Dimension Reduction for Polynomials over Gaussian Space and Applications

Badih Ghazi [*]   Pritish Kamath [†]   Prasad Raghavendra [‡]

August 7, 2017

Abstract

In this work 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 applications, we address the following problems:

(I) **Computability of the Approximately Optimal Noise Stable function over Gaussian space.** The goal here is to find a partition of $\mathbb{R}^n$ into k parts, that maximizes the noise stability. An ε-optimal partition is one which is within additive ε of the optimal noise stability.

De, Mossel & Neeman (CCC 2017) raised the question of an explicit (computable) bound on the dimension $n_0(\varepsilon)$ in which we can find an ε-optimal partition.

De et al. already provide such an explicit bound. Using our dimension reduction technique, we are able to obtain improved explicit bounds on the dimension $n_0(\varepsilon)$.

(II) **Decidability of Approximate Non-Interactive Simulation of Joint Distributions.** A *non-interactive simulation* problem is specified by two distributions $P(x, y)$ and $Q(u, v)$: The goal is to determine if two players, Alice and Bob, that observe sequences X^n and Y^n respectively where $\{(X_i, Y_i)\}_{i=1}^{n}$ are drawn i.i.d. from $P(x, y)$ can generate pairs U and V respectively (without communicating with each other) with a joint distribution that is arbitrarily close in total variation to $Q(u, v)$. Even when P and Q are extremely simple, it is open in several cases if P can simulate Q.

Ghazi, Kamath & Sudan (FOCS 2016) formulated a gap problem of deciding whether there exists a non-interactive simulation protocol that comes ε-close to simulating Q, or does every non-interactive simulation protocol remain 2ε-far from simulating Q? The main underlying challenge here is to determine an explicit (computable) upper bound on the number of samples $n_0(\varepsilon)$ that can be drawn from $P(x, y)$ to get ε-close to Q (if it were possible at all).

While Ghazi et al. answered the challenge in the special case where Q is a joint distribution over $\{0, 1\} \times \{0, 1\}$, it remained open to answer the case where Q is a distribution over larger alphabet, say $[k] \times [k]$ for $k > 2$. Recently De, Mossel & Neeman (in a follow-up work), address this challenge for all $k \geq 2$. In this work, we are able to recover this result as well, with improved explicit bounds on $n_0(\varepsilon)$.

Our technique of dimension reduction for low-degree polynomials is simple and analogous to the Johnson-Lindenstrauss lemma, and could be of independent interest.

Contents

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. In 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 be formulated in terms of the notion of *Noise stability*.

Fix a real number $\rho \in [0,1]$. Suppose $f : \mathbb{R}^n \to [0,1]$ denotes the indicator function of a subset (say $\mathcal{A}_f$) of the n-dimensional Gaussian space ($\mathbb{R}^n$ with the Gaussian measure), then its noise stability $\mathsf{Stab}_\rho(f)$ is the probability that two ρ-correlated Gaussians $\mathbf{X}, \mathbf{Y}$ both fall in $\mathcal{A}_f$. Specifically, if $\mathcal{G}_\rho^{\otimes n}$ denotes the distribution of ρ-correlated Gaussians in n dimensions, that is, $\mathbf{X} \sim \gamma_n$ and $(\mathbf{Y}|\mathbf{X}) \sim (\rho\mathbf{X} + \sqrt{1-\rho^2}\mathbf{Z})$ for $\mathbf{Z} \sim \gamma_n$. Then, we can equivalently define noise stability as, $\mathsf{Stab}_\rho(f) = \Pr_{(\mathbf{X},\mathbf{Y})\sim\mathcal{G}_\rho^{\otimes n}}[f(\mathbf{X}) = f(\mathbf{Y})]$. More formally, the Ornstein-Uhlenbeck operator U_ρ, defined for each $\rho \in [0,1]$, acts on any $f : \mathbb{R}^n \to \mathbb{R}$ as

$$(U_\rho f)(\mathbf{X}) = \int_{\mathbf{Z}\in\mathbb{R}^n} f\left(\rho \cdot \mathbf{X} + \sqrt{1-\rho^2} \cdot \mathbf{Z}\right) \, d\gamma_n(\mathbf{Z}),$$

The noise stability is then defined as $\mathsf{Stab}_\rho(f) \overset{\text{def}}{=} \mathbb{E}_{\mathbf{X}\sim\gamma_n}[f(\mathbf{X}) \cdot U_\rho f(\mathbf{X})]$.

In terms of noise stability, the simplest isoperimetric problem is to determine, what is the largest possible value of $\mathsf{Stab}_\rho(f)$ for a function $f : \mathbb{R}^n \to [0,1]$ with a given expectation $\mathbb{E}[f] = \mu$. The seminal isoperimetric theorem of Borell [Bor85] shows that indicator function of halfspaces are the most noise-stable among all functions $f : \mathbb{R}^n \to [0,1]$ with a given expectation.

Borell's theorem (along with the invariance principle of [MOO05, Mos10]) has had fundamental applications in theoretical computer science, e.g., in the hardness of approximation for Max-Cut under the Unique Games conjecture [KKMO07], and in voting theory [Mos10].

In this work, we will be interested in higher analogues of Borell's theorem for partitions of the Gaussian space in to more than two subsets, or equivalently noise stability of functions f taking values over $[k] = \{0, \dots, k-1\}$. Towards stating these higher analogues of Borell's theorem, let's state Borell's theorem in a more general notation. Let Δ_k be the probability simplex in $\mathbb{R}^k$ (i.e. convex hull of the basis vectors $\{e_1, \dots, e_k\}$). The Ornstein-Uhlenbeck operator naturally extends to vector valued functions $f : \mathbb{R}^n \to \mathbb{R}^k$ as $U_\rho f = (U_\rho f_1, \dots, U_\rho f_k)$ (where $f = (f_1, \dots, f_k)$). The noise stability of functions $f : \mathbb{R}^n \to \Delta_k$, is now defined as $\mathsf{Stab}_\rho(f) := \mathbb{E}_{\mathbf{X}\sim\gamma_n}[\langle f(\mathbf{X}), U_\rho f(\mathbf{X})\rangle]$ 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 $e_i \in \Delta_k$. We can now state Borell's theorem in this notation as follows:

Borell's Theorem [Bor85]. *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 $h_1(\mathbf{X}) = \mathbf{1}_{\{\langle a,\mathbf{X}\rangle \geq b\}}$ and $h_2(\mathbf{X}) = 1 - h_1(\mathbf{X})$, for suitable $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ such that $\mathbb{E}[f] = \mathbb{E}[h]$. Then, $\mathsf{Stab}_\rho(f) \leq \mathsf{Stab}_\rho(h)$.*

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

The above question remains open even for $k = 3$. In the particular case where $\alpha = (\frac{1}{k}, \dots, \frac{1}{k})$, the *Standard Simplex Conjecture*[1] posits that the maximum noise stability is achieved by a "standard simplex partition" [KKMO07, IM12]. 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. In fact, a suprising result of [HMN16] shows that when the α_i's are not all equal, the standard simplex partition (an appropriately shifted version thereof) *does not* achieve the maximum

---

[1]also referred to as the *Peace-Sign Conjecture* when $k = 3$.

noise stability. This indicates that the case $k \geq 3$ is fundamentally different than the case where $k = 2$. The fact that we don't understand optimal partitions for $k \geq 3$, led De, Mossel & Neeman [DMN17a] to ask whether the optimal partition is realized in any finite dimension. More formally:

Question. 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 the above question remains open as of now! In this light, De, Mossel & Neeman [DMN17a] 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 the maximum achievable noise stability. Note that the challenge is really about n_0 being "explicit", since some $n_0(\rho, k, \varepsilon)$ always exists, as $S_n(\alpha)$ is a converging sequence as $n \to \infty$.

Indeed, De, Mossel and Neeman obtain such an *explicitly computable* function. To do so, they use and build on the theory of *eigenregular polynomials* that were previously studied by [DS14], which in turn uses other tools such as Malliavin calculus.

In this work, we introduce fundamentally different techniques (elaborated on shortly), thereby recovering the result of [DMN17a]. In particular, we show the following (we use $\mathcal{R} : \mathbb{R}^k \to \Delta_k$ to denote the "rounding operator", as in Definition 2.3).

Theorem 1.1 (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 $n_0 = n_0(\rho, k, \varepsilon)$ such that the following holds: For any $n \in \mathbb{N}$, let $f : \mathbb{R}^n \to \Delta_k$. Then, there exists a function $\widetilde{f} : \mathbb{R}^{n_0} \to \Delta_k$ such that*

1. $\left\| \mathbb{E}[f] - \mathbb{E}[\widetilde{f}] \right\|_1 \leq \varepsilon$.

