldsymbol{M} \sim \gamma_1^{\otimes(n \times n_0)}$,*

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{M}} F(\boldsymbol{M}) - \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes n}} \right| \leq \delta \qquad \textbf{(Mean bound)}$$

$$\mathop{\mathsf{Var}}_{\boldsymbol{M}} \left(F(\boldsymbol{M})\right) \leq \delta \qquad \textbf{(Variance bound)}$$

In particular, one may take $n_0 = \frac{d^{O(d)}}{\delta^2}$.

We provide a little sketch of the proof of Lemma 10 below, with the full details in Appendix A. Assuming Lemma 10, we can easily prove Theorem 8.

Proof of Theorem 8. We invoke Lemma 10 with parameters d and $\delta^2/2$, to get a choice of $n_0 = \frac{d^{O(d)}}{\delta^4}$. Using Chebyshev's inequality and the Variance bound in Lemma 10, we have that for any $\eta > 0$,

$$\Pr_{\boldsymbol{M}} \left[\left| F(\boldsymbol{M}) - \mathbb{E}_{\boldsymbol{M}} F(\boldsymbol{M}) \right| > \eta \right] \leq \frac{\delta^2}{2\eta}\,.$$

Using the triangle inequality, and the Mean bound in Lemma 10, we get

$$\begin{aligned} &\Pr_{\boldsymbol{M}} \left[\left| F(\boldsymbol{M}) - \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes n}} \right| > \delta \right] \\ &\leq \Pr_{\boldsymbol{M}} \left[\left| F(\boldsymbol{M}) - \mathbb{E}_{\boldsymbol{M}} F(\boldsymbol{M}) \right| + \left| \mathbb{E}_{\boldsymbol{M}} F(\boldsymbol{M}) - \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes n}} \right| > \delta \right] \\ &\leq \Pr_{\boldsymbol{M}} \left[\left| F(\boldsymbol{M}) - \mathbb{E}_{\boldsymbol{M}} F(\boldsymbol{M}) \right| > \delta - \delta^2 \right] \leq \delta. \end{aligned}$$ ◀

3.1 Proof Sketch of Lemma 10

While the proof of Lemma 10 is somewhat technical as a whole, the main driver of the entire lemma is a simple combinatorial fact that if we sample d times with replacement from a bag with n_0 items, then the probability of not sampling distinct items is at most $O(d^2/n_0) = o_{n_0}(1)$. We briefly illustrate this idea at play by proving a *simpler* version of the mean bound. For this section, let's consider a different dimension reduction of setting $A_{\boldsymbol{M}}$ and $B_{\boldsymbol{M}}$ as, $A_{\boldsymbol{M}}(a) = A(\boldsymbol{M}a/\sqrt{n_0})$ and $B_{\boldsymbol{M}}(b) = B(\boldsymbol{M}b/\sqrt{n_0})$, where $\boldsymbol{M} \sim \gamma_1^{\otimes(n \times n_0)}$.

Let $\boldsymbol{m}_i \in \mathbb{R}^{n_0}$ denote the vector corresponding to the i-th row of $\boldsymbol{M}$. Consider the mean of $F(\boldsymbol{M}) = \langle A_{\boldsymbol{M}}, B_{\boldsymbol{M}} \rangle_{\mathcal{G}_\rho^{\otimes n_0}}$:

$$\begin{aligned}\mathop{\mathbb{E}}_{\boldsymbol{M}} F(\boldsymbol{M}) &= \mathop{\mathbb{E}}_{\boldsymbol{M}} \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} A\left(\frac{\boldsymbol{M}\boldsymbol{a}}{\sqrt{n_0}}\right) B\left(\frac{\boldsymbol{M}\boldsymbol{b}}{\sqrt{n_0}}\right) \\ &= \mathop{\mathbb{E}}_{\boldsymbol{M}} \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \sum_{\sigma,\kappa} \frac{\widehat{A}(\sigma)\widehat{B}(\kappa)}{n_0^{(|\sigma|+|\kappa|)/2}} \cdot \prod_{i\,:\,\sigma_i=1} \langle \boldsymbol{m}_i, \boldsymbol{a}\rangle \cdot \prod_{j\,:\,\kappa_j=1} \langle \boldsymbol{m}_j, \boldsymbol{b}\rangle\end{aligned}$$

where, recall that A and B are multilinear, and so the relevant σ and κ are in $\{0,1\}^n$, with $|\sigma|, |\kappa| \le d$. Next, observe that $\mathbb{E}\, \boldsymbol{m}_i \boldsymbol{m}_i^T = I_{n_0 \times n_0}$, and hence we get that,

$$\mathop{\mathbb{E}}_{\boldsymbol{M}} \prod_{i\,:\,\sigma_i=1} \langle \boldsymbol{m}_i, \boldsymbol{a}\rangle \cdot \prod_{j\,:\,\kappa_j=1} \langle \boldsymbol{m}_j, \boldsymbol{b}\rangle = \begin{cases} \langle \boldsymbol{a}, \boldsymbol{b}\rangle^{|\sigma|} & \text{if } \sigma = \kappa \\ 0 & \text{if } \sigma \neq \kappa \end{cases}.$$

Finally, we observe that if we expand $\langle \boldsymbol{a}, \boldsymbol{b}\rangle^d$ as $\sum_{i_1,\ldots,i_d \in [n_0]} \boldsymbol{a}_{i_1}\boldsymbol{b}_{i_1} \ldots \boldsymbol{a}_{i_d}\boldsymbol{b}_{i_d}$, then from the combinatorial fact above, except for a $O(d^2) \cdot n_0^{d-1}$ out of total n_0^d terms, the indices $i_1, \ldots, i_d$ are all distinct. It is immediate to see that if all the i_j's are distinct then $\mathbb{E}_{\boldsymbol{a},\boldsymbol{b}} \boldsymbol{a}_{i_1}\boldsymbol{b}_{i_1} \ldots \boldsymbol{a}_{i_d}\boldsymbol{b}_{i_d} = \rho^d$. Additionally, we show that if the i_j's are not all distinct then $|\mathbb{E}_{\boldsymbol{a},\boldsymbol{b}} \boldsymbol{a}_{i_1}\boldsymbol{b}_{i_1} \ldots \boldsymbol{a}_{i_d}\boldsymbol{b}_{i_d}| \le d^{O(d)}$ (this follows from the fact that for the d-th moments of γ_1 are at most $d^{O(d)}$). Putting this together we get for any σ (with $|\sigma| \le d$) that,

$$\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \frac{\langle \boldsymbol{a}, \boldsymbol{b}\rangle^{|\sigma|}}{n_0^{|\sigma|}} = \rho^{|\sigma|} \pm \frac{d^{O(d)}}{n_0}$$

Putting everything together we get,

$$\mathop{\mathbb{E}}_{\boldsymbol{M}} F(\boldsymbol{M}) = \sum_{\sigma} \widehat{A}(\sigma)\widehat{B}(\sigma) \cdot \left(\rho^{|\sigma|} \pm \frac{d^{O(d)}}{n_0}\right) = \langle A, B\rangle_{\mathcal{G}_\rho^{\otimes n}} \pm \sum_{\sigma} \widehat{A}(\sigma)\widehat{B}(\sigma) \cdot \frac{d^{O(d)}}{n_0}$$

And hence,

$$\left|\mathbb{E}_{\boldsymbol{M}} F(\boldsymbol{M}) - \langle A, B\rangle_{\mathcal{G}_\rho^{\otimes n}}\right| \le \frac{d^{O(d)}}{n_0} \cdot \sum_{\sigma} \widehat{A}(\sigma)\widehat{B}(\sigma) \le \frac{d^{O(d)}}{n_0} \cdot \|A\|_2 \cdot \|B\|_2 \le \delta,$$

where we use the Cauchy-Schwarz inequality and that $n_0 \ge d^{O(d)}/\delta$. This completes a proof sketch of the mean bound in Lemma 10. Replacing $\sqrt{n_0}$ by $\|\boldsymbol{a}\|_2$ introduces a minor technicality, but still works because $\|\boldsymbol{a}\|_2$ is tightly concentrated around $\sqrt{n_0}$. The variance bound is slightly more complicated with the use of a hypercontractive inequality instead of Cauchy-Schwarz. The full details of the proof are in Appendix A.

4 Transformation to Low-Degree Multilinear Polynomials

While Theorem 6 applies only for low-degree multilinear polynomials, we can extend it for all functions by using the following lemma that transforms k-dimensional ℓ_2-integrable functions $A : \mathbb{R}^n \to \mathbb{R}^k$ and $B : \mathbb{R}^n \to \mathbb{R}^k$ into low-degree multilinear polynomials while approximately preserving all correlations and also not deviating much from the simplex Δ_k (although slightly increasing the number of variables).

▶ **Lemma 11** (Low-Degree Multilinear Transformation). *Given parameters $\rho \in [0,1]$, $\delta > 0$, $k \in \mathbb{N}$, there exists an explicit $d = d(k,\rho,\delta)$ and $t := t(k,d,\delta)$ such that the following holds: Let $A : \mathbb{R}^n \to \mathbb{R}^k$ and $B : \mathbb{R}^n \to \mathbb{R}^k$, s.t. for any $i \in [k]$, it holds that $\mathsf{Var}(A_i), \mathsf{Var}(B_i) \le 1$. Then, there exist functions $\widetilde{A} : \mathbb{R}^{nt} \to \mathbb{R}^k$ and $\widetilde{B} : \mathbb{R}^{nt} \to \mathbb{R}^k$ such that the following statements hold.*

1. $\widetilde{A}$ *and* $\widetilde{B}$ *are multilinear with degree at most* d.
2. *For any* $i \in [k]$, *it holds that* $\mathsf{Var}(\widetilde{A}_i) \le \mathsf{Var}(A_i) \le 1$ *and* $\mathsf{Var}(\widetilde{B}_i) \le \mathsf{Var}(B_i) \le 1$.
3. $\left\|\mathcal{R}(\widetilde{A}) - \widetilde{A}\right\|_2 \le \left\|\mathcal{R}(A) - A\right\|_2 + \delta$ *and* $\left\|\mathcal{R}(\widetilde{B}) - \widetilde{B}\right\|_2 \le \left\|\mathcal{R}(B) - B\right\|_2 + \delta$
4. *For every* $i, j \in [k]$,

$$\left|\left\langle \widetilde{A}_i, \widetilde{B}_j \right\rangle_{\mathcal{G}_\rho^{\otimes nt}} - \langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes n}}\right| \le \frac{\delta}{\sqrt{k}}$$

In particular, one may take $d = O\left(\frac{\sqrt{k}\log^2(k/\delta)}{\delta(1-\rho)}\right)$ *and* $t = O\left(\frac{kd^2}{\delta^2}\right)$.

This lemma is itself proved in two stages. The first stage transforms general functions to low-degree polynomials by applying a small noise operator (making the functions have "decaying Hermite tails") followed by truncation of the higher degree terms. The second stage transforms low-degree polynomials into multilinear ones, by replacing each variable by a normalized sum of new variables (making the functions have low mass on non-multilinear terms) followed by truncation of the non-multilinear terms.

These techniques are quite standard in literature. For the use of noise operator in the first stage see e.g. [44, 42]. For the substitution of variables in the second stage see e.g. [19]. However, since we are stating particular quantitative versions of the lemmas, we provide the proofs in Appendix B for completeness.

5 Non-Interactive Simulation from Correlated Gaussian Sources

In this section, we complete the proof of our main theorem regarding non-interactive simulation from correlated Gaussian sources, i.e. Theorem 2. Recall that it immediately implies Theorem 1 by setting $A = B = f$ and obtaining $\widetilde{f} = \widetilde{A} = \widetilde{B}$.

Proof of Theorem 2. Starting with functions $A : \mathbb{R}^n \to \Delta_k$ and $B : \mathbb{R}^n \to \Delta_k$, we first apply Lemma 11 to transform A and B to low-degree and multilinear polynomials, and subsequently apply Theorem 8. Unfortunately after these transformations, the range is no longer restricted to Δ_k. Nevertheless, we do have that these transformations ensure that the functions still output something "close" to the simplex Δ_k. This allows us to apply the rounding operator and get the range as Δ_k again (using Lemma 7). An overview of the transformations done is presented in Figure 3.

We thus transform A and B through each of the following steps. At each step, we approximately preserve the correlation $\langle A_i, B_j \rangle$ for every $i, j \in [k]$. Additionally, in each step $\left\|\mathcal{R}(A) - A\right\|_2$ and $\left\|\mathcal{R}(B) - B\right\|_2$ doesn't increase significantly (note that, to begin with, the range of A and B is Δ_k and hence we start with $\left\|\mathcal{R}(A) - A\right\|_2 = \left\|\mathcal{R}(B) - B\right\|_2 = 0$).

1. **Transformation to Low-Degree & Multilinear**: We apply Lemma 11 on A and B with parameter δ (chosen later), setting $d = d(\rho, k, \delta)$ and $t = t(d, k, \delta)$ as required, to get degree-d and multilinear $A^{(1)} : \mathbb{R}^n \to \mathbb{R}^k$ and $B^{(1)} : \mathbb{R}^n \to \mathbb{R}^k$. Moreover, we have that for every $i, j \in [k]$,

$$\left|\left\langle A_i^{(1)}, B_j^{(1)} \right\rangle_{\mathcal{G}_\rho^{\otimes nt}} - \langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes n}}\right| \le \delta \tag{2}$$

   Additionally, we have $\left\|\mathcal{R}(A^{(1)}) - A^{(1)}\right\|_2 \le \left\|\mathcal{R}(A) - A\right\|_2 + \delta \le \delta$ and similarly for $B^{(1)}$.

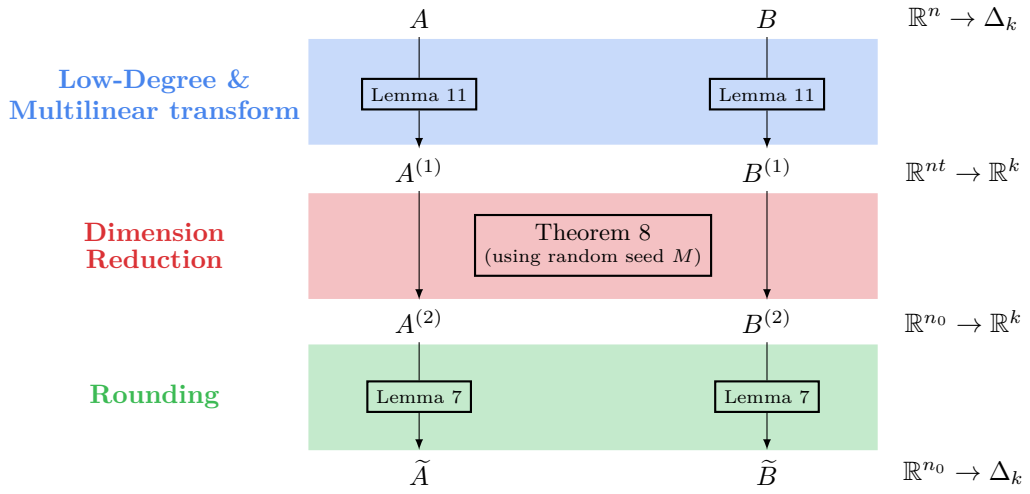


Figure 3 Transformations for Non-interactive simulation from Correlated Gaussian Sources

2. **Dimension reduction**: We apply Theorem 8 with parameter δ/k^2, setting $n_0 = n_0(d, \rho, \delta/k^2)$ as required, on individual coordinates of $A^{(1)}$ and $B^{(1)}$ to obtain functions $A^{(2)} : \mathbb{R}^{n_0} \to \mathbb{R}^k$ and $B^{(2)} : \mathbb{R}^{n_0} \to \mathbb{R}^k$. Taking a union bound, we have that with probability at least $1 - \delta$, it holds for every $i, j \in [k]$ that,

$$\left| \left\langle A_i^{(2)}, B_j^{(2)} \right\rangle_{\mathcal{G}_\rho^{\otimes n_0}} - \left\langle A_i^{(1)}, B_j^{(1)} \right\rangle_{\mathcal{G}_\rho^{\otimes nt}} \right| \leq \delta \tag{3}$$

From Proposition 9, we have that with probability $1 - 4\delta$,

$$\left\| \mathcal{R}(A^{(2)}) - A^{(2)} \right\|_2 \leq \sqrt{\left\| \mathcal{R}(A^{(1)}) - A^{(1)} \right\|_2} \leq \sqrt{\delta}$$

$$\left\| \mathcal{R}(B^{(2)}) - B^{(2)} \right\|_2 \leq \sqrt{\left\| \mathcal{R}(B^{(1)}) - B^{(1)} \right\|_2} \leq \sqrt{\delta}$$

Note that this is the only randomized step in the entire transformation, and it succeeds in obtaining the above three statements with probability at least $1 - 5\delta$.