2. $\mathsf{Stab}_\rho(\widetilde{f}) \geq \mathsf{Stab}_\rho(f) - \varepsilon$.

Moreover, there exists an explicitly computable $d_0 = d_0(\rho, k, \varepsilon)$ for which there is a degree-d_0 polynomial $g : \mathbb{R}^{n_0} \to \mathbb{R}^k$, such that, $\widetilde{f}(a) = \mathcal{R}\left(g\left(\frac{a}{\|a\|_2} \right) \right)$.

The explicit n_0 and d_0 are upper bounded as $n_0 \leq \exp\left(\mathrm{poly}\left(k, \frac{1}{1-\rho}, \frac{1}{\varepsilon} \right) \right)$ and $d_0 \leq \mathrm{poly}\left(k, \frac{1}{1-\rho}, \frac{1}{\varepsilon} \right)$.

Remarks.

(i) While we do obtain an actual *explicit* bound on n_0 and d_0, we skip it in the theorem statement in order to stress on the qualitative nature of the bound. In contrast, it is mentioned in [DMN17a] that their bound on n_0 is not primitive recursive and has an Ackermann-type growth (which is introduced by their application of the regularity lemma from [DS14]).

(ii) A subtle point in our theorem is that the range of $\widetilde{f}$ is Δ_k and not $[k]$. Interestingly however, it follows from a thresholding lemma in [DMN17a, 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.

The above theorem has an immediate application of showing that approximately most-stable voting schemes (among all low-influential voting schemes) can be computed efficiently. We refer the reader to [DMN17a] for the details of this application.

In order to prove Theorem 1.1, we in fact turn to the more general and seemingly harder problem of *non-interactive simulation of joint distributions*.

1.2 Non-Interactive Simulation of Joint Distributions

Suppose that two players, Alice and Bob, observe the sequence of random variables $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ respectively, where each pair (x_i, y_i) is independently drawn from a *source* joint distribution $\mu(x, y)$. 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 for the same, or in other words, what is the *simulation rate*?

This setup, referred to as the *Non-Interactive Simulation (NIS) of Joint Distributions*, has been extensively studied in Information Theory, and more recently in Theoretical Computer Science. The history of this problem goes back to the classical works of Gács and Körner [GK73] and Wyner [Wyn75]. Specifically, consider the distribution Eq over $\{0, 1\} \times \{0, 1\}$ where both marginals are $\mathsf{Ber}(1/2)$ and the bits 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 [Wit75] who studied the case where the target distribution $\nu = \mathcal{G}_\rho$ is the distribution of ρ-correlated Gaussians. In this case, he showed that the largest correlation (i.e., largest value of ρ) that can be simulated is exactly the well-known "maximal correlation coeffcient" $\rho(\mu)$ (see Definition 2.2) which was first introduced by Hirschfeld [Hir35] and Gebelein [Geb41] and then studied by Rényi [Rén59]. This immediately gives a polynomial time algorithm to decide if $\mathcal{G}_\rho$ can be simulated from samples from a given μ, since the maximal correlation coefficient $\rho(\mu)$ is efficiently computable. In the same work [Wit75], Witsenhausen also considered the case where the target distribution $\nu = \mathsf{DSBS}_\rho$, 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 $\mathsf{DSBS}_{\rho'}$ where $\rho' = 1 - \frac{2 \arccos \rho(\mu)}{\pi}$. Indeed, this is morally same as the rounding technique employed in Goemans-Williamson's approximation algorithm for MaxCut [GW95] 20 years later!

We will consider the modern formulation of the NIS question as defined in [KA16]. 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.

Definition 1.2 (Non-interactive Simulation of Joint Distributions [KA16])**.** *Let $(\mathcal{Z} \times \mathcal{Z}, \mu)$ and $([k] \times [k], \nu)$ be two joint probability spaces. We say that the distribution ν can be* non-interactively simulated *from distribution μ, if there exists a sequence of functions[2] $\left\{ A^{(n)} : \mathcal{Z}^n \to [k] \right\}_{n \in \mathbb{N}}$ and $\left\{ B^{(n)} : \mathcal{Z}^n \to [k] \right\}_{n \in \mathbb{N}}$ such that the joint distribution $\nu_n = (A^{(n)}(\boldsymbol{x}), B^{(n)}(\boldsymbol{y}))_{(\boldsymbol{x}, \boldsymbol{y}) \sim \mu^{\otimes n}}$ over $[k] \times [k]$ is such that $\lim_{n \to \infty} d_{\mathrm{TV}}(\nu_n, \nu) = 0$.*

The notion of non-interactive simulation is summarized in Figure 1. Note that even though the definition itself doesn't give Alice and Bob access to private randomness, they can nevertheless take extra samples and use them as private randomness. We will model the use of private randomness, by allowing $A^{(n)}$ and $B^{(n)}$ to map to the simplex Δ_k, instead of $[k]$. We will then interpret $A_i^{(n)}(\boldsymbol{x})$ (resp. $B_j^{(n)}(\boldsymbol{x})$) as the probability of Alice outputting i (resp. Bob outputting j). For convenience, we will still use $(A^{(n)}(\boldsymbol{x}), B^{(n)}(\boldsymbol{y}))_{(\boldsymbol{x}, \boldsymbol{y}) \sim \mu^{\otimes n}}$ to denote the joint distribution generated over $[k] \times [k]$.

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 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 $\mathsf{DSBS}_{0.49}$, it is open if μ can simulate ν. This problem was formalized as a natural

---

[2]we will often refer to such functions as *strategies* of players Alice and Bob.
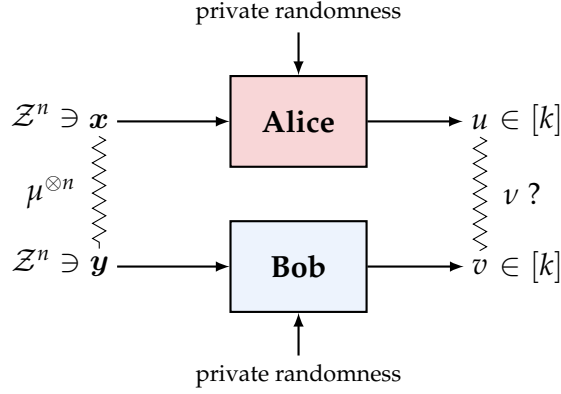
Figure 1: Non-Interactive Simulation, e.g., as studied in [KA16]

gap-version of the non-interactive simulation problem in a work by a subset of the authors along with Madhu Sudan [GKS16b]. Here we state a slightly more generalized version.

Problem 1.3 (GAP-NIS($(\mathcal{Z} \times \mathcal{Z}, \mu), V, k, \varepsilon$), cf. [GKS16b]). *Given a joint probability space $(\mathcal{Z} \times \mathcal{Z}, \mu)$ and another family of joint probability spaces V supported over $[k] \times [k]$, and an error parameter $\varepsilon > 0$, distinguish between the following cases:*

(i) *there exists N, and functions $A : \mathcal{Z}^N \to \Delta_k$ and $B : \mathcal{Z}^N \to \Delta_k$, for which the distribution ν' of $(A(\boldsymbol{x}), B(\boldsymbol{y}))_{(\boldsymbol{x}, \boldsymbol{y}) \sim \mu^{\otimes N}}$ is such that $d_{\mathrm{TV}}(\nu', \nu) \leq \varepsilon$ for some $\nu \in V$.*

(ii) *for all N and all functions $A : \mathcal{Z}^N \to \Delta_k$ and $B : \mathcal{Z}^N \to \Delta_k$, the distribution ν' of $(A(\boldsymbol{x}), B(\boldsymbol{y}))_{(\boldsymbol{x}, \boldsymbol{y}) \sim \mu^{\otimes N}}$ is such that $d_{\mathrm{TV}}(\nu', \nu) > 2\varepsilon$ for all $\nu \in V$.* [3]

1.3 NIS from Gaussian Sources & the MNS question

We now remark on why the NIS question is a more general question than the Maximum Noise Stability question. For any distribution ν, define the agreement probability of ν supported over $[k] \times [k]$ as $\mathrm{agr}(\nu) = \mathrm{Pr}_{(u,v) \sim \nu}[u = v]$. Recall that for $f : \mathbb{R}^n \to \Delta_k$, the stability can equivalently be defined as $\mathrm{Stab}_\rho(f) = \mathbb{E}_{(\mathbf{X}, \mathbf{Y}) \sim \mathcal{G}_\rho^{\otimes n}} \langle f(\mathbf{X}), f(\mathbf{Y}) \rangle = \sum_{i=1}^k \mathbb{E}_{(\mathbf{X}, \mathbf{Y}) \sim \mathcal{G}_\rho^{\otimes n}} [f_i(\mathbf{X}), f_i(\mathbf{Y})]$. Basically, the MNS question can be interpreted as asking: what is the maximum "agreement probability" of any distribution ν that can be non-interactively simulated from $\mu = \mathcal{G}_\rho$, with both marginal distributions given by $\boldsymbol{\alpha}$, and with an additional constraint that both Alice and Bob use the same strategy, i.e. $A = B = f$. Thus, to understand the MNS question, we turn to understanding which target distributions ν can be non-interactively simulated with the source distribution $\mu = \mathcal{G}_\rho$; we will ignore, for the moment, the restriction that Alice and Bob need to use the same strategy.

Recall that implicit in [Wit75], was an approach to non-interactively simulate target distributions ν over $\{0, 1\} \times \{0, 1\}$ from $\mu = \mathcal{G}_\rho$ using *half-space* functions (using only one sample of μ). Combining Witsenhausen's approach and Borell's theorem [Bor85] gives us an exact characterization of all distributions ν over $\{0, 1\} \times \{0, 1\}$ that can be simulated from $\mu = \mathcal{G}_\rho$. Moreover, any distribution ν that can be simulated from $\mathcal{G}_\rho$ can in fact be simulated using only one sample from $\mathcal{G}_\rho$ (potentially in addition to private randomness).

For $k > 2$, we do not have such an exact characterization of the distributions ν that can be simulated from $\mathcal{G}_\rho$. The challenges underlying here are the same as those underlying in the *Standard Simplex Conjecture*. Nevertheless, we prove a bound on the number of samples needed to come ε-close to simulating ν, if it were possible at all, in the form of the following theorem,

---

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

4

Theorem 1.4 (NIS from correlated Gaussian source). *Given parameters $k \geq 2$, $\rho \in (0, 1)$ and $\varepsilon > 0$, there exists an explicitly computable $n_0 = n_0(\rho, k, \varepsilon)$ such that the following holds:*
For any N, and any $A : \mathbb{R}^N \to \Delta_k$ and $B : \mathbb{R}^N \to \Delta_k$, there exist functions $\widetilde{A} : \mathbb{R}^{n_0} \to \Delta_k$ and $\widetilde{B} : \mathbb{R}^{n_0} \to \Delta_k$ such that,

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

Moreover, there exists an explicitly computable $d_0 = d_0(\rho, k, \varepsilon)$ for which there are degree-d_0 polynomials $A_0 : \mathbb{R}^{n_0} \to \mathbb{R}^k$ and $B_0 : \mathbb{R}^{n_0} \to \mathbb{R}^k$, such that, $\widetilde{A}(\boldsymbol{a}) = \mathcal{R}\left(A_0\left(\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2} \right) \right)$ and $\widetilde{B}(\boldsymbol{b}) = \mathcal{R}\left(B_0\left(\frac{\boldsymbol{b}}{\|\boldsymbol{b}\|_2} \right) \right)$.
The explicit n_0 and d_0 are upper bounded as $n_0 \leq \exp\left(\mathrm{poly}\left(k, \frac{1}{1-\rho}, \frac{1}{\varepsilon} \right) \right)$ and $d_0 \leq \mathrm{poly}\left(k, \frac{1}{1-\rho}, \frac{1}{\varepsilon} \right)$.

In fact, 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 simultaneously with the same random seed, if $A = B$, then the transformation gives $\widetilde{A} = \widetilde{B}$ as well.

It is now easy to see that Theorem 1.1 follows simply as a corollary of the above theorem, when applied on functions $A = B = f$.

By an "oblivious" randomized transformation, we mean that to obtain $\widetilde{A}$ from A, we only need to know A and a shared random seed M. That is, the transformation doesn't use the knowledge of B. Similarly, to obtain $\widetilde{B}$ from B, we only need to know B and the same shared random seed M. This hinted at in [GKS16b] as a potential barrier for showing decidability of GAP-NIS when $k \geq 2$. Indeed our transformation overcomes this barrier and we elaborate more on this in Section 1.8.

1.4 NIS from Arbitrary Discrete Sources

In prior work [GKS16b], 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 only 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 combining Witsenhausen [Wit75] and Borell's theorem [Bor85], 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 [GKS16b] as a barrier for extending their result to $k > 2$.

Following up on [DMN17a], De, Mossel & Neeman were able to extend their techniques to show the decidability of GAP-NIS for all $k \geq 2$ [DMN17b]. To do so, they follow the same high level framework of using a Regularity Lemma and Invariance Principle introduced in [GKS16b]. In addition, they build on the tools developed in [DMN17a] along with a new smoothing argument inspired by boosting procedures in learning theory and potential function arguments in complexity theory and additive combinatorics.

In this work, we are able to recover this result using a fundamentally different and more elementary approach, by only using Theorem 1.4 along with the framework introduced in [GKS16b], thereby showing decidability of GAP-NIS for all $k \geq 2$. The central underlying theorem to prove decidability of GAP-NIS is the following.

Theorem 1.5 (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 $n_0 = n_0(\mu, k, \varepsilon)$ such that the following holds:*
Let $A : \mathcal{Z}^N \to \Delta_k$ and $B : \mathcal{Z}^N \to \Delta_k$. Then there exist functions $\widetilde{A} : \mathcal{Z}^{n_0} \to \Delta_k$ and $\widetilde{B} : \mathcal{Z}^{n_0} \to \Delta_k$ such that,

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

In particular, n_0 is an explicit function upper bounded by $\exp\left(\mathrm{poly}\left(k, \frac{1}{\varepsilon}, \frac{1}{1-\rho_0}, \log\left(\frac{1}{\alpha} \right) \right) \right)$, where $\alpha = \alpha(\mu)$ is the smallest atom in μ and $\rho_0 = \rho(\mu)$ is the maximal correlation of μ.

5

The decidability of GAP-NIS follows quite easily from the above theorem. The main idea is, once we know a bound on the number of samples of μ that are needed to get ε-close to ν (if it were possible at all), we can brute force over all possible strategies of Alice and Bob. For completeness, we provide a proof of the following theorem in Appendix F.

Theorem 1.6 (Decidability of GAP-NIS). *Given a joint probability space $(\mathcal{Z} \times \mathcal{Z}, \mu)$ and a family of joint probability spaces V supported over $[k] \times [k]$, and an error parameter $\varepsilon > 0$, there exists an algorithm that runs in time $T(\mu, k, \varepsilon)$ (which is an explicitly computable function), and decides $\text{GAP-NIS}((\mathcal{Z} \times \mathcal{Z}, \mu), V, k, \varepsilon)$.*

The run time $T(\mu, k, \varepsilon)$ is upper bounded by $\exp \exp \exp \left(\text{poly} \left(k, \frac{1}{\varepsilon}, \frac{1}{1 - \rho_0}, \log \left(\frac{1}{\alpha} \right) \right) \right)$, where $\rho_0 = \rho(\mu)$ is the maximal correlation of $(\mathcal{Z} \times \mathcal{Z}, \mu)$ and $\alpha \stackrel{\text{def}}{=} \alpha(\mu)$ is the minimum non-zero probability in μ.

1.5 Dimension Reduction for Low-Degree 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.

Let's start with Theorem 1.4, and see how we might even begin proving it. We are given two vector-valued functions[4] $A : \mathbb{R}^n \to \Delta_k$ 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 $(A(\mathbf{X}), B(\mathbf{Y}))_{(\mathbf{X}, \mathbf{Y}) \sim \mathcal{G}_\rho^{\otimes n}}$ over $[k] \times [k]$. Observe that $\mathbb{E}_{(\mathbf{X}, \mathbf{Y}) \sim \mathcal{G}_\rho^{\otimes n}}[A_i(\mathbf{X}) \cdot B_j(\mathbf{Y})]$ is the probability of the event [*Alice outputs i and Bob outputs j*]. We succinctly write this expectation as $\langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes n}}$. In order to approximately preserve the joint distribution $(A(\mathbf{X}), B(\mathbf{Y}))_{(\mathbf{X}, \mathbf{Y}) \sim \mathcal{G}_\rho^{\otimes n}}$, it suffices to approximately preserve $\langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes n}}$ for each $(i, j) \in [k] \times [k]$.