3. **Rounding to Δ_k**: Finally, we set $\widetilde{A} = \mathcal{R}(A^{(2)})$ and $\widetilde{B} = \mathcal{R}(B^{(2)})$. Thus, assuming the previous step succeeds, we have that $\|\widetilde{A}_i - A_i^{(2)}\|_2 \leq \sqrt{\delta}$ and $\|\widetilde{B}_j - B_j^{(2)}\|_2 \leq \sqrt{\delta}$. Hence we can invoke Lemma 7, to conclude that,

$$\left| \left\langle \widetilde{A}_i, \widetilde{B}_j \right\rangle_{\mathcal{G}_\rho^{\otimes n_0}} - \left\langle A_i^{(2)}, B_j^{(2)} \right\rangle_{\mathcal{G}_\rho^{\otimes n_0}} \right| \leq 2\sqrt{\delta}. \tag{4}$$

Thus we started with functions $A : \mathbb{R}^n \to \Delta_k$ and $B : \mathbb{R}^n \to \Delta_k$ and ended with functions $\widetilde{A} : \mathbb{R}^{n_0} \to \Delta_k$ and $\widetilde{B} : \mathbb{R}^{n_0} \to \Delta_k$ such that for every $i, j \in [k]$ (by combining Equations 2, 3 and 4) it holds that,

$$\left| \left\langle \widetilde{A}_i, \widetilde{B}_j \right\rangle_{\mathcal{G}_\rho^{\otimes n_0}} - \langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes n}} \right| \leq O(\sqrt{\delta}) \, .$$

Thus, more strongly, if we instantiate $\delta = O(\varepsilon^2/k^4)$, then we get that our entire transformation succeeds with probability $1 - \varepsilon$ in obtaining $\widetilde{A}$ and $\widetilde{B}$ such that,

$$d_{\mathrm{TV}} \left((A(\boldsymbol{X}), B(\boldsymbol{Y}))_{\mathcal{G}_\rho^{\otimes n}}, \; (\widetilde{A}(\boldsymbol{a}), \widetilde{B}(\boldsymbol{b}))_{\mathcal{G}_\rho^{\otimes n_0}} \right) \leq \varepsilon \, .$$

It is easy to see that n_0 works out to be

$$n_0 = \frac{d^{O(d)}}{\delta^4} = \exp\left(\widetilde{O}\left(\frac{k^{4.5}}{\varepsilon^2(1-\rho)}\right)\right) = \exp\left(\mathrm{poly}\left(k, \frac{1}{\varepsilon}, \frac{1}{1-\rho}\right)\right) . \quad \blacktriangleleft$$

References

1 OpenQIProblemsWiki - All the Bell Inequalities. http://qig.itp.uni-hannover.de/qiproblems/1. Accessed: 2016-07-12.

2 Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography. part i: secret sharing. *IEEE Transactions on Information Theory*, 39(4), 1993.

3 Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography. ii. cr capacity. *Information Theory, IEEE Transactions on*, 44(1):225–240, 1998.

4 Noga Alon and Eyal Lubetzky. The shannon capacity of a graph and the independence numbers of its powers. *Information Theory, IEEE Transactions on*, 52(5):2172–2176, 2006.

5 Boaz Barak, Moritz Hardt, Ishay Haviv, Anup Rao, Oded Regev, and David Steurer. Rounding parallel repetitions of unique games. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 374–383. IEEE Computer Society, 2008. doi:10.1109/FOCS.2008.55.

6 Mohammad Bavarian, Dmitry Gavinsky, and Tsuyoshi Ito. On the role of shared randomness in simultaneous communication. In *Automata, Languages, and Programming*, pages 150–162. Springer, 2014.

7 Salman Beigi. A new quantum data processing inequality. *CoRR*, abs/1210.1689, 2012. arXiv:1210.1689.

8 Salman Beigi and Amin Gohari. On the duality of additivity and tensorization. *arXiv preprint arXiv:1502.00827*, 2015.

9 Andrej Bogdanov and Elchanan Mossel. On extracting common random bits from correlated sources. *Information Theory, IEEE Transactions on*, 57(10):6351–6355, 2011.

10 Christer Borell. Geometric bounds on the ornstein-uhlenbeck velocity process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 70(1):1–13, 1985.

11 Gilles Brassard and Louis Salvail. Secret-key reconciliation by public discussion. In *advances in Cryptology—EUROCRYPT'93*, pages 410–423. Springer, 1994.

12 Mark Braverman and Young Kun-Ko. Information value of two-prover games. In Anna R. Karlin, editor, *9th Innovations in Theoretical Computer Science Conference, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA*, volume 94 of *LIPIcs*, pages 12:1–12:15. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018. doi:10.4230/LIPIcs.ITCS.2018.12.

13 Mark Braverman and Anup Rao. Information equals amortized communication. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 748–757. IEEE, 2011.

14 Mark Braverman and Jon Schneider. Information complexity is computable. *arXiv preprint arXiv:1502.02971*, 2015.

15 Clement Cannonne, Venkat Guruswami, Raghu Meka, and Madhu Sudan. Communication with imperfectly shared randomness. *ITCS*, 2014.

16 Eric Chitambar, Runyao Duan, and Yaoyun Shi. Tripartite entanglement transformations and tensor rank. *Physical review letters*, 101(14):140502, 2008.

17 Imre Csiszár and Prakash Narayan. Common randomness and secret key generation with a helper. *Information Theory, IEEE Transactions on*, 46(2):344–366, 2000.

18 Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Struct. Algorithms*, 22(1):60–65, 2003. doi:10.1002/rsa.10073.

19 Anindya De, Elchanan Mossel, and Joe Neeman. Noise stability is computable and approximately low-dimensional. In Ryan O'Donnell, editor, *32nd Computational Complexity Conference, CCC 2017, July 6-9, 2017, Riga, Latvia*, volume 79 of *LIPIcs*, pages 10:1–10:11. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017. doi:10.4230/LIPIcs.CCC.2017.10.

20 Anindya De, Elchanan Mossel, and Joe Neeman. Non interactive simulation of correlated distributions is decidable. In Artur Czumaj, editor, *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 2728–2746. SIAM, 2018. doi:10.1137/1.9781611975031.174.

21 Anindya De and Rocco A Servedio. Efficient deterministic approximate counting for low-degree polynomial threshold functions. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 832–841. ACM, 2014.

22 Payam Delgosha and Salman Beigi. Impossibility of local state transformation via hypercontractivity. *CoRR*, abs/1307.2747, 2013. arXiv:1307.2747.

23 Payam Delgosha and Salman Beigi. Impossibility of local state transformation via hypercontractivity. *Communications in Mathematical Physics*, 332(1):449–476, 2014.

24 Peter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973.

25 Hans Gebelein. Das statistische problem der korrelation als variations-und eigenwertproblem und sein zusammenhang mit der ausgleichsrechnung. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 21(6):364–379, 1941.

26 Badih Ghazi and T. S. Jayram. Resource-efficient common randomness and secret-key schemes. In Artur Czumaj, editor, *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 1834–1853. SIAM, 2018. doi:10.1137/1.9781611975031.120.

27 Badih Ghazi, Pritish Kamath, and Prasad Raghavendra. Dimension reduction for polynomials over gaussian space and applications. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:125, 2017. URL: https://eccc.weizmann.ac.il/report/2017/125.

28 Badih Ghazi, Pritish Kamath, and Madhu Sudan. Communication complexity of permutation-invariant functions. In Robert Krauthgamer, editor, *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1902–1921. SIAM, 2016. doi:10.1137/1.9781611974331.ch134.

29 Badih Ghazi, Pritish Kamath, and Madhu Sudan. Decidability of non-interactive simulation of joint distributions. In Irit Dinur, editor, *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 545–554. IEEE Computer Society, 2016. doi:10.1109/FOCS.2016.65.

30 Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995. doi:10.1145/227683.227684.

31 Steven Heilman. Euclidean partitions optimizing noise stability. *CoRR*, abs/1211.7138, 2012. arXiv:1211.7138.

32 Steven Heilman, Elchanan Mossel, and Joe Neeman. Standard simplices and pluralities are not the most noise stable. *Israel Journal of Mathematics*, 213(1):33–53, 2016.

33 Hermann O Hirschfeld. A connection between correlation and contingency. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 31, pages 520–524. Cambridge Univ Press, 1935.

34 Thomas Holenstein. Parallel repetition: Simplification and the no-signaling case. *Theory of Computing*, 5(1):141–172, 2009. doi:10.4086/toc.2009.v005a008.

35 Marcus Isaksson and Elchanan Mossel. Maximally stable gaussian partitions with discrete applications. *Israel Journal of Mathematics*, 189(1):347–396, 2012.

36 William Johnson and Joram Lindenstrauss. Extensions of lipschitz maps into a hilbert space. 26:189–206, 01 1984.

37 Sudeep Kamath and Venkat Anantharam. On non-interactive simulation of joint distributions. *IEEE Trans. Information Theory*, 62(6):3419–3435, 2016. doi:10.1109/TIT.2016.2553672.

38 Julia Kempe, Hirotada Kobayashi, Keiji Matsumoto, Ben Toner, and Thomas Vidick. Entangled games are hard to approximate. *SIAM J. Comput.*, 40(3):848–877, 2011. doi:10.1137/090751293.

39 Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM Journal on Computing*, 37(1):319–357, 2007.

40 László Lovász. On the shannon capacity of a graph. *Information Theory, IEEE Transactions on*, 25(1):1–7, 1979.

41 Ueli M Maurer. Secret key agreement by public discussion from common information. *Information Theory, IEEE Transactions on*, 39(3):733–742, 1993.

42 Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010.

43 Elchanan Mossel and Ryan O'Donnell. Coin flipping from a cosmic source: On error correction of truly random bits. *arXiv preprint math/0406504*, 2004.

44 Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 21–30. IEEE, 2005.

45 Elchanan Mossel, Ryan O'Donnell, Oded Regev, Jeffrey E Steif, and Benny Sudakov. Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse bonami-beckner inequality. *Israel Journal of Mathematics*, 154(1):299–336, 2006.

46 Ilan Newman. Private vs. common random bits in communication complexity. *Information processing letters*, 39(2):67–71, 1991.

47 Michael A Nielsen. Conditions for a class of entanglement transformations. *Physical Review Letters*, 83(2):436, 1999.

48 Anup Rao. Parallel repetition in projection games and a concentration bound. *SIAM J. Comput.*, 40(6):1871–1891, 2011. doi:10.1137/080734042.

49 Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998. doi:10.1137/S0097539795280895.

50 Ran Raz. A counterexample to strong parallel repetition. *SIAM J. Comput.*, 40(3):771–777, 2011. doi:10.1137/090747270.

51 Renato Renner and Stefan Wolf. Simple and tight bounds for information reconciliation and privacy amplification. In *Advances in cryptology-ASIACRYPT 2005*, pages 199–216. Springer, 2005.

52 Alfréd Rényi. On measures of dependence. *Acta mathematica hungarica*, 10(3-4):441–451, 1959.

53 Claude E Shannon. The zero error capacity of a noisy channel. *Information Theory, IRE Transactions on*, 2(3):8–19, 1956.

54 Hans S Witsenhausen. On sequences of pairs of dependent random variables. *SIAM Journal on Applied Mathematics*, 28(1):100–113, 1975.

55 Pawel Wolff. Hypercontractivity of simple random variables. *Studia Mathematica*, 180(3):219–236, 2007.

56 Aaron D. Wyner. The common information of two dependent random variables. *IEEE Trans. Information Theory*, 21(2):163–179, 1975. doi:10.1109/TIT.1975.1055346.

A Proofs of Mean and Variance Bounds in Dimension Reduction

In this section, we provide the proof of Lemma 10. This is the main new technical component introduced in this paper. Even though the calculations might seem cumbersome, they involve mostly elementary steps. To understand the high level picture, we recommend the reader to go through a short proof sketch presented in subsection 3.1.

Recall that starting with degree d multilinear polynomials $A : \mathbb{R}^n \to \mathbb{R}$ and $B : \mathbb{R}^n \to \mathbb{R}$, we defined functions $A_{\boldsymbol{M}} : \mathbb{R}^{n_0} \to \mathbb{R}$ and $B_{\boldsymbol{M}} : \mathbb{R}^{n_0} \to \mathbb{R}$, for $\boldsymbol{M} \sim \gamma_1^{\otimes(n\times n_0)}$, as

$$A_{\boldsymbol{M}}(a) \;=\; A\left(\frac{\boldsymbol{M}a}{\|a\|_2}\right) \qquad \text{and} \qquad B_{\boldsymbol{M}}(\boldsymbol{b}) = B\left(\frac{\boldsymbol{M}b}{\|b\|_2}\right) \qquad \text{for } a, b \in \mathbb{R}^{n_0} \setminus \{0\}\ .$$

and we defined their correlation as $F(\boldsymbol{M}) \stackrel{\text{def}}{=} \langle A_{\boldsymbol{M}}, B_{\boldsymbol{M}} \rangle_{\mathcal{G}_\rho^{\otimes n_0}}$. Lemma 10 proves bounds on the mean and variance of $F(\boldsymbol{M})$, which we restate below for convenience.

▶ **Lemma 12.** (Mean & Variance Bound). *Given d and δ, there exists an explicitly computable $n_0 := n_0(d, \delta)$ such that for $\boldsymbol{M} \sim \gamma_1^{\otimes(n\times n_0)}$,*

$$\left|\mathop{\mathbb{E}}_{\boldsymbol{M}} F(\boldsymbol{M}) - \langle A, B\rangle_{\mathcal{G}_\rho^{\otimes n}}\right| \;\le\; \delta \qquad \textbf{(Mean bound)}$$

$$\mathop{\mathsf{Var}}_{\boldsymbol{M}}\left(F(\boldsymbol{M})\right) \;\le\; \delta \qquad \textbf{(Variance bound)}$$

In particular, one may take $n_0 = \frac{d^{O(d)}}{\delta^2}$.

We break down the full proof into the following three modular steps.

1. In subsection A.1, we prove a *meta-lemma* (Lemma 13) that will help us prove both the mean and variance bounds; indeed this meta-lemma is at the heart of why Theorem 8 holds. Morally, this lemma says that if we have an expectation of a product of a small number of inner products of normalized correlated Gaussian vectors, then, we can exchange the product and the expectations while incurring only a small additive error. Lemma 13 is the main take away from this subsection, and the reader may skip to subsection A.2 and subsection A.3 to see the rest of the proof.
2. In subsection A.2, we prove bounds on the mean and co-variances of degree-d multilinear monomials, under the above transformation of replacing $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^n$ (inputs to A and B) by $\frac{\boldsymbol{M}\boldsymbol{a}}{\|\boldsymbol{a}\|_2}$ and $\frac{\boldsymbol{M}\boldsymbol{b}}{\|\boldsymbol{b}\|_2}$ respectively.
3. In subsection A.3, we finally use the above bounds on mean and co-variances of degree-d multilinear monomials in order to prove Lemma 10.

▶ Remark. To make our notations convenient, we will often write equations such as $\alpha = \beta \pm \varepsilon$ which is to be interpreted as $\left|\alpha - \beta\right| \le \varepsilon$.

A.1 Product of Inner Products of Normalized Correlated Gaussian Vectors

The following is the main lemma in this subsection (this is the *meta-lemma* alluded to earlier).

▶ **Lemma 13.** *Given $d, D \in \mathbb{Z}_{\ge 0}$ and $\delta > 0$ (with D sufficiently larger than d), let $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_d)$ be a $2dD$-dimensional multivariate Gaussian distribution such that,*

- *each $\boldsymbol{u}_i, \boldsymbol{v}_i \in \mathbb{R}^D$ are marginally distributed as standard D-dimensional Gaussians γ_D.*
- *for each $j \in [D]$, the joint distribution $(\boldsymbol{u}_{1,j}, \ldots, \boldsymbol{u}_{d,j}, \boldsymbol{v}_{1,j}, \ldots, \boldsymbol{v}_{d,j})$, is independent across different values of j.*

Then,

$$\left| \mathop{\mathbb{E}}_{\{\boldsymbol{u}_i,\boldsymbol{v}_i\}_i} \left[\prod_{i=1}^{d} \frac{\langle \boldsymbol{u}_i,\boldsymbol{v}_i\rangle}{\|\boldsymbol{u}_i\|_2\|\boldsymbol{v}_i\|_2} \right] - \prod_{i=1}^{d} \mathop{\mathbb{E}}_{\{\boldsymbol{u}_i,\boldsymbol{v}_i\}_i} \left[\frac{\langle \boldsymbol{u}_i,\boldsymbol{v}_i\rangle}{D} \right] \right| \;\leq\; \frac{d^{O(d)}}{D}.$$

We point out that there are two steps taking place in Lemma 13:

(i) the replacement of $\|\boldsymbol{u}_i\|_2$ (and $\|\boldsymbol{v}_i\|_2$) by $\sqrt{D}$ (around which it is tightly concentrated),

(ii) the interchanging of the expectation and the product.

We will handle each of these changes one by one.

Product of Negative Moments of ℓ_2-norm of Correlated Gaussian vectors

In order to handle the replacement of $\|\boldsymbol{u}_i\|_2$ (and $\|\boldsymbol{v}_i\|_2$) by $\sqrt{D}$, we will prove some bounds on the mean and variance of products of negative powers of the ℓ_2-norm of a standard Gaussian vector.

▶ **Lemma 14.** *Let $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_\ell$ be (possibly correlated) multivariate Gaussians where each $\boldsymbol{w}_i \in \mathbb{R}^D$ is marginally distributed as γ_D, and let $d_1, d_2, \ldots, d_\ell$ be non-negative integers with $d := \sum_{i=1}^{\ell} d_i$. Then,*

$$\left| \mathbb{E}\left[\prod_{i=1}^{\ell} \frac{1}{\|\boldsymbol{w}_i\|_2^{d_i}} \right] - \frac{1}{D^{d/2}} \right| \;\leq\; O\left(\frac{d^5}{D^{\frac{d}{2}+1}}\right),$$

$$\mathsf{Var}\left[\prod_{i=1}^{\ell} \frac{1}{\|\boldsymbol{w}_i\|_2^{d_i}} \right] \;\leq\; O\left(\frac{d^5}{D^{d+1}}\right).$$

▶ Remark. It is conceivable that the bounds in Lemma 14 could be improved in terms of the dependence on d. However, this was not central to our application, so we go ahead with the stated bounds. The main point to note in the above lemma is the extra factor of D in the denominator.

We start out by first proving the base case where we have a single vector $\boldsymbol{w}$, that is, $\ell = 1$.

▶ **Proposition 15.** *There exists an absolute constant C such that for sufficiently large $d, D \in \mathbb{Z}_{>0}$ satisfying $D > Cd^2$, we have that for $\boldsymbol{w} \sim \gamma_D$,*

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{w}} \left[\frac{1}{\|\boldsymbol{w}\|_2^d} \right] - \frac{1}{D^{d/2}} \right| \;\leq\; C \cdot \left(\frac{d^2}{D^{\frac{d}{2}+1}}\right), \tag{5}$$

$$\mathop{\mathsf{Var}}_{\boldsymbol{w}} \left[\frac{1}{\|\boldsymbol{w}\|_2^d} \right] \;\leq\; 8C \cdot \left(\frac{d^2}{D^{d+1}}\right). \tag{6}$$

Proof. It is well-known that the distribution of $\|\boldsymbol{w}\|_2$ follows a χ-distribution with parameter D, and whose probability density function is given by

$$f_D(x) \;=\; \frac{x^{D-1} \cdot e^{-\frac{x^2}{2}}}{2^{\frac{D}{2}-1} \cdot \Gamma(\frac{D}{2})}, \qquad (x \in \mathbb{R}_{\geq 0})$$

where $\Gamma(\cdot)$ denotes the Gamma function. Thus, we have that

$$\mathop{\mathbb{E}}_{\boldsymbol{w}} \left[\frac{1}{\|\boldsymbol{w}\|^d} \right] \;=\; \int_0^\infty \frac{1}{x^d} \cdot f_D(x) dx \;=\; \int_0^\infty \frac{x^{D-d-1} \cdot e^{-\frac{x^2}{2}}}{2^{\frac{D}{2}-1} \cdot \Gamma(\frac{D}{2})} dx$$

$$= \frac{2^{\frac{D-d-1}{2}} \cdot \Gamma\left(\frac{D-d}{2}\right)}{2^{\frac{D}{2}-1} \cdot \Gamma(\frac{D}{2})} \;=\; \frac{1}{D^{d/2}} \cdot \left(1 \pm O\left(\frac{d^2}{D}\right)\right),$$

where the last equality follows from the Stirling's approximation of the Gamma function, which holds for every real number $z > 0$:

$$\Gamma(z+1) = \sqrt{2\pi z} \cdot \left(\frac{z}{e}\right)^z \cdot \left(1 \pm O\left(\frac{1}{z}\right)\right).$$

This completes the proof of Equation 5, for the explicit constant C that can be derived from the Stirling's approximation. Now, Equation 6 immediately follows as:

$$\begin{aligned}
\mathop{\mathsf{Var}}_{\boldsymbol{w}}\left[\frac{1}{\|\boldsymbol{w}\|^d}\right] &= \mathop{\mathbb{E}}_{\boldsymbol{w}}\left[\frac{1}{\|\boldsymbol{w}\|^{2d}}\right] - \mathop{\mathbb{E}}_{\boldsymbol{w}}\left[\frac{1}{\|\boldsymbol{w}\|^d}\right]^2 \\
&= \left(\frac{1}{D^d} \pm C \cdot \left(\frac{(2d)^2}{D^{d+1}}\right)\right) - \left(\frac{1}{D^{d/2}} \pm C \cdot \left(\frac{d^2}{D^{d/2+1}}\right)\right)^2 \\
&\leq 8C \cdot \left(\frac{d^2}{D^{d+1}}\right),
\end{aligned}$$

where, we use that D is sufficiently large that $C^2\left(\frac{d^4}{D^{d+2}}\right) < 2C \cdot \left(\frac{d^2}{D^{d+1}}\right)$, i.e. $D > Cd^2$. ◀

We now show how to generalize the above to prove Lemma 14.

Proof of Lemma 14. More specifically, we will show that,

$$\left|\mathbb{E}\left[\prod_{i=1}^{\ell} \frac{1}{\|\boldsymbol{w}_i\|_2^{d_i}}\right] - \frac{1}{D^{d/2}}\right| \leq C \cdot \ell^3 \cdot \left(\frac{d^2}{D^{\frac{d}{2}+1}}\right) \tag{7}$$

$$\mathsf{Var}\left[\prod_{i=1}^{\ell} \frac{1}{\|\boldsymbol{w}_i\|_2^{d_i}}\right] \leq 8C \cdot \ell^3 \cdot \left(\frac{d^2}{D^{d+1}}\right) \tag{8}$$

where C is the absolute constant (as obtained in Proposition 15). This implies the lemma since $\ell \leq d$.

We proceed by induction on ℓ (more specifically on $\log \ell$). For $\ell = 1$, the bound immediately follows from Proposition 15. For the inductive step, we assume that the bound in Equations 7 and 8 holds for ℓ, and we prove that the bound also holds for 2ℓ. While it may seem that our bounds are being proven only when ℓ is a power of 2, it is not hard to see that our proof could be done for non powers of 2 as well, giving a bound that is monotonically increasing in ℓ and hence it suffices having proved it for ℓ that are powers of 2. Let $d_1, d_2, \ldots, d_{2\ell}$ be non-negative integers with $d := \sum_{i=1}^{2\ell} d_i$. For notational convenience, let $s_1 = \sum_{i=1}^{\ell} d_i$ and $s_2 = \sum_{i=\ell+1}^{2\ell} d_i$, and so $d = s_1 + s_2$.

We will first prove Equation 7 inductively by using the following idea: for any two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$, we have $\mathbb{E}[\boldsymbol{XY}] = \mathbb{E}[\boldsymbol{X}]\,\mathbb{E}[\boldsymbol{Y}] + \mathsf{Cov}[\boldsymbol{X}, \boldsymbol{Y}]$ and $|\,\mathsf{Cov}[\boldsymbol{X}, \boldsymbol{Y}]| \leq \sqrt{\mathsf{Var}[\boldsymbol{X}] \cdot \mathsf{Var}[\boldsymbol{Y}]}$ and hence $\mathbb{E}[\boldsymbol{XY}] = \mathbb{E}[\boldsymbol{X}]\,\mathbb{E}[\boldsymbol{Y}] \pm \sqrt{\mathsf{Var}[\boldsymbol{X}] \cdot \mathsf{Var}[\boldsymbol{Y}]}$. Thus, we get,

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^{2\ell} \frac{1}{\|\boldsymbol{w}_i\|_2^{d_i}}\right] &= \mathbb{E}\left[\prod_{i=1}^{\ell} \frac{1}{\|\boldsymbol{w}_i\|_2^{d_i}}\right] \cdot \mathbb{E}\left[\prod_{i=\ell+1}^{2\ell} \frac{1}{\|\boldsymbol{w}_i\|_2^{d_i}}\right] \\
&\quad \pm \sqrt{\mathsf{Var}\left[\prod_{i=1}^{\ell} \frac{1}{\|\boldsymbol{w}_i\|_2^{d_i}}\right] \cdot \mathsf{Var}\left[\prod_{i=\ell+1}^{2\ell} \frac{1}{\|\boldsymbol{w}_i\|_2^{d_i}}\right]}.
\end{aligned} \tag{9}$$

Using the inductive assumption w.r.t. ℓ, we get that,

$$\mathbb{E}\left[\prod_{i=1}^{\ell}\frac{1}{\left\|\boldsymbol{w}_i\right\|_2^{d_i}}\right] = \frac{1}{D^{s_1/2}}\left(1 \pm C\cdot\ell^3\cdot\left(\frac{s_1^2}{D}\right)\right) \tag{10}$$

$$\mathbb{E}\left[\prod_{i=\ell+1}^{2\ell}\frac{1}{\left\|\boldsymbol{w}_i\right\|_2^{d_i}}\right] = \frac{1}{D^{s_2/2}}\left(1 \pm C\cdot\ell^3\cdot\left(\frac{s_2^2}{D}\right)\right) \tag{11}$$

and

$$\mathsf{Var}\left[\prod_{i=1}^{\ell}\frac{1}{\left\|\boldsymbol{w}_i\right\|_2^{d_i}}\right] \leq \frac{1}{D^{s_1}}\cdot 8C\cdot\ell^3\cdot\left(\frac{s_1^2}{D}\right) \tag{12}$$

$$\mathsf{Var}\left[\prod_{i=\ell+1}^{2\ell}\frac{1}{\left\|\boldsymbol{w}_i\right\|_2^{d_i}}\right] \leq \frac{1}{D^{s_2}}\cdot 8C\cdot\ell^3\cdot\left(\frac{s_2^2}{D}\right) \tag{13}$$

Plugging Equations 10, 11, 12 and 13 in Equation 9, it is not hard to see that,

$$\mathbb{E}\left[\prod_{i=1}^{2\ell}\frac{1}{\left\|\boldsymbol{w}_i\right\|_2^{d_i}}\right] = \frac{1}{D^{d/2}}\left(1 \pm C\cdot(2\ell)^3\cdot\left(\frac{d^2}{D}\right)\right) .$$

This completes the proof of Equation 7. Now, Equation 8 follows easily as,

$$\begin{aligned}\mathsf{Var}\left[\prod_{i=1}^{2\ell}\frac{1}{\left\|\boldsymbol{w}_i\right\|_2^{d_i}}\right] &= \mathbb{E}\left[\prod_{i=1}^{2\ell}\frac{1}{\left\|\boldsymbol{w}_i\right\|_2^{2d_i}}\right] - \mathop{\mathbb{E}}_{\boldsymbol{w}}\left[\prod_{i=1}^{2\ell}\frac{1}{\left\|\boldsymbol{w}_i\right\|_2^{d_i}}\right]^2 \\ &= \left(\frac{1}{D^d} \pm C\cdot(2\ell)^3\left(\frac{(2d)^2}{D^{d+1}}\right)\right) - \left(\frac{1}{D^{d/2}} \pm C\cdot(2\ell)^3\cdot\left(\frac{d^2}{D^{d/2+1}}\right)\right)^2 \\ &\leq 8C\cdot(2\ell)^3\cdot\left(\frac{d^2}{D^{d+1}}\right) .\end{aligned}$$

◀

Interchanging Product and Expectation

In order to handle the interchanging of the product and expectation operations, we will show the following lemma.

▶ **Lemma 16.** *Let $(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_d,\boldsymbol{v}_1,\ldots,\boldsymbol{v}_d)$ be distributed as in Lemma 13. Then,*

$$\left|\mathop{\mathbb{E}}_{\{\boldsymbol{u}_i,\boldsymbol{v}_i\}_i}\left[\prod_{i=1}^{d}\langle\boldsymbol{u}_i,\boldsymbol{v}_i\rangle\right] - \prod_{i=1}^{d}\mathop{\mathbb{E}}_{\{\boldsymbol{u}_i,\boldsymbol{v}_i\}_i}\left[\langle\boldsymbol{u}_i,\boldsymbol{v}_i\rangle\right]\right| \leq d^{O(d)}\cdot D^{d-1}.$$

▶ Remark. The $d^{O(d)}$ term has an explicit expression, although we only highlight its qualitative nature for clarity. Again, it is conceivable that the bounds in Lemma 16 could be improved in terms of the dependence on d, although we suspect that it is tight upto constant factors in the exponent. Anyhow, this was not central to our application, so we go ahead with the stated bounds. The main point to note in the above lemma is that the exponent of D is $(d-1)$ instead of d.