Thus, to prove Theorem 1.4, 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 $\left\langle \widetilde{A}_i, \widetilde{B}_j \right\rangle_{\mathcal{G}_\rho^{\otimes n_0}} \approx_\varepsilon \left\langle A_i, B_j \right\rangle_{\mathcal{G}_\rho^{\otimes n}}$. 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 do a structural transformation on A and B to make them low-degree multilinear polynomials (see Section 2.2 for formal definitions) – such transformations are described in Sections 4 and 5. This however creates a new problem that the transformed A and B no longer map to Δ_k. Nevertheless, we show that after the said transformations we still have that the outputs of A and B are close to Δ_k in expected ℓ_2^2 distance (for now, let's informally denote this by $\text{dist}(A, \Delta_k)$). We show that this ensures 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 explicitly computable function D of k, d, and δ, along with polynomials $\widetilde{A} : \mathbb{R}^D \to \mathbb{R}^k$ and $\widetilde{B} : \mathbb{R}^D \to \mathbb{R}^k$ that δ-approximately preserves (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 in expected ℓ_2^2 distance, that is, $\text{dist}(A, \Delta_k)$ and $\text{dist}(B, \Delta_k)$.

We introduce a very simple and natural dimension-reduction procedure for low-degree multilinear polynomials over Gaussian space. Specifically, for an i.i.d. sequence of ρ-correlated Gaussians (a_1, b_1), $(a_2, b_2), \cdots, (a_D, b_D)$, we set

$$\widetilde{A}(\boldsymbol{a}) := A \left(\frac{M\boldsymbol{a}}{\|\boldsymbol{a}\|_2} \right) \qquad \text{and} \qquad \widetilde{B}(\boldsymbol{b}) := B \left(\frac{M\boldsymbol{b}}{\|\boldsymbol{b}\|_2} \right) \tag{1}$$

where M is a randomly sampled $N \times D$ matrix with i.i.d. standard Gaussian entries. Our main dimension-reduction theorem for low-degree polynomials is stated as follows,

---

[4]recall that we think of a vector valued function $A : \mathbb{R}^n \to \mathbb{R}^k$ as a tuple $(A_1, \ldots, A_k)$, where each $A_i : \mathbb{R}^n \to \mathbb{R}$

Theorem 1.7 (Dimension Reduction Over Gaussian Space). *Given parameters $k \geq 2$, $d \in \mathcal{Z}_{\geq 0}$, $\rho \in (0,1)$ and $\delta > 0$, there exists an* explicitly computable $D = D(d, k, \delta)$, *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 $\widetilde{A} : \mathbb{R}^D \to \mathbb{R}^k$ and $\widetilde{B} : \mathbb{R}^D \to \mathbb{R}^k$ as defined in* Equation (1). *With* probability *at least $1 - 3\delta$ over the random choice of $M \sim \mathcal{N}(0,1)^{N \times D}$, the following will hold:*

- *For every $i, j \in [k]$:* $\left| \left\langle A_i, B_j \right\rangle_{\mathcal{G}_\rho^{\otimes N}} - \left\langle \widetilde{A}_i, \widetilde{B}_j \right\rangle_{\mathcal{G}_\rho^{\otimes D}} \right| \leq \delta$.

- $\operatorname{dist}(\widetilde{A}, \Delta_k) \leq \sqrt{\delta}$ *and* $\operatorname{dist}(\widetilde{B}, \Delta_k) \leq \sqrt{\delta}$.

In particular, D is an explicit function upper bounded by $\exp\left(\operatorname{poly}\left(d, \log k, \log(\frac{1}{\delta})\right)\right)$.

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 1.4. The proof of Theorem 1.7 is obtained by combining Theorem 3.1 and Proposition 3.2 in Section 3.

Analogy with the Johnson-Lindenstrauss lemma. We will now highlight a few parallels between Theorem 1.7 and the proof of the Johnson-Lindenstrauss lemma. Suppose we have two unit vectors $u, v \in \mathbb{R}^n$. We wish to obtain a randomized transformation $\Psi_s : \mathbb{R}^n \to \mathbb{R}^D$ (where s is the random seed), such that, $\langle u, v \rangle \approx_\delta \langle \Psi_s(u), \Psi_s(v) \rangle$ holds with probability $1 - \delta$, over the randomness of seed s; note that here $\langle \cdot, \cdot \rangle$ denotes the inner product over $\mathbb{R}^n$ and $\mathbb{R}^D$ respectively. Indeed, there is such a transformation, namely, $\Psi_M(u) = \frac{M \cdot u}{\sqrt{D}}$ where $M \sim \mathcal{N}(0,1)^{\otimes D \times n}$. Let $F(M) = \langle \Psi_M(u), \Psi_M(v) \rangle$. Such a transformation satisfies that,

$$\mathbb{E}_M[F(M)] = \langle u, v \rangle \quad \text{and} \quad \operatorname{Var}_M(F(M)) = \frac{\langle u, v \rangle^2 + \|u\|_2^2 \|v\|_2^2}{D} \leq \frac{2}{D},$$

where we use that u and v are unit vectors. Thus, if we choose $D = 2/\delta^3$, then we can make the variance smaller than δ^3. Thereby, using Chebyshev's inequality, we get that with probability $1 - \delta$, it holds that $|\langle \Psi_M(u), \Psi_M(v) \rangle - \langle u, v \rangle| \leq \delta$. Thus, we have a *oblivious* randomized dimension reduction that reduced 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 $D = O(1/\varepsilon^2 \log(1/\delta))$ suffices to preserve the inner product upto an additive ε, with probability $1 - \delta$. However, we described the Chebyshev's inequality version as this is what our proof of Theorem 1.7 does at a high level.

The problem we are facing, although morally similar, is technically entirely different. We want the reduce the dimension of the domain of a pair of polynomials $A : \mathbb{R}^n \to \mathbb{R}$ and $B : \mathbb{R}^n \to \mathbb{R}$. For the moment, consider the transformation such that $\Psi_M(A) : \mathbb{R}^D \to \mathbb{R}$ is given by $A(Ma/\sqrt{D})$. Similarly, $\Psi_M(B) = B(Mb/\sqrt{D})$. Our proof of Theorem 1.7 proceeds along similar lines as the above proof of Johnson-Lindenstrauss Lemma, that is, by considering $F(M) = \mathbb{E}_{(a,b) \sim \mu^{\otimes D}}[\Psi_M(A)(a) \cdot \Psi_M(B)(b)]$, and proving bounds on $\mathbb{E}_M[F(M)]$ and $\operatorname{Var}(F(M))$. This turns out to be quite delicate! We don't even have $\mathbb{E}_M[F(M)] = \mathbb{E}_{(x,y) \sim \mu^{\otimes n}}[A(x) \cdot B(y)]$. What we do show is that,

$$\left| \mathbb{E}_M[F(M)] - \mathbb{E}_{(x,y) \sim \mu^{\otimes n}}[A(x) \cdot B(y)] \right| \leq o_D(1) \quad \text{and} \quad \operatorname{Var}_M(F(M)) \leq o_D(1),$$

that is, both are decreasing functions in D (with some dependence on d, which is the degree of A and B). Interestingly however, in the case of $d = 1$, it turns out that $F(M)$ is in fact an unbiased estimator. Indeed, this is not a coincidence! We leave it to the interested reader to figure out that in the case of $d = 1$, our tranformation is in fact identical to the above described Johnson-Lindenstrauss transformation on the n-dimensional space of Hermite coefficients of A and B.

Our actual transformation is slightly different, namely $\Psi_M(A) = A(Ma/\|a\|_2)$. This is to ensure the second point in Theorem 1.7, about preserving the closeness of the output of A to Δ_k. The proof gets a little more technical, but this is intuitively similar to the above transformation since $\|a\|_2$ is tightly concentrated around $\sqrt{D}$.

1.6 Related Work and Other Motivations

Information Theory. We point out that 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[5] which was studied by [MO04, MOR+06]. Moreover, recent works in the information theory community [KA16, BG15] 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 do not give the optimal bounds in general. Finally, the "non-interactive agreement distillation" problem studied by [BM11] can also be viewed as a particular case of the NIS setup.

    Randomness in Computation. As discussed in [GKS16b], one motivation for studying NIS problems stems from the study of the role of randomness in distributed computing. Specifically, recent works in cryptography [AC93, AC98, BS94, CN00, Mau93, RW05], quantum computing [Nie99, CDS08, DB14] and communication complexity [BGI14, CGMS14, GKS16a] study how the ability to solve various computational tasks gets affected by weakening the source of 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.