To prove the lemma, we first obtain the following proposition on moments of a multivariate Gaussian.

▶ **Proposition 17.** *Let $\boldsymbol{w} \in \mathbb{R}^\ell$ be any multivariate Gaussian vector with each coordinate marginally distributed according to γ_1. Let $d_1, d_2, \ldots, d_\ell$ be non-negative integers such that $d := \sum_{i=1}^{\ell} d_i$. Then,*

$$\left| \mathbb{E}\left[\prod_{i=1}^{\ell} \boldsymbol{w}_i^{d_i} \right] \right| \leq (2d)^{3d}.$$

Proof. More specifically we will show that when ℓ is a power of 2,

$$\left| \mathbb{E}\left[\prod_{i=1}^{\ell} \boldsymbol{w}_i^{d_i} \right] \right| \leq 2^{\ell-1}(\ell d)^d. \tag{14}$$

It is easy to see that this immediately implies the bound of $2^d \cdot d^{2d}$ in the main lemma, since $\ell \leq d$. However if ℓ is not a power of 2 we can round it up to the nearest power of 2, which amounts to substituting $\ell \leq 2d$ in the above, obtaining a bound of $2^{3d} \cdot d^{2d} \leq (2d)^{3d}$.
We proceed by induction on ℓ (more specifically on $\log \ell$). For $\ell = 1$, we use the well-known fact that for $w \sim \gamma_1$,

$$\left|\mathbb{E}[\boldsymbol{w}^d]\right| \;=\; \left\{ \begin{matrix} 0 & \text{if } d \text{ is odd} \\ (d-1)!! & \text{if } d \text{ is even} \end{matrix} \right\} \;\leq\; d^d,$$

where $(d-1)!!$ denotes the double factorial of $(d-1)$, i.e., the product of all integers from 1 to $d-1$ that have the same parity as $d-1$. For the inductive step, we assume that the bound in (14) holds for ℓ and we show that it also holds for 2ℓ. For notational convenience, let $s_1 = \sum_{i=1}^{\ell} d_i$ and $s_2 = \sum_{i=\ell+1}^{2\ell} d_i$, and so $d = s_1 + s_2$.

The main idea to prove the inductive step is simply the Cauchy-Schwarz inequality.

$$\begin{aligned} \left| \mathbb{E}\left[\prod_{i=1}^{2\ell} \boldsymbol{w}_i^{d_i} \right] \right| \;&\leq\; \sqrt{\mathbb{E}\left[\prod_{i=1}^{\ell} \boldsymbol{w}_i^{2d_i} \right] \cdot \mathbb{E}\left[\prod_{i=\ell+1}^{2\ell} \boldsymbol{w}_i^{2d_i} \right]} \\ &\leq\; \sqrt{2^{\ell-1}(2\ell s_1)^{2s_1} \cdot 2^{\ell-1}(2\ell s_2)^{2s_2}} \;\leq\; 2^{2\ell-1}(2\ell d)^d \,, \end{aligned}$$

where, we use the inductive assumption regarding product of ℓ terms and that $s_1 + s_2 = d$. ◀

Using the above proposition, we are now able to prove Lemma 16.

Proof of Lemma 16. Let $S \subseteq [D]^d$ be the set of all tuples $c \in [D]^d$ such that $c_j \neq c_k$ for all $j \neq k \in [d]$. Let $\overline{S}$ denote the complement of S in $[D]^d$. Note that $|\overline{S}| \leq d^2 \cdot D^{d-1}$. We have that

$$\begin{aligned} \mathbb{E}\left[\prod_{i=1}^{d} \langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle \right] \;&=\; \mathbb{E}\left[\prod_{i=1}^{d} \sum_{k=1}^{D} \boldsymbol{u}_{i,k} \boldsymbol{v}_{i,k} \right] \;=\; \sum_{c \in [D]^d} \mathbb{E}\left[\prod_{i=1}^{d} \boldsymbol{u}_{i,c_i} \boldsymbol{v}_{i,c_i} \right] \\ &=\; \sum_{c \in S} \mathbb{E}\left[\prod_{i=1}^{d} \boldsymbol{u}_{i,c_i} \boldsymbol{v}_{i,c_i} \right] \;+\; \sum_{c \in \overline{S}} \mathbb{E}\left[\prod_{i=1}^{d} \boldsymbol{u}_{i,c_i} \boldsymbol{v}_{i,c_i} \right] \\ &=\; \sum_{c \in S} \prod_{i=1}^{d} \mathbb{E}[\boldsymbol{u}_{i,c_i} \boldsymbol{v}_{i,c_i}] \;+\; \sum_{c \in \overline{S}} \mathbb{E}\left[\prod_{i=1}^{d} \boldsymbol{u}_{i,c_i} \boldsymbol{v}_{i,c_i} \right], \end{aligned} \tag{15}$$

where the last equality follows from the assumption that the distribution of the j-th coordinates $(\boldsymbol{u}_{1,j}, \ldots, \boldsymbol{u}_{d,j}, \boldsymbol{v}_{1,j}, \ldots, \boldsymbol{v}_{d,j})$ is independent across $j \in [D]$. On the other hand, we have

that

$$\prod_{i=1}^{d} \mathbb{E}[\langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle] \;=\; \prod_{i=1}^{d} \mathbb{E}\left[\sum_{k=1}^{D} \boldsymbol{u}_{i,k}\boldsymbol{v}_{i,k}\right] \;=\; \sum_{c\in[D]^d} \prod_{i=1}^{d} \mathbb{E}[\boldsymbol{u}_{i,c_i}\boldsymbol{v}_{i,c_i}]$$
$$= \sum_{c\in S} \prod_{i=1}^{d} \mathbb{E}[\boldsymbol{u}_{i,c_i}\boldsymbol{v}_{i,c_i}] \;+\; \sum_{c\in \overline{S}} \prod_{i=1}^{d} \mathbb{E}[\boldsymbol{u}_{i,c_i}\boldsymbol{v}_{i,c_i}] \tag{16}$$

Combining Equations 15 and 16, we get

$$\left| \mathbb{E}\left[\prod_{i=1}^{d} \langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle\right] - \prod_{i=1}^{d} \mathbb{E}[\langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle] \right| \;=\; \left| \sum_{c\in\overline{S}} \left(\mathbb{E}\left[\prod_{i=1}^{d} \boldsymbol{u}_{i,c_i}\boldsymbol{v}_{i,c_i}\right] - \prod_{i=1}^{d} \mathbb{E}[\boldsymbol{u}_{i,c_i}\boldsymbol{v}_{i,c_i}] \right) \right|$$
$$\leq\; |\overline{S}| \cdot \max_{c\in\overline{S}} \left| \mathbb{E}\left[\prod_{i=1}^{d} \boldsymbol{u}_{i,c_i}\boldsymbol{v}_{i,c_i}\right] - \prod_{i=1}^{d} \mathbb{E}[\boldsymbol{u}_{i,c_i}\boldsymbol{v}_{i,c_i}] \right|$$
$$\leq\; d^2 \cdot D^{d-1} \cdot \left((2d)^{3d} + 1\right) \;\leq\; d^{O(d)} \cdot D^{d-1},$$

where the second last inequality follows from the fact that $|\overline{S}| \leq d^2 \cdot D^{d-1}$ and from Proposition 17. ◀

Putting things together to prove Lemma 13

Proof of Lemma 13. We show the following bounds, which immediately imply Lemma 13.

$$\left| \mathop{\mathbb{E}}_{\{\boldsymbol{u}_i,\boldsymbol{v}_i\}} \left[\prod_{i=1}^{d} \frac{\langle \boldsymbol{u}_i,\boldsymbol{v}_i\rangle}{\|\boldsymbol{u}_i\|_2\|\boldsymbol{v}_i\|_2}\right] - \mathop{\mathbb{E}}_{\{\boldsymbol{u}_i,\boldsymbol{v}_i\}} \left[\prod_{i=1}^{d} \frac{\langle \boldsymbol{u}_i,\boldsymbol{v}_i\rangle}{D}\right] \right| \;\leq\; \frac{d^{O(d)}}{D}. \tag{17}$$

$$\left| \mathop{\mathbb{E}}_{\{\boldsymbol{u}_i,\boldsymbol{v}_i\}} \left[\prod_{i=1}^{d} \frac{\langle \boldsymbol{u}_i,\boldsymbol{v}_i\rangle}{D}\right] - \prod_{i=1}^{d} \mathop{\mathbb{E}}_{\{\boldsymbol{u}_i,\boldsymbol{v}_i\}} \left[\frac{\langle \boldsymbol{u}_i,\boldsymbol{v}_i\rangle}{D}\right] \right| \;\leq\; \frac{d^{O(d)}}{D}. \tag{18}$$

Note that Equation 18 is simply a restatement of Lemma 16. To prove Equation 17, we define the random variables

$$\boldsymbol{W} \;:=\; \prod_{i=1}^{d} \langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle \qquad \text{and} \qquad \boldsymbol{Z} \;:=\; \prod_{i=1}^{d} \frac{1}{\|\boldsymbol{u}_i\|_2 \|\boldsymbol{v}_i\|_2} \;-\; \frac{1}{D^d}\,.$$

Note that Equation 17 is equivalent to showing bounds on $|\,\mathbb{E}[\boldsymbol{W}\cdot\boldsymbol{Z}]|$. In order to do so, we use the following four bounds:

1. $\left|\mathbb{E}[\boldsymbol{W}]\right| \leq D^d + d^{O(d)} \cdot D^{d-1}$. Since, by Lemma 16, we have that

$$|\,\mathbb{E}[\boldsymbol{W}]| \;\leq\; \left| \prod_{i=1}^{d} \mathbb{E}[\langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle] \right| + d^{O(d)} \cdot D^{d-1} \;\leq\; D^d + d^{O(d)} \cdot D^{d-1}$$

2. $\mathsf{Var}[\boldsymbol{W}] \leq d^{O(d)} \cdot D^{2d-1}$. Since,

$$\mathsf{Var}[\boldsymbol{W}] \;=\; \mathbb{E}[\boldsymbol{W}^2] - [\mathbb{E}\,\boldsymbol{W}]^2$$
$$=\; \mathbb{E}\left[\prod_{i=1}^{d} \langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle^2\right] - \left[\mathbb{E}\prod_{i=1}^{d} \langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle\right]^2$$
$$\leq\; d^{O(d)} \cdot D^{2d-1} \qquad \ldots \text{(from Lemma 16)}$$

3. $\left|\mathbb{E}[\boldsymbol{Z}]\right| = O\left(\frac{d^5}{D^{d+1}}\right)$ (follows exactly from Lemma 14).

4. $\mathsf{Var}[\boldsymbol{Z}] = O\left(\frac{d^5}{D^{2d+1}}\right)$ (follows exactly from Lemma 14).

Thus, we can bound $|\mathbb{E}[\boldsymbol{W}\cdot\boldsymbol{Z}]|$ as,

$$\left|\mathbb{E}[\boldsymbol{W}\cdot\boldsymbol{Z}]\right| \;\le\; \left|\mathbb{E}[\boldsymbol{W}]\right|\cdot\left|\mathbb{E}[\boldsymbol{Z}]\right| + \sqrt{\mathsf{Var}[\boldsymbol{W}]\cdot\mathsf{Var}[\boldsymbol{Z}]} \;\le\; \frac{d^{O(d)}}{D}\,.$$

This completes the proof of Equation 17 and hence of Lemma 13. ◀

A.2 Mean & Variance Bounds for Multilinear Monomials

For the rest of this section, we simplify our notations as follows:

- For $(\boldsymbol{a},\boldsymbol{b})\sim\mathcal{G}_\rho^{\otimes n_0}$, we will use $\widetilde{\boldsymbol{a}}$ and $\widetilde{\boldsymbol{b}}$ to denote the normalized vectors $\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2}$ and $\frac{\boldsymbol{b}}{\|\boldsymbol{b}\|_2}$ respectively.
- We will use $\boldsymbol{U}\in\mathbb{R}^n$ to denote $\boldsymbol{M}\widetilde{\boldsymbol{a}}$ and similarly $\boldsymbol{V}\in\mathbb{R}^n$ to denote $\boldsymbol{M}\widetilde{\boldsymbol{b}}$. We will also have independent variables $(\boldsymbol{a}',\boldsymbol{b}')\sim\mathcal{G}_\rho^{\otimes n_0}$, for which we use $\boldsymbol{U}'=\boldsymbol{M}\widetilde{\boldsymbol{a}}'$ and $\boldsymbol{V}'=\boldsymbol{M}\widetilde{\boldsymbol{b}}'$.
- $\boldsymbol{U}_i$ denotes the i-th coordinate of $\boldsymbol{U}$. Similarly, $\boldsymbol{m}_i\in\mathbb{R}^{n_0}$ is the i-th row of $\boldsymbol{M}$. Note that $\boldsymbol{U}_i=\langle\boldsymbol{m}_i,\widetilde{\boldsymbol{a}}\rangle$. For $S\subseteq[n]$, let $\boldsymbol{U}_S$ denote $\prod_{i\in S}\boldsymbol{U}_i=\prod_{i\in S}\langle\boldsymbol{m}_i,\widetilde{\boldsymbol{a}}\rangle$. Similarly for $\boldsymbol{V}_S$.
- We will take expectations over random variables $\boldsymbol{M}$, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{a}'$, $\boldsymbol{b}'$. It will be understood that we are sampling $M\sim\gamma_1^{\otimes(n\times n_0)}$. Also, $(\boldsymbol{a},\boldsymbol{b})$ and $(\boldsymbol{a}',\boldsymbol{b}')$ are independently sampled from $\mathcal{G}_\rho^{\otimes n_0}$.

▶ **Lemma 18** (Mean bounds for monomials)**.** *Given parameter d and δ, there exists an explicitly computable $n_0:=n_0(d,\delta)$ such that the following holds: For any subsets $S,T\subseteq[n]$ satisfying $|S|,|T|\le d$, it holds that,*

$$\mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}\boldsymbol{U}_S\boldsymbol{V}_T \;=\; \begin{cases}0 & \text{if } S\neq T\\ \rho^{|S|}\pm\delta & \text{if } S=T\end{cases}\,.$$

In particular, one may take $n_0=\frac{d^{O(d)}}{\delta}$.