    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* [Sha56, Lov79] 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 [AL06]. Other examples of such open problems (which are more closely related to NIS) arise in the problems of *local state transformation of quantum entanglement* [Bei12, DB13], the problem of computing the *entangled value of a 2-prover 1-round game* (see for, e.g., [KKM+11] and also the open problems [ope]) and the problem of computing the *amortized value of parallel repetitions of a 2-prover 1-round game* [Raz98, Hol09, Rao11, Raz11, BHH+08]. Yet another example of a tensor-power problem is the task of computing the *amortized communication complexity of a communication problem*. Braverman-Rao [BR11] showed that this equals the information complexity of the communication problem, however the computability of information complexity was shown only recently [BS15].

    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.

1.7 Comparisons with recent works of De, Mossel & Neeman

Our main theorems Theorem 1.1 and Theorem 1.5 were proved by De, Mossel & Neeman [DMN17a, DMN17b] (only qualitatively, although with worse explicit bounds on n_0). Our work was inspired by [DMN17a, DMN17b] through several high-level ideas, such as the use of smoothing and multilinearization transformations (although these tranformations are technically different in our case, so we state and prove our lemmas from scratch). However, the authors hold the opinion 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 [DMN17a]): "*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*". In our work, the main intuition for the dimension reduction is an "oblivious" dimension reduction technique, much similar to the Johnson-Lindenstrauss Lemma, as described in Section 1.5.

---

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

While inspired by the works of De, Mossel & Neeman, we believe that our technique offers a fresh perspective on why it is possible to obtain explicit bounds for the above problems. Moreover, our bound on $n_0(\varepsilon)$ in both cases is "merely" exponential in the parameters of the problem, whereas, the bounds in the works of De et al. are not primitive recursive and have an Ackermann-type growth.

1.8 Outline of Proofs

Dimension Reduction for Polynomials. We being with describing the main ideas behind Theorem 1.7. For polynomials $A : \mathbb{R}^N \to \mathbb{R}$ and $B : \mathbb{R}^N \to \mathbb{R}$, we apply a second-moment argument to the random variable

$$F(M) := \langle A_M, B_M \rangle_{\mathcal{G}_\rho^{\otimes D}} ,$$

where A_M and B_M are the substitutions in Equation (1). Specifically, we compute bounds on the mean and variance of $F(M)$ (Lemma 3.3); the key point being that these bounds go to 0 as D gets larger. Thus, we can get an explicit bound on how large D needs to be in order to make the mean deviation and the variance small. Assuming Lemma 3.3, we simply apply Chebyshev's inequality in order to upper-bound the probability that this random variable significantly deviates from its mean.

Lemma 3.3 is the most technical and novel part of this work, and is proved in Appendix A. To prove these mean and variance bounds, we first analyze the case when A and B are multi-linear monomials (Appendix A.2). Then, via a simple application of hypercontractivity, we use the monomial calculations in order to obtain bounds on the mean and variance for general multilinear polynomials (Appendix A.3).

NIS from Gaussian Sources. We now turn to the proof of Theorem 1.4 (which immediately implies Theorem 1.1). We are given $A : \mathbb{R}^N \to \Delta_k$ and $B : \mathbb{R}^N \to \Delta_k$, and we want to construct functions $\widetilde{A} : \mathbb{R}^{n_0} \to \Delta_k$ and $\widetilde{B} : \mathbb{R}^{n_0} \to \Delta_k$ such that the joint distribution of $(\widetilde{A}, \widetilde{B})$ is close (in total variation distance) to that of (A, B).

For any $i, j \in [k]$, we consider the quantity $\mathbb{E}_{XY}[A_i(X) \cdot B_j(Y)]$ which is the probability of the event that [*Alice outputs i and Bob outputs j*]. Across multiple steps, we modify Alice's and Bob's strategies while approximately preserving this quantity for every i, j. Note that if we preserve this quantity for every i, j up to an additive ε/k^2, then this ensures that the joint distribution of Alice and Bob's outputs is preserved up to a total variation distance of ε. The first step is a smoothing operation (Lemma 4.1) that transforms A and B into polynomials $A^{(1)}, B^{(1)} : \mathbb{R}^N \to \mathbb{R}^k$ that are guaranteed to have (constant) degree d. In the second step, we use a multilinearization operation (Lemma 5.1) to convert $A^{(1)}, B^{(1)}$ into *multilinear* degree-d polynomials $A^{(2)}, B^{(2)} : \mathbb{R}^{Nt} \to \mathbb{R}^k$ (this operation increases the number of variables by a multiplicative t factor). Both these operations preserve the correlation $\langle A_i, B_j \rangle$, and keeps the expected ℓ_2^2 distance of A and B from Δ_k small. We then apply our main dimension-reduction procedure (Theorem 1.7) to obtain *constant-dimensional* functions $A^{(3)}, B^{(3)} : \mathbb{R}^{n_0} \to \mathbb{R}^k$ that preserve the structure and correlations of $A^{(2)}$ and $B^{(2)}$. At the final step, we set $\widetilde{A}$ and $\widetilde{B}$ to be the roundings of $A^{(3)}$ and $B^{(3)}$ (repsectively) to the closest functions mapping to Δ_k. Our analysis ensures that in each of the above steps, the two correlations of the two functions as well as their individual distances to the probability simplex are approximately preserved.

NIS from Arbitrary Discrete Sources. Our proof of Theorem 1.5 proceeds by a reduction to Theorem 1.4. The reduction uses the framework already developed in [GKS16b]) of using the invariance principle (Appendix E) as obtained in [Mos10, IM12] and a Regularity Lemma (Appendix D) inspired by [DSTW10]. We also need to use additional smoothing and multilinearization steps (the full details are in Section 7).

One key point about Theorem 1.4 that is crucial for this application is the "*oblivious*" nature of the dimension reduction. In the framework of [GKS16b], we need to apply Theorem 1.4 on a family of strategies $\left\{ A^{(1)}, \ldots, A^{(T)} \right\}$ and $\left\{ B^{(1)}, \ldots, B^{(T)} \right\}$, where each $A^{(i)}, B^{(i)} : \mathbb{R}^N \to \mathbb{R}^k$. That is, we want
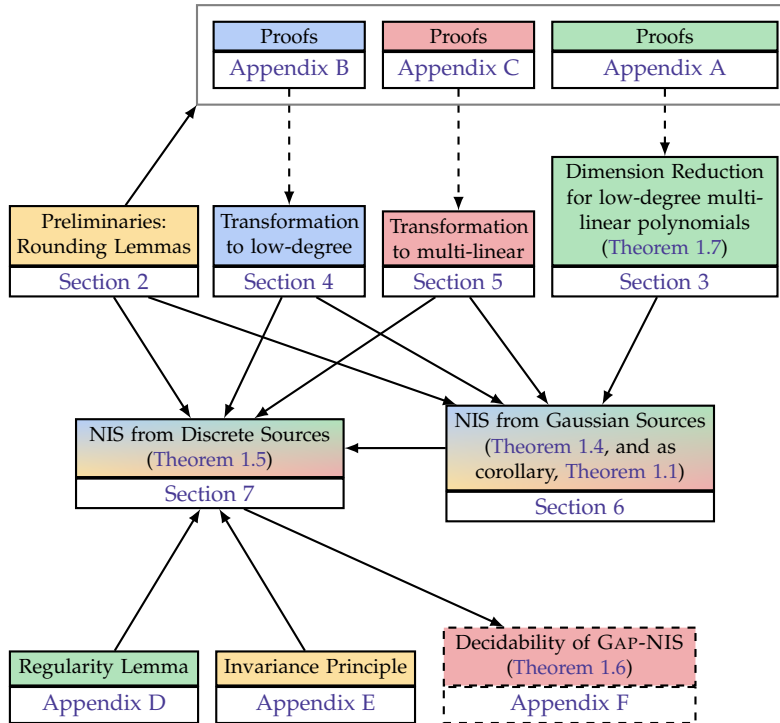
Figure 2: Organization of the paper, where arrows indicate the dependencies of the Sections/Appendices on each other (dashed lines indicate the dependencies that could be skipped on first reading; dashed box indicates an optional appendix)

to be able to reduce the dimensionality of all the $A^{(i)}$'s and $B^{(i)}$'s while simultaneously preserving the joint distribution $(A^{(i)}, B^{(j)})$ for at least a $(1 - \varepsilon)$ fraction of the pairs $(i, j) \in [T] \times [T]$. The *oblivious* randomized transformation in Theorem 1.4 gives us that the tranformation done on $A^{(i)}$ depends only on the random seed and not on which $B^{(j)}$ we are comparing it against. Moreover, this transformation works with "high" probability, so in expectation we get that the joint distribution is approximately preserved for atleast a $(1 - \varepsilon)$-fraction of the pairs $(i, j) \in [T] \times [T]$.