Proof. We have that

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}\boldsymbol{U}_S\boldsymbol{V}_T &= \mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}\left[\prod_{i\in S}\boldsymbol{U}_i\cdot\prod_{i\in T}\boldsymbol{V}_i\right]\\
&= \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}\mathop{\mathbb{E}}_{\boldsymbol{M}}\left[\prod_{i\in S\cap T}\boldsymbol{U}_i\boldsymbol{V}_i\cdot\prod_{i\in S\setminus T}\boldsymbol{U}_i\cdot\prod_{i\in T\setminus S}\boldsymbol{V}_i\right]\\
&= \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}\mathop{\mathbb{E}}_{\boldsymbol{M}}\left[\prod_{i\in S\cap T}\langle\boldsymbol{m}_i,\widetilde{\boldsymbol{a}}\rangle\left\langle\boldsymbol{m}_i,\widetilde{\boldsymbol{b}}\right\rangle\cdot\prod_{i\in S\setminus T}\langle\boldsymbol{m}_i,\widetilde{\boldsymbol{a}}\rangle\cdot\prod_{i\in T\setminus S}\left\langle\boldsymbol{m}_i,\widetilde{\boldsymbol{b}}\right\rangle\right]\\
\mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}\boldsymbol{U}_S\boldsymbol{V}_T &= \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}\left[\prod_{i\in S\cap T}\mathop{\mathbb{E}}_{\boldsymbol{m}_i}\langle\boldsymbol{m}_i,\widetilde{\boldsymbol{a}}\rangle\left\langle\boldsymbol{m}_i,\widetilde{\boldsymbol{b}}\right\rangle\cdot\prod_{i\in S\setminus T}\mathop{\mathbb{E}}_{\boldsymbol{m}_i}\langle\boldsymbol{m}_i,\widetilde{\boldsymbol{a}}\rangle\cdot\prod_{i\in T\setminus S}\mathop{\mathbb{E}}_{\boldsymbol{m}_i}\left\langle\boldsymbol{m}_i,\widetilde{\boldsymbol{b}}\right\rangle\right], \qquad (19)
\end{aligned}$$

where the last equality follows from the independence of the $\boldsymbol{m}_i$'s.

If $S\neq T$, one of $\prod_{i\in S\setminus T}\mathbb{E}_{\boldsymbol{m}_i}[\langle\boldsymbol{m}_i,\widetilde{\boldsymbol{a}}\rangle]$ or $\prod_{i\in T\setminus S}\mathbb{E}_{\boldsymbol{m}_i}[\left\langle\boldsymbol{m}_i,\widetilde{\boldsymbol{b}}\right\rangle]$ is 0. This is because even for any fixed vector a and for each $i\in[n]$, the random variable $\langle\boldsymbol{m}_i,\widetilde{a}\rangle$ has zero-mean (and similarly for $\left\langle\boldsymbol{m}_i,\widetilde{b}\right\rangle$). The first part of the lemma now follows from Equation 19.

If $S = T$, Equation 19 becomes

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \boldsymbol{U}_S\boldsymbol{V}_T &= \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}\left[\prod_{i\in S}\mathop{\mathbb{E}}_{\boldsymbol{m}_i}\frac{\langle \boldsymbol{m}_i,\boldsymbol{a}\rangle}{\|\boldsymbol{a}\|_2}\frac{\langle \boldsymbol{m}_i,\boldsymbol{b}\rangle}{\|\boldsymbol{b}\|_2}\right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}\left[\prod_{i\in S}\frac{\langle \boldsymbol{a},\boldsymbol{b}\rangle}{\|\boldsymbol{a}\|_2\|\boldsymbol{b}\|_2}\right] \qquad \left[\text{since } \mathop{\mathbb{E}}_{\boldsymbol{m}_i}\boldsymbol{m}_i\cdot\boldsymbol{m}_i^T = I_{n_0\times n_0}.\right] \\
&= \prod_{i\in S}\left[\frac{\mathbb{E}_{\boldsymbol{a},\boldsymbol{b}}\langle \boldsymbol{a},\boldsymbol{b}\rangle}{n_0}\right]\pm\delta \qquad \left[\text{from Lemma 13, for } n_0 = \frac{d^{O(d)}}{\delta}\right] \\
&= \rho^{|S|}\pm\delta.
\end{aligned}$$

◀

▶ **Lemma 19** (Covariance bounds for monomials)**.** *Given parameters d and δ, there exists an explicitly computable $n_0 := n_0(d,\delta)$ such that the following holds: For any subsets $S,T,S',T'\subseteq[n]$ satisfying $|S|,|T|,|S'|,|T'|\le d$, it holds that,*

$$\left|\mathop{\mathbb{E}}_{\substack{\boldsymbol{M}\\ \boldsymbol{a},\boldsymbol{b},\boldsymbol{a}',\boldsymbol{b}'}}[\boldsymbol{U}_S\boldsymbol{V}_T\boldsymbol{U}'_{S'}\boldsymbol{V}'_{T'}] - \left(\mathop{\mathbb{E}}_{\substack{\boldsymbol{M}\\ \boldsymbol{a},\boldsymbol{b}}}[\boldsymbol{U}_S\boldsymbol{V}_T]\right)\cdot\left(\mathop{\mathbb{E}}_{\substack{\boldsymbol{M}\\ \boldsymbol{a}',\boldsymbol{b}'}}[\boldsymbol{U}'_{S'}\boldsymbol{V}'_{T'}]\right)\right| \begin{cases} = 0 & \text{if } S\triangle T\triangle S'\triangle T'\neq\emptyset \\ \le\delta & \text{if } S\triangle T\triangle S'\triangle T' = \emptyset\end{cases}$$

Here, $S\triangle T\triangle S'\triangle T'$ is the symmetric difference of the sets S,T,S',T', equivalently, the set of all $i\in[n]$ which appear an odd number of times in the multiset $S\sqcup T\sqcup S'\sqcup T'$.
In particular, one may take $n_0 = \frac{d^{O(d)}}{\delta^2}$.

In order to prove Lemma 19, we need the following lemma.

▶ **Lemma 20.** *For $\boldsymbol{m}\sim\gamma_{n_0}$,*

$$\mathop{\mathbb{E}}_{\substack{\boldsymbol{a},\boldsymbol{b},\\ \boldsymbol{a}',\boldsymbol{b}'}}\left[\left(\mathop{\mathbb{E}}_{\boldsymbol{m}}[\langle\boldsymbol{m},\widetilde{\boldsymbol{a}}\rangle\langle\boldsymbol{m},\widetilde{\boldsymbol{b}}\rangle\langle\boldsymbol{m},\widetilde{\boldsymbol{a}}'\rangle\langle\boldsymbol{m},\widetilde{\boldsymbol{b}}'\rangle] - \mathop{\mathbb{E}}_{\boldsymbol{m}}[\langle\boldsymbol{m},\widetilde{\boldsymbol{a}}\rangle\langle\boldsymbol{m},\widetilde{\boldsymbol{b}}\rangle]\cdot\mathop{\mathbb{E}}_{\boldsymbol{m}}[\langle\boldsymbol{m},\widetilde{\boldsymbol{a}}'\rangle\langle\boldsymbol{m},\widetilde{\boldsymbol{b}}'\rangle]\right)^2\right] \le O\left(\frac{1}{n_0}\right)$$

and

$$\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{a}'}\left[\left(\mathop{\mathbb{E}}_{\boldsymbol{m}}[\langle\boldsymbol{m},\widetilde{\boldsymbol{a}}\rangle\,\langle\boldsymbol{m},\widetilde{\boldsymbol{a}}'\rangle] - \mathop{\mathbb{E}}_{\boldsymbol{m}}[\langle\boldsymbol{m},\widetilde{\boldsymbol{a}}\rangle]\cdot\mathop{\mathbb{E}}_{\boldsymbol{m}}[\langle\boldsymbol{m},\widetilde{\boldsymbol{a}}'\rangle]\right)^2\right] \;\le\; O\left(\frac{1}{n_0}\right).$$

Proof. To prove the first part of the lemma, consider the quantity

$$\begin{aligned}
&T(\boldsymbol{a},\boldsymbol{b},\boldsymbol{a}',\boldsymbol{b}') \\
&:= \mathop{\mathbb{E}}_{\boldsymbol{m}}[\langle\boldsymbol{m},\widetilde{\boldsymbol{a}}\rangle\langle\boldsymbol{m},\widetilde{\boldsymbol{b}}\rangle\langle\boldsymbol{m},\widetilde{\boldsymbol{a}}'\rangle\langle\boldsymbol{m},\widetilde{\boldsymbol{b}}'\rangle] - \mathop{\mathbb{E}}_{\boldsymbol{m}}[\langle\boldsymbol{m},\widetilde{\boldsymbol{a}}\rangle\langle\boldsymbol{m},\widetilde{\boldsymbol{b}}\rangle]\cdot\mathop{\mathbb{E}}_{\boldsymbol{m}}[\langle\boldsymbol{m},\widetilde{\boldsymbol{a}}'\rangle\langle\boldsymbol{m},\widetilde{\boldsymbol{b}}'\rangle] \\
&= \langle\widetilde{\boldsymbol{a}},\widetilde{\boldsymbol{b}}\rangle\langle\widetilde{\boldsymbol{a}}',\widetilde{\boldsymbol{b}}'\rangle + \langle\widetilde{\boldsymbol{a}},\widetilde{\boldsymbol{a}}'\rangle\langle\widetilde{\boldsymbol{b}},\widetilde{\boldsymbol{b}}'\rangle + \langle\widetilde{\boldsymbol{a}},\widetilde{\boldsymbol{b}}'\rangle\langle\widetilde{\boldsymbol{a}}',\widetilde{\boldsymbol{b}}\rangle - \langle\widetilde{\boldsymbol{a}},\widetilde{\boldsymbol{b}}\rangle\langle\widetilde{\boldsymbol{a}}',\widetilde{\boldsymbol{b}}'\rangle \\
&= \langle\widetilde{\boldsymbol{a}},\widetilde{\boldsymbol{a}}'\rangle\langle\widetilde{\boldsymbol{b}},\widetilde{\boldsymbol{b}}'\rangle + \langle\widetilde{\boldsymbol{a}},\widetilde{\boldsymbol{b}}'\rangle\langle\widetilde{\boldsymbol{a}}',\widetilde{\boldsymbol{b}}\rangle.
\end{aligned}$$

where we use that for any $j\in[n_0]$, it holds that $\mathbb{E}_{\boldsymbol{m}}[\boldsymbol{m}_j^4] = 3$ and $\mathbb{E}_{\boldsymbol{m}}[\boldsymbol{m}_j^2] = 1$. Thus,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{a}',\boldsymbol{b}'}\left[T(\boldsymbol{a},\boldsymbol{b},\boldsymbol{a}',\boldsymbol{b}')^2\right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{a}',\boldsymbol{b}'}\left[\left[\langle\widetilde{\boldsymbol{a}},\widetilde{\boldsymbol{a}}'\rangle\langle\widetilde{\boldsymbol{b}},\widetilde{\boldsymbol{b}}'\rangle + \langle\widetilde{\boldsymbol{a}},\widetilde{\boldsymbol{b}}'\rangle\langle\widetilde{\boldsymbol{a}}',\widetilde{\boldsymbol{b}}\rangle\right]^2\right] \\
&\le 2\cdot\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{a}',\boldsymbol{b}'}\left[\langle\widetilde{\boldsymbol{a}},\widetilde{\boldsymbol{a}}'\rangle^2\langle\widetilde{\boldsymbol{b}},\widetilde{\boldsymbol{b}}'\rangle^2\right] \;+\; 2\cdot\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{a}',\boldsymbol{b}'}\left[\langle\widetilde{\boldsymbol{a}},\widetilde{\boldsymbol{b}}'\rangle^2\langle\widetilde{\boldsymbol{a}}',\widetilde{\boldsymbol{b}}\rangle^2\right] \\
&\le O\left(\frac{1}{n_0}\right),
\end{aligned}$$

where the last step follows by two applications of Lemma 13 (with $d = 4$). This completes the proof of the first part of the lemma. The second part of the lemma similarly follows from Lemma 13 (with $d = 2$) along with the fact that $\mathbb{E}_{\boldsymbol{m}}[\langle \boldsymbol{m}, \widetilde{\boldsymbol{a}} \rangle] = 0$. ◀

Proof of Lemma 19. Let $\mathbf{1}(E)$ denote the 0/1 indicator function of an event E. We have that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\boldsymbol{M}} \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'} [\boldsymbol{U}_S \boldsymbol{V}_T \boldsymbol{U}'_{S'} \boldsymbol{V}'_{T'}] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{M}} \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'} \left[\prod_{i \in S \cup T \cup S' \cup T'} \boldsymbol{U}_i^{\mathbf{1}(i \in S)} \boldsymbol{V}_i^{\mathbf{1}(i \in T)} {U'}_i^{\mathbf{1}(i \in S')} {V'}_i^{\mathbf{1}(i \in T')} \right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'} \left[\prod_{i \in S \cup T \cup S' \cup T'} \mathop{\mathbb{E}}_{\boldsymbol{m}_i} \left[\boldsymbol{U}_i^{\mathbf{1}(i \in S)} \boldsymbol{V}_i^{\mathbf{1}(i \in T)} {U'}_i^{\mathbf{1}(i \in S')} {V'}_i^{\mathbf{1}(i \in T')} \right] \right]. \qquad (20)
\end{aligned}$$

On the other hand, we have that

$$\mathop{\mathbb{E}}_{\boldsymbol{M}} \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} [\boldsymbol{U}_S \boldsymbol{V}_T] = \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \mathop{\mathbb{E}}_{\boldsymbol{M}} \left[\prod_{i \in S \cup T} \boldsymbol{U}_i^{\mathbf{1}(i \in S)} \boldsymbol{V}_i^{\mathbf{1}(i \in T)} \right] = \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \left[\prod_{i \in S \cup T} \mathop{\mathbb{E}}_{\boldsymbol{m}_i} \left[\boldsymbol{U}_i^{\mathbf{1}(i \in S)} \boldsymbol{V}_i^{\mathbf{1}(i \in T)} \right] \right], \qquad (21)$$

$$\text{and similarly,} \qquad \mathop{\mathbb{E}}_{\boldsymbol{M}} \mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'} [\boldsymbol{U}'_{S'} \boldsymbol{V}'_{T'}] = \mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'} \left[\prod_{i \in S' \cup T'} \mathop{\mathbb{E}}_{\boldsymbol{m}_i} \left[\boldsymbol{U}_i^{\mathbf{1}(i \in S')} \boldsymbol{V}_i^{\mathbf{1}(i \in T')} \right] \right]. \qquad (22)$$

If there exists $i \in S \cup T \cup S' \cup T'$ that appears in an odd number of S, T, S' and T', then it can be seen that the expectation in Equation 20 is equal to 0, and that at least one of the expectations in Equations 21 and 22 is equal to 0. This already handles the case that $S \triangle T \triangle S' \triangle T' \neq \emptyset$.