1.9 Organization of the Paper

In Section 2, we summarize some useful definitions, and prove a couple of simple lemmas that will be useful in the paper. In Section 3, we state our main technique of dimension reduction for polynomials, i.e. Theorem 1.7, with the key lemmas and proofs in Appendix A. In Sections 4 and 5 we describe the transformations to make functions low-degree and multilinear, with proofs deferred to Appendices B and C.

In Section 6, we prove Theorem 1.4, deriving Theorem 1.1 as a corollary. Finally, in Section 7, we prove Theorem 1.5, for which we need more tools such as the Regularity Lemma and the Invariance Principle, which we provide in Appendices D and E. Finally, for sake of completeness, the proof of Theorem 1.6 is provided in Appendix F.

To ease the task of navigating the paper, we provide an outline of the paper in Figure 2.

2 Preliminaries

2.1 Probability Spaces : Discrete and Gaussian

We will mostly use script letter $\mathcal{Z}$ to denote a finite set of size q, and μ will usually denote a probability distribution. We use small letters x, y, etc. to denote elements of $\mathcal{Z}$, and bold small letters $\boldsymbol{x}$, $\boldsymbol{y}$, etc. to denote elements in $\mathcal{Z}^n$. We use x_i, y_i to denote individual coordinates of $\boldsymbol{x}$, $\boldsymbol{y}$ respectively. For a probability space $(\mathcal{Z}, \mu)$, we will use the following definitions and notations,

- The pair $(\mathcal{Z}^n, \mu^{\otimes n})$ denotes the product space $\mathcal{Z} \times \mathcal{Z} \times \cdots \times \mathcal{Z}$ endowed with the product distribution.

- The support of μ is $\mathrm{Supp}(\mu) := \{x : \mu(x) > 0\}$. We assume w.l.o.g. that $\mathrm{Supp}(\mu) = \mathcal{Z}$.

- $\alpha(\mu)$ denotes the minimum non-zero probability atom in μ.

- $L^2(\mathcal{Z}, \mu)$ denotes the space of all functions from $\mathcal{Z}$ to $\mathbb{R}$.

- The inner product on $L^2(\mathcal{Z}, \mu)$ is denoted by $\langle f, g \rangle_\mu := \mathbb{E}_{x \sim \mu}[f(x)g(x)]$.

- The ℓ_p-norm by $\|f\|_p := \left[\mathbb{E}_{x \sim \mu} |f(x)|^p \right]^{1/p}$. Also, $\|f\|_\infty := \max_{\mu(x) > 0} |f(x)|$.

- For two distributions μ and ν, $d_{\mathrm{TV}}(\mu, \nu)$ is the total variation distance between μ and ν.

$(\mathcal{Z} \times \mathcal{Z}, \mu)$ denotes a joint probability space. We use μ_A and μ_B to denote the marginal distributions of μ. The correlation between functions acting on parts of a joint distribution is defined as follows.

Definition 2.1 (Correlation between strategies). *Let $(\mathcal{Z} \times \mathcal{Z}, \mu)$ be any joint probability space. For functions $A \in L^2(\mathcal{Z}, \mu_A)$ and $B \in L^2(\mathcal{Z}, \mu_B)$, the correlation between A and B over distribution μ is defined as,*

$$\langle A, B \rangle_\mu \;=\; \mathbb{E}_{(x,y) \sim \mu} \left[A(x) \cdot B(y) \right] .$$

More generally, if we have functions $A \in L^2(\mathcal{Z}^n, \mu_A^{\otimes n})$ and $B \in L^2(\mathcal{Z}^n, \mu_B^{\otimes n})$, the correlation between A and B over distribution $\mu^{\otimes n}$ is defined as,

$$\langle A, B \rangle_{\mu^{\otimes n}} \;=\; \mathbb{E}_{(\boldsymbol{x}, \boldsymbol{y}) \sim \mu^{\otimes n}} \left[A(\boldsymbol{x}) \cdot B(\boldsymbol{y}) \right] .$$

Remark. While $\langle A, B \rangle_{\mu^{\otimes n}}$ is the correlation over the joint distribution, the term $\langle A, A' \rangle_{\mu_A^{\otimes n}}$ is the correlation as defined earlier over the marginal space. To make this distinction clear, from now on, μ always refers to the joint distribution on $\mathcal{Z} \times \mathcal{Z}$, and we will use μ_A or μ_B to indicate distributions over $\mathcal{Z}$.
An important quantity associated to any joint distribution is that of the *maximal correlation coefficient*, which was first introduced by Hirschfeld [Hir35] and Gebelein [Geb41] and then studied by Rényi [Rén59].

Definition 2.2 (Maximal correlation). *Given a joint probability space $(\mathcal{Z} \times \mathcal{Z}, \mu)$, we define its* maximal correlation $\rho(\mathcal{Z} \times \mathcal{Z}; \mu)$ *(or simply $\rho(\mu)$) as follows,*

$$\rho(\mathcal{Z} \times \mathcal{Z}; \mu) \;=\; \sup_{f,g} \left\{ \langle f, g \rangle_\mu \;\middle|\; \begin{array}{ll} f : \mathcal{Z} \to \mathbb{R}, & \mathbb{E}[f] = \mathbb{E}[g] = 0 \\ g : \mathcal{Z} \to \mathbb{R}, & \mathrm{Var}(f) = \mathrm{Var}(g) = 1 \end{array} \right\}$$

Although the above definitions were stated for distributions over finite sets, they extend naturally to the case where $\mathcal{Z} = \mathbb{R}$, equipped with the Gaussian measure $\mathcal{N}(0, 1)$ (also denoted as γ_1). To distinguish between discrete and Gaussian spaces, we will use capital letters X, Y, etc. to denote elements of $\mathbb{R}$, and

bold letters $\mathbf{X}$, $\mathbf{Y}$, etc. to denote elements in $\mathbb{R}^n$. In this case, we use X_i, Y_i to denote individual coordinates of $\mathbf{X}$, $\mathbf{Y}$ respectively. The pair $(\mathbb{R}^n, \gamma_n)$ denotes the product space $\mathbb{R}^n$ endowed with the standard n-dimensional Gaussian measure. Unless explicitly mentioned otherwise, all the functions with domain $\mathbb{R}^n$ that we consider will be in $L^2(\mathbb{R}^n, \gamma_n)$, which is the space of ℓ_2-integrable functions with respect to the γ_n measure.

Over the space of reals, we will primarily consider the joint distribution of ρ-correlated Gaussians $(\mathbb{R} \times \mathbb{R}, \mathcal{G}_\rho)$. This is a 2-dimensional Gaussian distributions (X, Y), where X and Y are marginally distributed according to γ_1, with $\mathbb{E}[XY] = \rho$. It is well-known that the maximal correlation of $\mathcal{G}_\rho$ is ρ.

2.2 Fourier & Hermite Analysis

Fourier analysis for discrete product spaces. We recall some background in Fourier analysis that will be useful to us. Let $(\mathcal{Z}, \mu_A)$ be a finite probability space with $|\mathcal{Z}| = q$. Let $\mathcal{X}_0, \cdots, \mathcal{X}_{q-1} : \mathcal{Z} \to \mathbb{R}$ be an orthonormal basis for the space $L^2(\mathcal{Z}, \mu_A)$ with respect to the inner product $\langle ., . \rangle_{\mu_A}$. Furthermore, we require that this basis has the property that $\mathcal{X}_0 = \mathbf{1}$, i.e., the function that is identically 1 on every element of $\mathcal{Z}$.

For $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_n) \in \mathbb{Z}_q^n$, define $\mathcal{X}_{\boldsymbol{\sigma}} : \mathcal{Z}^n \to \mathbb{R}^n$ as follows,

$$\mathcal{X}_{\boldsymbol{\sigma}}(x_1, \ldots, x_n) = \prod_{i \in [n]} \mathcal{X}_{\sigma_i}(x_i)$$

It is easily seen that the functions $\left\{ \mathcal{X}_{\boldsymbol{\sigma}} : \boldsymbol{\sigma} \in \mathbb{Z}_q^n \right\}$ form an orthonormal basis for the product space $L^2(\mathcal{Z}^n, \mu_A^{\otimes n})$. Thus, every function $A \in L^2(\mathcal{Z}^n, \mu_A^{\otimes n})$ has a *Fourier expansion* given by

$$A(\boldsymbol{x}) = \sum_{\boldsymbol{\sigma} \in \mathbb{Z}_q^n} \widehat{A}(\boldsymbol{\sigma}) \mathcal{X}_{\boldsymbol{\sigma}}(\boldsymbol{x}) \,,$$

where $\widehat{A}(\boldsymbol{\sigma})$'s are the *Fourier coefficients* of A, which can be obtained as $\widehat{A}(\boldsymbol{\sigma}) = \langle A, \mathcal{X}_{\boldsymbol{\sigma}} \rangle_{\mu_A}$. Although we will work with an arbitrary (albeit fixed) basis, many of the important properties of the Fourier transform are basis-independent. For example, Parseval's identity states that $\|A\|_2^2 = \sum_{\boldsymbol{\sigma} \in \mathbb{Z}_{\geq 0}^n} \widehat{A}(\boldsymbol{\sigma})^2$.

For a joint probability space $(\mathcal{Z} \times \mathcal{Z}, \mu)$, we let $\mathcal{X}_0, \cdots, \mathcal{X}_{q-1} : \mathcal{Z} \to \mathbb{R}$ be an orthonormal basis for the space $L^2(\mathcal{Z}, \mu_A)$, and $\mathcal{Y}_0, \cdots, \mathcal{Y}_{q-1} : \mathcal{Z} \to \mathbb{R}$ be an orthonormal basis for the space $L^2(\mathcal{Z}, \mu_B)$. Although we could choose these basis independently, it is helpful to choose the basis such that, $\langle \mathcal{X}_i, \mathcal{Y}_j \rangle_\mu = \rho_i \cdot \mathbf{1}_{i=j}$, where $\rho_{q-1} \leq \cdots \leq \rho_1 = \rho(\mu)$ (here, $\rho(\mu)$ is the maximal correlation of μ as in Definition 2.2).

For $\boldsymbol{\sigma} \in \mathbb{Z}_q^n$, the *degree* of $\boldsymbol{\sigma}$ is denoted by $|\boldsymbol{\sigma}| \overset{\text{def}}{=} |\{i \in [n] : \sigma_i \neq 0\}|$. We say that the degree of a function[6] $A \in L^2(\mathcal{Z}^n, \mu_A^{\otimes n})$, denoted by $\deg(A)$, is the largest value of $|\boldsymbol{\sigma}|$ such that $\widehat{A}(\boldsymbol{\sigma}) \neq 0$.

Hermite Analysis for Gaussian space. Analogous to discrete spaces, 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 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 $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n) \in \mathbb{Z}_{\geq 0}^n$, define $H_{\boldsymbol{\sigma}} : \mathbb{R}^n \to \mathbb{R}$ as

$$H_{\boldsymbol{\sigma}}(\mathbf{X}) = \prod_{i=1}^{n} H_{\sigma_i}(X_i).$$

---

[6]we will interchangeably use the word *polynomial* to talk about any function in $L^2(\mathcal{A}^n, \mu^{\otimes n})$.

It is easily follows that the set $\{H_{\boldsymbol{\sigma}} : \boldsymbol{\sigma} \in \mathbb{Z}_{\geq 0}^n\}$ 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

$$A(\mathbf{X}) = \sum_{\sigma \in \mathbb{Z}_{\geq 0}^n} \widehat{A}(\boldsymbol{\sigma}) \cdot H_{\boldsymbol{\sigma}}(\mathbf{X}) ,$$

where the $\widehat{A}(\boldsymbol{\sigma})$'s are the *Hermite coefficients* of A, which can be obtained as $\widehat{A}(\boldsymbol{\sigma}) = \langle A, H_{\boldsymbol{\sigma}} \rangle_{\gamma_n}$. The degree of $\boldsymbol{\sigma}$ is defined as $|\boldsymbol{\sigma}| := \sum_{i \in [n]} \sigma_i$, and the degree of A is the largest $|\boldsymbol{\sigma}|$ for which $\widehat{A}(\boldsymbol{\sigma}) \neq 0$. Analogous to Boolean functions, we have Parseval's identity, that is, $\|A\|_2^2 = \sum_{\boldsymbol{\sigma} \in \mathbb{Z}_{\geq 0}^n} \widehat{A}(\boldsymbol{\sigma})^2$. We say that $A \in L^2(\mathbb{R}^n, \gamma_n)$ is *multilinear* if $\widehat{A}(\boldsymbol{\sigma})$ is non-zero only if $\sigma_i \in \{0, 1\}$ for all $i \in [n]$.

2.3 Vector-valued functions

We will extensively work with vector-valued functions. For any function $A : \mathcal{D} \to \mathbb{R}^k$ (for any domain $\mathcal{D}$), we will write $A = (A_1, \cdots, A_k)$, where $A_i : \mathcal{D} \to \mathbb{R}$ is the i-th coordinate of the output of A. That is, $A_i(x) = (A(x))_i$ for any $x \in \mathcal{D}$.

The definitions of Fourier analysis and Hermite analysis extend naturally to vector-valued functions. For $A : \mathbb{R}^n \to \mathbb{R}^k$, we use $\widehat{A}(\boldsymbol{\sigma})$ to denote the vector $\left(\widehat{A}_1(\boldsymbol{\sigma}), \ldots, \widehat{A}_k(\boldsymbol{\sigma}) \right)$. In this setting, $\|A\|_2^2 :=$ $\mathbb{E}_{\mathbf{X} \sim \gamma_n} \|A(\mathbf{X})\|_2^2 = \|A_1\|_2^2 + \cdots + \|A_k\|_2^2 = \sum_{\boldsymbol{\sigma} \in \mathbb{Z}_{\geq 0}^n} \|\widehat{A}(\boldsymbol{\sigma})\|_2^2$. Also, $\deg(A)$ is defined as $\max_{i \in [k]} \deg(A_i)$. Again, unless explicitly mentioned otherwise, 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, $\Delta_k = \{v \in \mathbb{R}^k : \|v\|_1 = 1\}$ is the set of probability distributions over $[k]$. While we consider vector-valued functions mapping to $\mathbb{R}^k$, we are primarily interested in functions which map to Δ_k. We use the rouding operator, defined as follows, in order to change the range of a function from $\mathbb{R}^k$ to Δ_k.

Definition 2.3 (Rounding operator)**.** *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, as k is fixed throughout this paper.*

As for vector-valued functions, we use $\mathcal{R}_i$ to denote 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 $\mathcal{R}_i(A)$.

Useful lemmas for ℓ_2-close strategies

An important relaxation in our work is to consider strategies that do not map to Δ_k, but instead map to $\mathbb{R}^k$. For such strategies to be meaningful, we will require that the outputs are *usually close* to Δ_k. In this case, we will be *rounding* them to the simplex Δ_k.

The following simple lemmas are going to be very useful. The first lemma says that if we modify the strategies of Alice and Bob such that they remain close in ℓ_2-distance, then the correlation between their strategies does not change significantly.

Lemma 2.4 (Close strategies in ℓ_2, have similar correlations)**.** *Given any joint probability space $(\mathcal{Z} \times \mathcal{Z}, \mu)$. Let $A, \widetilde{A} \in L^2(\mathcal{Z}^n, \mu_A^{\otimes n})$ and $B, \widetilde{B} \in L^2(\mathcal{Z}^n, \mu_B^{\otimes n})$ such that $\|A\|_2, \|\widetilde{A}\|_2, \|B\|_2, \|\widetilde{B}\|_2 \leq 1$. If $\|A - \widetilde{A}\|_2 \leq \varepsilon$ and $\|B - \widetilde{B}\|_2 \leq \varepsilon$, then it holds that,*

$$\left| \left\langle \widetilde{A}, \widetilde{B} \right\rangle_{\mu^{\otimes n}} - \langle A, B \rangle_{\mu^{\otimes n}} \right| \leq 2\varepsilon$$

Proof. The proof follows very easily from the Cauchy-Schwarz inequality. In particular,

$$
\begin{aligned}
\left| \left\langle \widetilde{A}, \widetilde{B} \right\rangle_{\mu^{\otimes n}} - \left\langle A, B \right\rangle_{\mu^{\otimes n}} \right|
&= \left| \left\langle (\widetilde{A} - A), \widetilde{B} \right\rangle_{\mu^{\otimes n}} + \left\langle A, (\widetilde{B} - B) \right\rangle_{\mu^{\otimes n}} \right| \\
&\leq \left| \left\langle (\widetilde{A} - A), \widetilde{B} \right\rangle_{\mu^{\otimes n}} \right| + \left| \left\langle A, (\widetilde{B} - B) \right\rangle_{\mu^{\otimes n}} \right| \\
&\leq \left\| \widetilde{A} - A \right\|_2 \cdot \left\| \widetilde{B} \right\|_2 + \left\| \widetilde{B} - B \right\|_2 \cdot \left\| A \right\|_2 \qquad \ldots \text{(Cauchy-Schwarz inequality)} \\
&\leq 2\varepsilon . \hspace{6cm} \square
\end{aligned}
$$

The second lemma 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.