Henceforth, we assume that each $i \in S \cup T \cup S' \cup T'$ appears in an even number of S, T, S' and T'. Assume for ease of notation that $S \cup T \cup S' \cup T' \subseteq [4d]$. Define

$$\begin{aligned}
g_i(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{a}', \boldsymbol{b}') &:= \mathop{\mathbb{E}}_{\boldsymbol{m}_i} \left[\boldsymbol{U}_i^{\mathbf{1}(i \in S)} \boldsymbol{V}_i^{\mathbf{1}(i \in T)} {\boldsymbol{U}'}_i^{\mathbf{1}(i \in S')} {\boldsymbol{V}'}_i^{\mathbf{1}(i \in T')} \right] \qquad (23) \\
h_i(\boldsymbol{a}, \boldsymbol{b}) &:= \mathop{\mathbb{E}}_{\boldsymbol{m}_i} \left[\boldsymbol{U}_i^{\mathbf{1}(i \in S)} \boldsymbol{V}_i^{\mathbf{1}(i \in T)} \right]. \qquad (24) \\
h'_i(\boldsymbol{a}', \boldsymbol{b}') &:= \mathop{\mathbb{E}}_{\boldsymbol{m}_i} \left[{\boldsymbol{U}'}_i^{\mathbf{1}(i \in S')} {\boldsymbol{V}'}_i^{\mathbf{1}(i \in T')} \right]. \qquad (25)
\end{aligned}$$

Combining Equations 20, 21 and 22 along with the definitions in 23, 24 and 25, we get

$$\begin{aligned}
&\left| \mathop{\mathbb{E}}_{\boldsymbol{M}} \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'} [\boldsymbol{U}_S \boldsymbol{V}_T \boldsymbol{U}'_{S'} \boldsymbol{V}'_{T'}] - \mathop{\mathbb{E}}_{\boldsymbol{M}} \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} [\boldsymbol{U}_S \boldsymbol{V}_T] \cdot \mathop{\mathbb{E}}_{\boldsymbol{M}} \mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'} [\boldsymbol{U}'_{S'} \boldsymbol{V}'_{T'}] \right| \\
&= \left| \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'} \left[\prod_{i=1}^{4d} g_i(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{a}', \boldsymbol{b}') - \prod_{i=1}^{4d} h_i(\boldsymbol{a}, \boldsymbol{b}) \cdot h'_i(\boldsymbol{a}', \boldsymbol{b}') \right] \right| \\
&= \left| \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'} \left[\sum_{j=1}^{4d} \begin{bmatrix} \prod_{i=1}^{j-1} h_i(\boldsymbol{a}, \boldsymbol{b}) \cdot h'_i(\boldsymbol{a}', \boldsymbol{b}') \prod_{i=j}^{4d} g_i(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{a}', \boldsymbol{b}') \\ - \prod_{i=1}^{j} h_i(\boldsymbol{a}, \boldsymbol{b}) \cdot h'_i(\boldsymbol{a}', \boldsymbol{b}') \prod_{i=j+1}^{4d} g_i(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{a}', \boldsymbol{b}') \end{bmatrix} \right] \right| \\
&\leq \sum_{j=1}^{4d} \left| \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'} \left[\prod_{i=1}^{j-1} h_i(\boldsymbol{a}, \boldsymbol{b}) \cdot h'_i(\boldsymbol{a}', \boldsymbol{b}') \prod_{i=j+1}^{4d} g_i(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{a}', \boldsymbol{b}') \cdot \begin{bmatrix} g_j(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{a}', \boldsymbol{b}') \\ -h_j(\boldsymbol{a}, \boldsymbol{b}) \cdot h'_j(\boldsymbol{a}', \boldsymbol{b}') \end{bmatrix} \right] \right| \\
&\leq 4 \cdot d \cdot \sqrt{\tau \cdot \kappa},
\end{aligned}$$

where the last inequality follows from the Cauchy-Schwarz inequality with

$$\tau := \max_{j\in[4d]} \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'} \left[\prod_{i=1}^{j-1} h_i(\boldsymbol{a},\boldsymbol{b})^2 \cdot h_i(\boldsymbol{a}',\boldsymbol{b}')^2 \prod_{i=j+1}^{4d} g_i(\boldsymbol{a},\boldsymbol{b},\boldsymbol{a}',\boldsymbol{b}')^2\right]$$
$$\kappa := \max_{j\in[4d]} \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} \mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'} \left[g_j(\boldsymbol{a},\boldsymbol{b},\boldsymbol{a}',\boldsymbol{b}') - h_j(\boldsymbol{a},\boldsymbol{b})\cdot h_j(\boldsymbol{a}',\boldsymbol{b}')\right]^2$$

Lemma 20 implies that $\kappa \le O(1/n_0)$. We now show that $\tau \le 2^{O(d)}$. Note that for any $i \in [n_0]$, it holds that,

$$h_i(\boldsymbol{a},\boldsymbol{b}) = \begin{cases} \langle \widetilde{\boldsymbol{a}}, \widetilde{\boldsymbol{b}}\rangle & \text{if } i \in S \text{ and } i \in T \\ 1 & \text{if } i \notin S \text{ and } i \notin T \\ 0 & \text{otherwise} \end{cases}$$

$$h'_i(\boldsymbol{a}',\boldsymbol{b}') = \begin{cases} \langle \widetilde{\boldsymbol{a}}', \widetilde{\boldsymbol{b}}'\rangle & \text{if } i \in S' \text{ and } i \in T' \\ 1 & \text{if } i \notin S' \text{ and } i \notin T' \\ 0 & \text{otherwise} \end{cases}$$

$$g_i(\boldsymbol{a},\boldsymbol{b},\boldsymbol{a}',\boldsymbol{b}') = \begin{cases} \langle \widetilde{\boldsymbol{a}}, \widetilde{\boldsymbol{b}}\rangle\langle \widetilde{\boldsymbol{a}}', \widetilde{\boldsymbol{b}}'\rangle + \langle \widetilde{\boldsymbol{a}}, \widetilde{\boldsymbol{a}}'\rangle\langle \widetilde{\boldsymbol{b}}, \widetilde{\boldsymbol{b}}'\rangle + \langle \widetilde{\boldsymbol{a}}, \widetilde{\boldsymbol{b}}'\rangle\langle \widetilde{\boldsymbol{a}}', \widetilde{\boldsymbol{b}}\rangle & \text{if } i \in S\cap T\cap S'\cap T' \\ \langle \widetilde{\boldsymbol{a}}, \widetilde{\boldsymbol{b}}\rangle & \text{if } i \in S\cap T, i \notin S'\cup T' \\ \langle \widetilde{\boldsymbol{a}}, \widetilde{\boldsymbol{a}}'\rangle & \text{if } i \in S\cap S', i \notin T\cup T' \\ \langle \widetilde{\boldsymbol{a}}, \widetilde{\boldsymbol{b}}'\rangle & \text{if } i \in S\cap T', i \notin S'\cup T \\ \langle \widetilde{\boldsymbol{a}}', \widetilde{\boldsymbol{b}}\rangle & \text{if } i \in S'\cap T, i \notin S\cup T' \\ \langle \widetilde{\boldsymbol{a}}', \widetilde{\boldsymbol{b}}'\rangle & \text{if } i \in S'\cap T', i \notin S\cup T \\ \langle \widetilde{\boldsymbol{b}}, \widetilde{\boldsymbol{b}}'\rangle & \text{if } i \in T\cap T', i \notin S\cup S' \\ 1 & \text{otherwise} \end{cases}$$

Thus, if we expand out a single term $\prod_{i=1}^{j-1} h_i(\boldsymbol{a},\boldsymbol{b})^2 \cdot h_i(\boldsymbol{a}',\boldsymbol{b}')^2 \prod_{i=j+1}^{4d} g_i(\boldsymbol{a},\boldsymbol{b},\boldsymbol{a}',\boldsymbol{b}')^2$, we get at most 3^{8d} terms (since each g_i can increase the number of terms by a factor of at most 3). Each of these terms is the expectation of the product of inner product of some correlated Gaussian vectors. We have from Lemma 13 that each such term is at most $1+\delta$ and thus $\tau \le 2^{O(d)}$. Thus, for an explicit choice of n_0 that is upper bounded by $d^{O(d)}/\delta^2$, we get that $4d\sqrt{\tau\kappa} \le \delta$, which concludes the proof of the lemma. ◀

A.3 Mean & Variance Bounds for Multilinear Polynomials

We are now ready to prove Lemma 10. Recall again that,

$$F(M) = \mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} [A(\boldsymbol{U})\cdot B(\boldsymbol{V})] \qquad \text{where, } \boldsymbol{U} = \frac{M\boldsymbol{a}}{\|\boldsymbol{a}\|_2} \text{ and } \boldsymbol{V} = \frac{M\boldsymbol{b}}{\|\boldsymbol{b}\|_2}.$$

We wish to bound the mean and variance of $F(\boldsymbol{M})$. These proofs work by considering the Hermite expansions of A and B given by,

$$A(\boldsymbol{X}) = \sum_{S\subseteq[n]} \widehat{A}_S \boldsymbol{X}_S \qquad \text{and} \qquad B(\boldsymbol{X}) = \sum_{T\subseteq[n]} \widehat{B}_T \boldsymbol{Y}_T.$$

The basic definitions and facts related to Hermite polynomials were given in section 2.

Proof of Lemma 10. We start out by proving the bound on $\left|\mathbb{E}_{\boldsymbol{M}} F(\boldsymbol{M}) - \langle A, B\rangle_{\mathcal{G}_\rho^{\otimes n}}\right|$. To this end, we will use Lemma 18 with parameters d and δ. Thus, for a choice of $n_0 = d^{O(d)}/\delta^2$,

we have that,

$$\begin{aligned}
&\left|\mathbb{E}_{\boldsymbol{M}} F(\boldsymbol{M}) - \langle A, B\rangle_{\mathcal{G}_\rho^{\otimes n}}\right| \\
&= \left|\mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}[A(\boldsymbol{U})\cdot B(\boldsymbol{V})] - \mathop{\mathbb{E}}_{\boldsymbol{X},\boldsymbol{Y}\sim\mathcal{G}_\rho^{\otimes n}}[A(\boldsymbol{X})\cdot B(\boldsymbol{Y})]\right| \\
&= \left|\sum_{S,T\subseteq[n]} \widehat{A}_S\widehat{B}_T\cdot\left(\mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}[\boldsymbol{U}_S\cdot\boldsymbol{V}_T] - \mathbb{E}_{\boldsymbol{X},\boldsymbol{Y}\sim\mathcal{G}_\rho^{\otimes n}}[\boldsymbol{X}_S\cdot\boldsymbol{Y}_T]\right)\right| \\
&= \left|\sum_{S\subseteq[n]} \widehat{A}_S\widehat{B}_S\cdot\left(\mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}[\boldsymbol{U}_S\cdot\boldsymbol{V}_S] - \rho^{|S|}\right)\right| \quad \ldots \text{(terms corresponding to } S\neq T \text{ are 0.)} \\
&\leq \sum_{S\subseteq[n]}\left|\widehat{A}_S\widehat{B}_S\right|\cdot\delta \qquad \ldots\ldots \text{ (using Lemma 18)} \\
&\leq \left\|A\right\|_2\cdot\left\|B\right\|_2\cdot\delta \qquad \ldots\ldots \text{ (Cauchy-Schwarz inequality)} \\
&\leq \delta \qquad \ldots\ldots \text{ } (\left\|A\right\|_2, \left\|B\right\|_2 \leq 1)
\end{aligned}$$

◀

We now move to proving the bound on $\mathsf{Var}_{\boldsymbol{M}}(F(\boldsymbol{M}))$. To this end, we will use Lemma 19 with parameters d and $\delta/9^d$. Thus, for a choice of $n_0 = d^{O(d)}/\delta^2$, we have that,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\boldsymbol{M}}\left(\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} A(\boldsymbol{U})\cdot B(\boldsymbol{V})\right)^2 - \left(\mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}} A(\boldsymbol{U})\cdot B(\boldsymbol{V})\right)^2 \\
&= \left|\mathop{\mathbb{E}}_{\substack{\boldsymbol{M}\\ \boldsymbol{a},\boldsymbol{b}\\ \boldsymbol{a}',\boldsymbol{b}'}}[A(\boldsymbol{U})B(\boldsymbol{V})A(\boldsymbol{U}')B(\boldsymbol{V}')] - \left(\mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}[A(\boldsymbol{U})B(\boldsymbol{V})]\right)\cdot\left(\mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'}[A(\boldsymbol{U}')B(\boldsymbol{V}')]\right)\right| \\
&\leq \sum_{\substack{S,T\subseteq[n]\\ S',T'\subseteq[n]}}\left|\widehat{A}_S\widehat{B}_T\widehat{A}_{S'}\widehat{B}_{T'}\right|\cdot\left|\begin{array}{c}\mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}\mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'}[\boldsymbol{U}_S\boldsymbol{V}_T\boldsymbol{U}'_{S'}\boldsymbol{V}'_{T'}] \\ -\left(\mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a},\boldsymbol{b}}[\boldsymbol{U}_S\boldsymbol{V}_T]\right)\cdot\left(\mathop{\mathbb{E}}_{\boldsymbol{M}}\mathop{\mathbb{E}}_{\boldsymbol{a}',\boldsymbol{b}'}[\boldsymbol{U}'_{S'}\boldsymbol{V}'_{T'}]\right)\end{array}\right| \\
&\leq \frac{\delta}{9^d}\cdot\sum_{\substack{S,T,S',T'\subseteq[n]\\ S\triangle T\triangle S'\triangle T'=\emptyset}}\left|\widehat{A}_S\widehat{B}_T\widehat{A}_{S'}\widehat{B}_{T'}\right| .
\end{aligned}$$

To finish the proof, we will show that,

$$\sum_{\substack{S,T,S',T'\subseteq[n]\\ S\triangle T\triangle S'\triangle T'=\emptyset}}\left|\widehat{A}_S\widehat{B}_T\widehat{A}_{S'}\widehat{B}_{T'}\right| \leq 9^d\cdot\left\|A\right\|_2^2\cdot\left\|B\right\|_2^2 .$$

Define functions $f:\{1,-1\}^n\to\mathbb{R}$, $g:\{1,-1\}^n\to\mathbb{R}$ over the boolean hypercube as,

$$f(x) = \sum_{\substack{S\subseteq[n]\\ |S|\leq d}}\widehat{A}_S\mathcal{X}_S(x) \quad \text{and} \quad g(x) = \sum_{\substack{S\subseteq[n]\\ |S|\leq d}}\widehat{B}_S\mathcal{X}_S(x) .$$

Hypercontractivity bounds [55] for degree-d polynomials over the boolean hypercube imply that,

$$\mathop{\mathbb{E}}_x\left[f(x)^4\right] \leq 9^d\left(\mathop{\mathbb{E}}_x\left[f(x)^2\right]\right)^2 \quad \text{and} \quad \mathop{\mathbb{E}}_x\left[g(x)^4\right] \leq 9^d\left(\mathop{\mathbb{E}}_x\left[g(x)^2\right]\right)^2 .$$

We now finish the proof as follows,

$$\begin{aligned}\sum_{\substack{S,T,S',T'\subseteq[n]\\ S\triangle T\triangle S'\triangle T'=\emptyset}} \left|\widehat{A}_S\widehat{B}_T\widehat{A}_{S'}\widehat{B}_{T'}\right| &= \mathop{\mathbb{E}}_x\left[f(x)^2g(x)^2\right]\\ &\le \left(\mathop{\mathbb{E}}_x\left[f(x)^4\right]\right)^{1/2}\cdot\left(\mathop{\mathbb{E}}_x\left[g(x)^4\right]\right)^{1/2} \quad \dots\text{(Cauchy-Schwarz)}\\ &\le 9^d\cdot\left(\mathop{\mathbb{E}}_x\left[f(x)^2\right]\right)\cdot\left(\mathop{\mathbb{E}}_x\left[g(x)^2\right]\right) \quad \dots\text{(Hypercontractivity)}\\ &= 9^d\cdot\left\|A\right\|_2^2\cdot\left\|B\right\|_2^2.\end{aligned}$$

Thus, overall we get that, $\mathsf{Var}_{\boldsymbol{M}}(F(\boldsymbol{M}))\le\delta$.
This completes the proof of Lemma 10 for an explicit choice of $n_0\le d^{O(d)}/\delta^2$.

B Proof of Low-Degree Multilinear Transformation Lemma

The goal of this section is to prove Lemma 11, which follows immediately by putting together the following two lemmas. The first lemma transforms general functions to low-degree polynomials and second lemma subsequently transforms it to multilinear polynomials.

▶ **Lemma 21** (Low Degree Transformation). *Given parameters $\rho\in[0,1]$, $\delta>0$, $k\in\mathbb{N}$, there exists an explicit $d=d(\rho,k,\delta)$ such that the following holds:*
Let $A:\mathbb{R}^n\to\mathbb{R}^k$ and $B:\mathbb{R}^n\to\mathbb{R}^k$, such that, for any $j\in[k]$: $\mathsf{Var}(A_j),\mathsf{Var}(B_j)\le 1$.
Then, there exist functions $\widetilde{A}:\mathbb{R}^n\to\mathbb{R}^k$ and $\widetilde{B}:\mathbb{R}^n\to\mathbb{R}^k$ such that the following hold.

1. *$\widetilde{A}$ and $\widetilde{B}$ have degree at most d.*
2. *For any $i\in[k]$, it holds that $\mathsf{Var}(\widetilde{A}_i)\le\mathsf{Var}(A_i)\le 1$ and $\mathsf{Var}(\widetilde{B}_i)\le\mathsf{Var}(B_i)\le 1$.*
3. *$\left\|\mathcal{R}(\widetilde{A})-\widetilde{A}\right\|_2\le\left\|\mathcal{R}(A)-A\right\|_2+\delta$ and $\left\|\mathcal{R}(\widetilde{B})-\widetilde{B}\right\|_2\le\left\|\mathcal{R}(B)-B\right\|_2+\delta$*
4. *For every $i,j\in[k]$,*

$$\left|\left\langle\widetilde{A}_i,\widetilde{B}_j\right\rangle_{\mathcal{G}_\rho^{\otimes n}}-\langle A_i,B_j\rangle_{\mathcal{G}_\rho^{\otimes n}}\right|\le\frac{\delta}{\sqrt{k}}$$

In particular, one may take $d=O\left(\frac{\sqrt{k}\log^2(k/\delta)}{\delta(1-\rho)}\right)$.

▶ **Lemma 22** (Multi-linear Transformation). *Given parameters $\rho\in[0,1]$, $\delta>0$, $d,k\in\mathbb{Z}_{\ge 0}$, there exists an explicit $t=t(k,d,\delta)$ such that the following holds:*
Let $A:\mathbb{R}^n\to\mathbb{R}^k$ and $B:\mathbb{R}^n\to\mathbb{R}^k$ be degree-d polynomials, such that, for any $j\in[k]$: $\mathsf{Var}(A_j),\mathsf{Var}(B_j)\le 1$.
Then, there exist functions $\widetilde{A}:\mathbb{R}^{nt}\to\mathbb{R}^k$ and $\widetilde{B}:\mathbb{R}^{nt}\to\mathbb{R}^k$ such that the following hold:

1. *$\widetilde{A}$ and $\widetilde{B}$ are multilinear with degree at most d.*
2. *For any $i\in[k]$, it holds that $\mathsf{Var}(\widetilde{A}_i)\le\mathsf{Var}(A_i)\le 1$ and $\mathsf{Var}(\widetilde{B}_i)\le\mathsf{Var}(B_i)\le 1$.*
3. *$\left\|\mathcal{R}(\widetilde{A})-\widetilde{A}\right\|_2\le\left\|\mathcal{R}(A)-A\right\|_2+\delta$ and $\left\|\mathcal{R}(\widetilde{B})-\widetilde{B}\right\|_2\le\left\|\mathcal{R}(B)-B\right\|_2+\delta$*
4. *For every $i,j\in[k]$,*

$$\left|\left\langle\widetilde{A}_i,\widetilde{B}_j\right\rangle_{\mathcal{G}_\rho^{\otimes nt}}-\langle A_i,B_j\rangle_{\mathcal{G}_\rho^{\otimes n}}\right|\le\frac{\delta}{\sqrt{k}}$$

In particular, one may take $t=O\left(\frac{kd^2}{\delta^2}\right)$.

Simple Proposition for Rounding

Before getting to the proofs of the above lemmas, we present a simple proposition that will be useful. It says that if we have two strategies which are close in ℓ_2-distance, and one of them is *close* to the simplex Δ_k, then so is the other. The proof follows by a straightforward triangle inequality.

▶ **Proposition 23.** *For $A : \mathbb{R}^n \to \mathbb{R}^k$ and $\widetilde{A} : \mathbb{R}^n \to \mathbb{R}^k$ s.t. $\|A\|_2, \|\widetilde{A}\|_2 \le 1$, it holds that,*

$$\|\mathcal{R}(\widetilde{A}) - \widetilde{A}\|_2 \;\le\; \|\mathcal{R}(A) - A\|_2 + \|A - \widetilde{A}\|_2 \,.$$

B.1 Transformation to Low-Degree

The key idea behind Lemma 21 is quite standard, that applying a "small" amount of noise (via the Ornstein-Uhlenbeck operator) to a pair of functions doesn't hurt their correlation "significantly". In particular, we have the following lemma.

▶ **Lemma 24.** *Let $P, Q \in L^2(\mathbb{R}^n, \gamma_n)$ and $\varepsilon > 0$. There exists $\nu = \nu(\rho, \varepsilon)$ such that,*

$$\left|\langle P, Q\rangle_{\mathcal{G}_\rho^{\otimes n}} - \langle U_\nu P, U_\nu Q\rangle_{\mathcal{G}_\rho^{\otimes n}}\right| \le \varepsilon \cdot \sqrt{\mathsf{Var}[P]\,\mathsf{Var}[Q]}$$

In particular, one may take $\nu := (1-\varepsilon)^{\log\rho/(\log\varepsilon+\log\rho)}$, or even $\nu := 1 - C\frac{(1-\rho)\varepsilon}{\log(1/\varepsilon)}$ for some constant $C > 0$.

Proof. Consider the Hermite expansions of P and Q. That is,

$$P(\boldsymbol{X}) \;=\; \sum_{\sigma\in\mathbb{Z}^n_{\ge 0}} \widehat{P}(\sigma) H_\sigma(\boldsymbol{X}) \qquad \text{and} \qquad Q(\boldsymbol{Y}) \;=\; \sum_{\sigma\in\mathbb{Z}^n_{\ge 0}} \widehat{Q}(\sigma) H_\sigma(\boldsymbol{Y}).$$

Using properties of Hermite polynomials, namely, $U_\nu H_\sigma = \nu^{|\sigma|} H_\sigma$, we get that,

$$U_\nu P(\boldsymbol{X}) \;=\; \sum_{\sigma\in\mathbb{Z}^n_{\ge 0}} \nu^{|\sigma|}\widehat{P}(\sigma) H_\sigma(\boldsymbol{X}) \qquad \text{and} \qquad U_\nu Q(\boldsymbol{Y}) \;=\; \sum_{\sigma\in\mathbb{Z}^n_{\ge 0}} \nu^{|\sigma|}\widehat{Q}(\sigma) H_\sigma(\boldsymbol{Y}).$$

Our choice of ν was to ensures that $\rho^d\left(1-\nu^{2d}\right) \le \varepsilon$ for all $d \in \mathbb{N}$. Thus, we get that,

$$\begin{aligned}
&\left|\langle P, Q\rangle_{\mathcal{G}_\rho^{\otimes n}} - \langle U_\nu P, U_\nu Q\rangle_{\mathcal{G}_\rho^{\otimes n}}\right| \\
&= \left|\sum_{\sigma\neq\mathbf{0}} \rho^{|\sigma|}\cdot\widehat{P}(\sigma)\widehat{Q}(\sigma)\cdot\left(1-\nu^{2|\sigma|}\right)\right| \\
&\le \sum_{\sigma\neq\mathbf{0}} \left|\widehat{P}(\sigma)\widehat{Q}(\sigma)\right|\cdot\rho^{|\sigma|}\left(1-\nu^{2|\sigma|}\right) \\
&\le \varepsilon\cdot\sum_{\sigma\neq\mathbf{0}} \left|\widehat{P}(\sigma)\widehat{Q}(\sigma)\right| \qquad \ldots(\text{since, } \rho^d\left(1-\nu^{2d}\right) \le \varepsilon \text{ for all } d\in\mathbb{N}) \\
&\le \varepsilon\cdot\sqrt{\mathsf{Var}[P]\,\mathsf{Var}[Q]} \qquad \ldots(\text{Cauchy-Schwarz inequality})
\end{aligned}$$

◀

The above lemma transforms general functions into functions which are concentrated on low-degree. Thus, to complete the proof of Lemma 21, we consider the definition of *low-degree truncation.*

▶ **Definition 25** (Low-degree truncation)**.** Let $A \in L^2(\mathbb{R}^n, \gamma_n)$ is given by the Hermite expansion $A(\boldsymbol{X}) = \sum_{\sigma \in \mathbb{Z}^n_{\geq 0}} \widehat{A}_\sigma H_\sigma(\boldsymbol{X})$. The *degree-$d$ truncation* of A is defined as the function $A^{\leq d} \in L^2(\mathbb{R}^n, \gamma_n)$ given by

$$A^{\leq d}(\boldsymbol{X}) \;:=\; \sum_{\substack{\sigma \in \mathbb{Z}^n_{\geq 0} \\ |\sigma| \leq d}} \widehat{A}_\sigma H_\sigma(\boldsymbol{X}).$$

That is, $A^{\leq d}$ is obtained by retaining only the terms with degree at most d in the Hermite expansion of A, where recall that for $\sigma \in \mathbb{Z}^n_{\geq 0}$, its degree is defined as $|\sigma| = \sum_{i=1}^n \sigma_i$.
For convenience, define $A^{>d} := A - A^{\leq d}$. Also, for vector valued functions A, we define $A^{\leq d}$ as the function obtained by applying the above low-degree truncation on each coordinate.

Proof of Lemma 21. We obtain $\widetilde{A}$ and $\widetilde{B}$ by first applying some suitable amount of noise to the functions such that the functions have decaying Hermite tails and then truncating the Hermite coefficients corresponding to terms larger than degree d.

In particular, given parameter δ, we first choose ε and ν in Lemma 24, such that $\varepsilon = \frac{\delta}{2\sqrt{k}}$ and then $\nu = 1 - C\frac{(1-\rho)\varepsilon}{\log(1/\varepsilon)}$ as required. We choose d to be large enough such that $\nu^{2d} \leq \frac{\delta}{4\sqrt{k}}$, that is, $d = O\left(\frac{\log(k/\delta)}{\log(1/\nu)}\right) = O\left(\frac{\sqrt{k}\log^2(k/\delta)}{\delta(1-\rho)}\right)$. Finally, we let $\widetilde{A} := (U_\nu A)^{\leq d}$ and $\widetilde{B} := (U_\nu B)^{\leq d}$.

We now verify the four properties required of the lemma.

1. By definition, $\widetilde{A}$ and $\widetilde{B}$ have degree at most d.
2. $\mathsf{Var}(\widetilde{A}_i) \;=\; \sum_{\substack{\sigma \neq \mathbf{0} \\ |\sigma| \leq d}} \nu^{2|\sigma|} \cdot \widehat{A}_i(\sigma)^2 \;\leq\; \mathsf{Var}(A_i)$. Similarly, $\mathsf{Var}(\widetilde{B}_i) \;\leq\; \mathsf{Var}(B_i)$.
3. For convenience, define $\overline{A} := U_\nu A$, and hence $\widetilde{A} = \overline{A}^{\leq d}$. Observe that, since Δ_k is a convex body, $\left\|\mathcal{R}(v) - v\right\|_2^2$ is a convex function in $v \in \mathbb{R}^k$. Thus, we have that,

$$\begin{aligned}
\left\|\mathcal{R}(\overline{A}) - \overline{A}\right\|_2^2 \;&=\; \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\gamma_n} \left\|\mathcal{R}(\overline{A}(\boldsymbol{X})) - \overline{A}(\boldsymbol{X})\right\|_2^2 \\
&=\; \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\gamma_n} \left\| \mathcal{R}\left(\mathop{\mathbb{E}}_{\boldsymbol{X}'\sim U_\nu(\boldsymbol{X})} A(\boldsymbol{X}') \right) - \mathop{\mathbb{E}}_{\boldsymbol{X}'\sim U_\nu(\boldsymbol{X})} A(\boldsymbol{X}') \right\|_2^2 \\
&\leq\; \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\gamma_n} \mathop{\mathbb{E}}_{\boldsymbol{X}'\sim U_\nu(\boldsymbol{X})} \left\| \mathcal{R}\left(A(\boldsymbol{X}')\right) - A(\boldsymbol{X}') \right\|_2^2 \quad \ldots \text{(using convexity of } \left\|\mathcal{R}(v) - v\right\|_2^2) \\
&=\; \mathop{\mathbb{E}}_{\boldsymbol{X}'\sim\gamma_n} \left\| \mathcal{R}\left(A(\boldsymbol{X}')\right) - A(\boldsymbol{X}') \right\|_2^2 \\
&=\; \left\|\mathcal{R}(A) - A\right\|_2^2 .
\end{aligned}$$

Next, observe that, $\left\|\overline{A}^{>d}\right\|_2^2 \;=\; \sum_{|\sigma|>d} \nu^{2|\sigma|} \cdot \|\widehat{A}(\sigma)\|_2^2 \;\leq\; \nu^{2d} \cdot \sqrt{k} \;\leq\; \frac{\delta}{4}$. Thus, we get that,

$$\begin{aligned}
\left\|\mathcal{R}(\widetilde{A}) - \widetilde{A}\right\|_2 \;&\leq\; \left\|\mathcal{R}(\overline{A}) - \overline{A}\right\|_2 + \left\|\overline{A} - \widetilde{A}\right\|_2 \qquad \ldots \text{(Proposition 23)} \\
&=\; \left\|\mathcal{R}(\overline{A}) - \overline{A}\right\|_2 + \left\|\overline{A}^{>d}\right\|_2 \\
&\leq\; \left\|\mathcal{R}(A) - A\right\|_2 \;+\; \delta/4 \,.
\end{aligned}$$

Similar argument holds for $\widetilde{B}$.

4. For every $i, j \in [k]$, we simply have from Lemma 24 that

$$\left| \left\langle \overline{A}_i, \overline{B}_j \right\rangle_{\mathcal{G}_\rho^{\otimes n}} - \left\langle A_i, B_j \right\rangle_{\mathcal{G}_\rho^{\otimes n}} \right| \;\leq\; \varepsilon \;=\; \frac{\delta}{2\sqrt{k}}.$$

Additionally, since $\left\|\widetilde{A}_i - \overline{A}_i\right\|_2 \leq \frac{\delta}{4\sqrt{k}}$ and $\left\|\widetilde{B}_j - \overline{B}_j\right\|_2 \leq \frac{\delta}{4\sqrt{k}}$, we get using Lemma 7 that $\left|\left\langle \widetilde{A}_i, \widetilde{B}_j \right\rangle_{\mathcal{G}_\rho^{\otimes n}} - \left\langle \overline{A}_i, \overline{B}_j \right\rangle_{\mathcal{G}_\rho^{\otimes n}}\right| \leq \frac{\delta}{2\sqrt{k}}$. We get the desired statement by combining the two above statements. ◀

B.2 Transformation to Multi-linear

The key idea behind Lemma 22 is similar to that of Lemma 21 in that, we first apply a transformation on our polynomials that makes it concentrated on multilinear terms, while slightly increasing the number of variables. Subsequently, we apply a *multi-linear truncation* defined as follows.

▶ **Definition 26** (Multilinear truncation)**.** Suppose $A \in L^2(\mathbb{R}^n, \gamma_n)$ is given by the Hermite expansion $A(x) = \sum_{\sigma \in \mathbb{Z}^n_{\geq 0}} \widehat{A}_\sigma H_\sigma(x)$. The *multilinear truncation* of A is defined as the function $A^{\mathrm{ml}} \in L^2(\mathbb{R}^n, \gamma_n)$ given by

$$A^{\mathrm{ml}}(x) := \sum_{\sigma \in \{0,1\}^n} \widehat{A}_\sigma H_\sigma(x).$$

That is, A^{ml} is obtained by retaining only the multilinear terms in the Hermite expansion of A.
For convenience, also define $A^{\mathrm{nml}} := A - A^{\mathrm{ml}}$. Also, for vector valued functions A, we define A^{ml} as the function obtained by applying the above multilinear truncation on each coordinate.

▶ **Lemma 27.** *Given parameters $\rho \in [0,1]$, $\delta > 0$ and $d \in \mathbb{Z}_{\geq 0}$, there exists $t = t(d, \delta)$ such that the following holds:*

Let $A, B \in L^2(\mathbb{R}^n, \gamma_n)$ be degree-d polynomials, such that $\|A\|_2, \|B\|_2 \leq 1$. Define polynomials $\overline{A}, \overline{B} \in L^2(\mathbb{R}^{nt}, \gamma_{nt})$ over variables $\overline{\boldsymbol{X}} := \left\{\boldsymbol{X}_j^{(i)} : (i,j) \in [n] \times [t]\right\}$ and $\overline{\boldsymbol{Y}} := \left\{\boldsymbol{Y}_j^{(i)} : (i,j) \in [n] \times [t]\right\}$ respectively, as,

$$\overline{A}\left(\overline{\boldsymbol{X}}\right) := A(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(n)}) \qquad \text{and} \qquad \overline{B}\left(\overline{\boldsymbol{Y}}\right) := B(\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(n)})$$

where $\boldsymbol{X}^{(i)} = \left(\boldsymbol{X}_1^{(i)} + \cdots + \boldsymbol{X}_t^{(i)}\right)/\sqrt{t}$ and $\boldsymbol{Y}^{(i)} = \left(\boldsymbol{Y}_1^{(i)} + \cdots + \boldsymbol{Y}_t^{(i)}\right)/\sqrt{t}$.
Since $(\boldsymbol{X}^{(i)}, \boldsymbol{Y}^{(i)})$ is distributed according to $\mathcal{G}_\rho$, this transformation doesn't change the "structure" of A and B. In particular, it is follows that,

$$\left\langle \overline{A}, \overline{B} \right\rangle_{\mathcal{G}_\rho^{\otimes nt}} = \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes n}} \qquad \text{and} \qquad \left\|\overline{A}\right\|_2 = \|A\|_2 \qquad \text{and} \qquad \left\|\overline{B}\right\|_2 = \|B\|_2$$

Next, let $\overline{A}^{\mathrm{ml}}, \overline{B}^{\mathrm{ml}} \in L^2(\mathbb{R}^{nt}, \gamma_{nt})$ be the multilinear truncations of $\overline{A}$ and $\overline{B}$ respectively. Then the following hold,

1. *$\overline{A}^{\mathrm{ml}}$ and $\overline{B}^{\mathrm{ml}}$ are multilinear with degree at most d.*
2. *$\mathsf{Var}(\overline{A}^{\mathrm{ml}}) \leq \mathsf{Var}(A) \leq 1$ and $\mathsf{Var}(\overline{B}^{\mathrm{ml}}) \leq \mathsf{Var}(B) \leq 1$.*
3. *$\left\|\overline{A}^{\mathrm{ml}} - \overline{A}\right\|_2, \left\|\overline{B}^{\mathrm{ml}} - \overline{B}\right\|_2 \leq \delta/2$.*
4. *$\left|\left\langle \overline{A}^{\mathrm{ml}}, \overline{B}^{\mathrm{ml}} \right\rangle_{\mathcal{G}_\rho^{\otimes nt}} - \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes n}}\right| \leq \delta$.*

In particular, one may take $t = O\left(\frac{d^2}{\delta^2}\right)$.

In order to prove Lemma 27, we will need the following multinomial theorem for Hermite polynomials. It can be proved quite easily using the generating function for Hermite polynomials.

▶ **Fact 28** (Multinomial theorem for Hermite polynomials). *Let $\beta_1, \ldots, \beta_t \in \mathbb{R}$ satisfying $\sum_{i=1}^{t} \beta_i^2 = 1$. Then, for any $d \in \mathbb{N}$, it holds that*

$$H_d\left(\beta_1 X_1 + \cdots + \beta_t X_t\right) \;=\; \sum_{\substack{d_1,\ldots,d_t \in \mathbb{Z}_{\geq 0} \\ d_1+\cdots+d_t=d}} \sqrt{\frac{d!}{d_1!\cdots d_t!}} \cdot \prod_{i=1}^{t} \beta_i^{d_i} H_{d_i}(X_i)\ .$$

Proof of Lemma 27. Before we prove the theorem, we will first understand the effect of the transformation from X to $\overline{X}$ for a univariate Hermite polynomial. Instantiating β_i's in Fact 28 with $1/\sqrt{t}$, we get that,

$$H_d\left(\frac{X_1 + \cdots + X_t}{\sqrt{t}}\right) \;=\; \sum_{\substack{d_1,\ldots,d_t \in \mathbb{Z}_{\geq 0} \\ d_1+\cdots+d_t=d}} \sqrt{\frac{d!}{d_1!\cdots d_t!}} \cdot \frac{\prod_{i=1}^{t} H_{d_i}(X_i)}{t^{d/2}}\ .$$

We will split the terms into multilinear and non-multilinear terms, writing the above as $H_d^{\mathrm{ml}} + H_d^{\mathrm{nml}}$. Note that there are at most $O(\frac{d^2 t^{d-1}}{d!})$ non-multilinear terms (for $t \gg d^2$). Also, note that each coefficient $\frac{1}{t^{d/2}} \cdot \sqrt{\frac{d!}{d_1!\cdots d_t!}}$ is at most $\sqrt{\frac{d!}{t^d}}$. Thus, we can bound $\left\|H_d^{\mathrm{nml}}\right\|_2$ as follows,

$$\left\|H_d^{\mathrm{nml}}\right\|_2^2 \;=\; \sum_{\substack{d_1,\ldots,d_t \in \mathbb{Z}_{\geq 0} \\ d_1+\cdots+d_t=d \\ \exists i \;\; d_i \geq 2}} \left(\frac{1}{t^{d/2}} \cdot \sqrt{\frac{d!}{d_1!\cdots d_t!}}\right)^2 \;\leq\; O\left(\frac{d^2 t^{d-1}}{d!}\right) \cdot \frac{d!}{t^d} \;\leq\; O\left(\frac{d^2}{t}\right) \tag{26}$$

More generally, if we consider a term $\overline{H}_\sigma\left(\overline{X}\right) = H_{\sigma_1}(X^{(1)}) \cdot H_{\sigma_2}(X^{(2)}) \cdots H_{\sigma_n}(X^{(n)})$, where each $X^{(i)} = \left(X_1^{(i)} + \cdots + X_t^{(i)}\right)/\sqrt{t}$. Let's write $\overline{H}_\sigma\left(\overline{X}\right) = \overline{H}_\sigma^{\mathrm{ml}}\left(\overline{X}\right) + \overline{H}_\sigma^{\mathrm{nml}}\left(\overline{X}\right)$, that is, separating out the multilinear and non-multilinear terms. Similarly, for any i, let $H_{\sigma_i}(X^{(i)}) = H_{\sigma_i}^{\mathrm{ml}}(X^{(i)}) + H_{\sigma_i}^{\mathrm{nml}}(X^{(i)})$. We wish to bound $\left\|\overline{H}_\sigma^{\mathrm{nml}}\right\|_2$, which can be done as follows,

$$\begin{aligned}
\left\|\overline{H}_\sigma^{\mathrm{nml}}\right\|_2^2 \;&=\; \left\|\prod_{i=1}^{n}(H_{\sigma_i}^{\mathrm{ml}} + H_{\sigma_i}^{\mathrm{nml}}) - \prod_{i=1}^{n} H_{\sigma_i}^{\mathrm{ml}}\right\|_2^2 \\
&\leq\; \prod_{i=1}^{n}\left(1 + O\left(\frac{\sigma_i^2}{t}\right)\right) - 1 && \text{(from Equation 26)} \\
&\leq\; O\left(\frac{|\sigma|^2}{t}\right) && \text{(since, } t \gg |\sigma|^2\text{)} \\
\text{Thus,}\quad \left\|\overline{H}_\sigma^{\mathrm{nml}}\right\|_2^2 \;&<\; \delta^2/4. && \text{(for } t = \Theta(d^2/\delta^2)\text{)} \qquad (27)
\end{aligned}$$

We are now ready to prove the parts of Lemma 27.

1. It holds by definition that $\overline{A}^{\mathrm{ml}}$ and $\overline{B}^{\mathrm{ml}}$ are multilinear. Also, note that the transformation from A to $\overline{A}$ and finally to $\overline{A}^{\mathrm{ml}}$ does not increase the degree. So both $\overline{A}^{\mathrm{ml}}$ and $\overline{B}^{\mathrm{ml}}$ have degree at most d.

2. It is easy to see that $\mathsf{Var}(\overline{A}) = \mathsf{Var}(A)$. Since $\overline{A}^{\mathrm{ml}}$ is obtained by truncating certain Hermite coefficients of $\overline{A}$, it immediately follows that $\mathsf{Var}(\overline{A}^{\mathrm{ml}}) \leq \mathsf{Var}(\overline{A}) = \mathsf{Var}(A) \leq 1$. Similarly, $\mathsf{Var}(\overline{B}^{\mathrm{ml}}) \leq \mathsf{Var}(B) \leq 1$.
3. Recall that $\overline{A}^{\mathrm{nml}} = \overline{A} - \overline{A}^{\mathrm{ml}}$. We wish to bound $\left\|\overline{A}^{\mathrm{nml}}\right\|_2^2 \leq \delta^2/4$. Consider the Hermite expansion of A, namely $A(\boldsymbol{X}) = \sum_{\sigma \in \mathbb{Z}^n_{\geq 0}} \widehat{A}(\sigma) \cdot H_\sigma(\boldsymbol{X})$. Note that, $\overline{A}^{\mathrm{nml}}\left(\overline{\boldsymbol{X}}\right) = \sum_{\sigma \in \mathbb{Z}^n_{\geq 0}} \widehat{A}(\sigma) \cdot \overline{H}^{\mathrm{nml}}_\sigma\left(\overline{\boldsymbol{X}}\right)$, where recall that $\overline{H}^{\mathrm{nml}}_\sigma$ is the non-multilinear part of $\overline{H}_\sigma\left(\overline{\boldsymbol{X}}\right) = H_{\sigma_1}(X^{(1)}) \cdot H_{\sigma_2}(X^{(2)}) \cdots H_{\sigma_n}(X^{(n)})$, where each $X^{(i)} = \left(X_1^{(i)} + \cdots + X_t^{(i)}\right)/\sqrt{t}$.

   From Equation 27, we have that for any $\sigma \in \mathbb{Z}^n_{\geq 0}$, it holds that $\left\|\overline{H}^{\mathrm{nml}}_\sigma\right\|_2^2 < \delta^2/4$. And hence we get that,

$$\left\|\overline{A}^{\mathrm{nml}}\right\|_2^2 = \sum_\sigma \widehat{A}(\sigma)^2 \cdot \left\|\overline{H}^{\mathrm{nml}}_\sigma\right\|_2^2 \leq \sum_\sigma \widehat{A}(\sigma)^2 \cdot (\delta^2/4) = (\delta^2/4)\left\|A\right\|_2^2 \leq (\delta^2/4).$$

   Note that, here we use that $\overline{H}_\sigma(\overline{\boldsymbol{X}})$ are mutually orthogonal for different σ. Similarly, we can also get that $\left\|\overline{B}^{\mathrm{nml}}\right\|_2^2 \leq \delta^2/4$.
4. Note that we already have,

$$\left\langle \overline{A}, \overline{B} \right\rangle_{\mathcal{G}_\rho^{\otimes nt}} = \left\langle A, B \right\rangle_{\mathcal{G}_\rho^{\otimes n}} .$$

   And combining Part 3 and Lemma 7, we immediately get that

$$\left| \left\langle \overline{A}^{\mathrm{ml}}, \overline{B}^{\mathrm{ml}} \right\rangle_{\mathcal{G}_\rho^{\otimes nt}} - \left\langle \overline{A}, \overline{B} \right\rangle_{\mathcal{G}_\rho^{\otimes nt}} \right| \leq \delta$$

   where we use that $\left\|\overline{B}^{\mathrm{ml}}\right\|_2 \leq \left\|\overline{B}\right\|_2 \leq 1$ and $\left\|\overline{A}^{\mathrm{ml}}\right\|_2 \leq \left\|\overline{A}\right\|_2 \leq 1$. ◀

Proof of Lemma 22. We apply the transformation in Lemma 27, with parameter δ being $\delta/\sqrt{k}$, to each of the k-coordinates of $A : \mathbb{R}^n \to \mathbb{R}^k$ and $B : \mathbb{R}^n \to \mathbb{R}^k$ to get polynomials $\widetilde{A} : \mathbb{R}^{nt} \to \mathbb{R}^k$ and $\widetilde{B} : \mathbb{R}^n \to \mathbb{R}^k$. Namely, for any $j \in [k]$, we set $\widetilde{A}_j(\overline{\boldsymbol{X}}) = \overline{A}_j^{\mathrm{ml}}(\overline{\boldsymbol{X}})$ and $\widetilde{B}_j(\overline{\boldsymbol{Y}}) = \overline{B}_j^{\mathrm{ml}}(\overline{\boldsymbol{Y}})$as described in Lemma 27.

It is easy to see that parts 1, 2, 4 follow immediately from the conditions satisfied in Lemma 27. For part 3, we have that $\left\|\overline{A}_j^{\mathrm{ml}} - \overline{A}_j\right\|_2 \leq \delta/\sqrt{k}$ for every $j \in [k]$, which implies that $\left\|\overline{A}^{\mathrm{ml}} - \overline{A}\right\|_2 \leq \delta$. Using Proposition 23, we immediately get that,

$$\left\|\mathcal{R}(\overline{A}^{\mathrm{ml}}) - \overline{A}^{\mathrm{ml}}\right\|_2 \leq \left\|\mathcal{R}(\overline{A}) - \overline{A}\right\|_2 + \delta .$$

Finally, it is a simple observation that $\left\|\mathcal{R}(\overline{A}) - \overline{A}\right\|_2 = \left\|\mathcal{R}(A) - A\right\|_2$, and hence,

$$\left\|\mathcal{R}(\widetilde{A}) - \widetilde{A}\right\|_2 \leq \left\|\mathcal{R}(A) - A\right\|_2 + \delta .$$

Similarly, $\left\|\mathcal{R}(\widetilde{B}) - \widetilde{B}\right\|_2 \leq \left\|\mathcal{R}(B) - B\right\|_2 + \delta$. This concludes the proof. ◀