Lemma 2.5. *Given any joint probability space $(\mathcal{Z} \times \mathcal{Z}, \mu)$. Let $A : \mathcal{Z}^n \to \mathbb{R}^k$ and $\widetilde{A} : \mathcal{Z}^n \to \mathbb{R}^k$ such that $\|A\|_2, \|\widetilde{A}\|_2 \leq 1$. Then, for the rounding operator $\mathcal{R} : \mathbb{R}^k \to \Delta_k$, it holds that,*

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

Proof. The proof follows very easily from a triangle inequality. In particular,

$$
\begin{aligned}
\left\| \mathcal{R}(\widetilde{A}) - \widetilde{A} \right\|_2 &\leq \left\| \mathcal{R}(A) - \widetilde{A} \right\|_2 && \text{(since } \mathcal{R}(\widetilde{A}(\boldsymbol{x})) \text{ closest in } \Delta_k \text{ to } \widetilde{A}(\boldsymbol{x})\text{)} \\
&\leq \left\| \mathcal{R}(A) - A \right\|_2 + \left\| A - \widetilde{A} \right\|_2 && \text{(Triangle inequality)} \hspace{2cm} \square
\end{aligned}
$$

3 Dimension Reduction for Low-Degree Multilinear Polynomials

In this section, we present our main technique which is a dimension reduction for low-degree multilinear polynomials over Gaussian space, and prove Theorem 1.7, which is obtained immediately as a combination of Theorem 3.1 and Proposition 3.2 stated below.

Theorem 3.1. *Given parameters $d \in \mathbb{Z}_{>0}$, $\rho \in [0,1]$ and $\delta > 0$, there exists an explicitly computable $D = D(d, \delta)$, 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, such that $\|A\|_2, \|B\|_2 \leq 1$. For column vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^D$ and $M \in \mathbb{R}^{N \times D}$, define the functions $A_M : \mathbb{R}^D \to \mathbb{R}$ and $B_M : \mathbb{R}^D \to \mathbb{R}$ as

$$
A_M(\boldsymbol{a}) = A\left(\frac{M\boldsymbol{a}}{\|\boldsymbol{a}\|_2} \right) \qquad \text{and} \qquad B_M(\boldsymbol{b}) = B\left(\frac{M\boldsymbol{b}}{\|\boldsymbol{b}\|_2} \right)
$$

Sample $M \sim \mathcal{N}(0,1)^{\otimes(N \times D)}$. Then, with probability at least $1 - \delta$ over the choice of M, it holds that,

$$
\left| \langle A_M, B_M \rangle_{\mathcal{G}_\rho^{\otimes D}} - \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes N}} \right| < \delta .
$$

In particular, one may take $D = \frac{d^{O(d)}}{\delta^4}$.

In other words, for a typical choice of $M \sim \mathcal{N}(0,1)^{\otimes(N \times D)}$, the correlation between A and B is approximately preserved if we replace $(\mathbf{X}, \mathbf{Y}) \sim \mathcal{G}_\rho^{\otimes N}$ by $(M\boldsymbol{a}/\|\boldsymbol{a}\|_2, M\boldsymbol{b}/\|\boldsymbol{b}\|_2)$, where $(\boldsymbol{a}, \boldsymbol{b}) \sim \mathcal{G}_\rho^{\otimes D}$. Intuitively, M can be thought of as a means to "stretch" D 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 which shows 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 also have outputs close to the simplex. More formally,

Proposition 3.2. *Let $A : \mathbb{R}^N \to \mathbb{R}^k$ and $B : \mathbb{R}^N \to \mathbb{R}^k$, such that $\|\mathcal{R}(A) - A\|_2, \|\mathcal{R}(B) - B\|_2 \leq \delta$. Let $A_M : \mathbb{R}^D \to \mathbb{R}^k$ and $B_M : \mathbb{R}^D \to \mathbb{R}^k$ be defined analogously to Theorem 3.1. For $M \sim \mathcal{N}(0,1)^{\otimes(N \times D)}$, with probability at least $1 - 2\delta$, it holds that,*

$$\|\mathcal{R}(A_M) - A_M\|_2 \leq \sqrt{\delta} \qquad \text{and} \qquad \|\mathcal{R}(B_M) - B_M\|_2 \leq \sqrt{\delta} .$$

Proof. We first observe that for any fixed $\boldsymbol{a} \in \mathbb{R}^D$, the distribution of $\frac{M\boldsymbol{a}}{\|\boldsymbol{a}\|_2}$ is identical to that of a standard N-variate Gaussian distribution. Thus, we immediately have that,

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

Alternately,

$$\mathbb{E}_M \|\mathcal{R}(A_M) - A_M\|_2^2 = \|\mathcal{R}(A) - A\|_2^2 \leq \delta^2$$

Thus, using Markov's inequality, we get that with probability at least $1 - \delta$,

$$\|\mathcal{R}(A_M) - A_M\|_2 \leq \sqrt{\delta}.$$

We similarly argue for B_M, and a union bound completes the proof. $\qquad \square$

To prove Theorem 3.1, we primarily use the second moment method (i.e., Chebyshev's inequality). In particular, let $F(M)$ be defined as,

$$F(M) \stackrel{\text{def}}{=} \langle A_M, B_M \rangle_{\mathcal{G}_\rho^{\otimes D}}$$

The most technical part of this work is to show sufficiently good bounds on the mean and variance of $F(M)$ for a random choice of $M \sim \mathcal{N}(0,1)^{\otimes(N \times D)}$, given by the following lemma.

Lemma 3.3. *(Bound on Mean & Variance). Given parameters d and δ, there exists $D := D(d, \delta)$ such that the following holds: For $M \sim \mathcal{N}(0,1)^{\otimes(N \times D)}$,*

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

$$\mathrm{Var}_M (F(M)) \leq \delta \qquad \textbf{(Variance bound)}$$

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

The proof of Lemma 3.3 appears in Appendix A. Assuming Lemma 3.3, we can easily prove Theorem 3.1.

Proof of Theorem 3.1. We invoke Lemma 3.3 with parameters d and $\delta^2/2$, and we get a choice of $D = \frac{d^{O(d)}}{\delta^4}$. Using Chebyshev's inequality and using the Variance bound in Lemma 3.3, we have that for any $\eta > 0$,

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

Using the triangle inequality, and the Mean bound in Lemma 3.3, we get

$$\Pr_M \left[\left| F(M) - \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes N}} \right| > \delta \right]$$

$$\leq \Pr_M \left[|F(M) - \mathbb{E}_M F(M)| + \left| \mathbb{E}_M F(M) - \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes N}} \right| > \delta \right]$$

$$\leq \Pr_M \left[|F(M) - \mathbb{E}_M F(M)| > \delta - \delta^2 \right]$$

$$\leq \delta. \qquad \square$$

4 Reduction from General Polynomials to Low-Degree Polynomials

In this section, we state a lemma which says that functions in $L^2(\mathbb{R}^n, \gamma_n)$, and more generally in $L^2(\mathcal{Z}^n, \mu^{\otimes n})$, can be converted to low-degree polynomials while approximately preserving correlations with other functions and also not deviating much from the simplex Δ_k. This technique is considered quite standard and was also used in [GKS16b, DMN17a, DMN17b] for the same reason. For completeness, we provide the proof in Appendix B.

Lemma 4.1 (Main Smoothing Lemma). *Let $\rho \in [0,1], \delta > 0, k \in \mathbb{N}$ be any given constant parameters. There exists an explicit $d = d(\rho, k, \delta)$ such that the following holds:*

(Correlated Discrete Hypercube): *Let $(\mathcal{Z} \times \mathcal{Z}, \mu)$ be a joint probability space, with $\rho(\mathcal{Z}, \mathcal{Z}; \mu) = \rho$. Let $A : \mathcal{Z}^n \to \mathbb{R}^k$ and $B : \mathcal{Z}^n \to \mathbb{R}^k$, such that, for any $j \in [k] : \mathsf{Var}(A_j), \mathsf{Var}(B_j) \leq 1$.*
Then, there exist functions $A^{(1)} : \mathcal{Z}^n \to \mathbb{R}^k$ and $B^{(1)} : \mathcal{Z}^n \to \mathbb{R}^k$ such that statements 1-4 below hold.

(Correlated Gaussian): *Let $A : \mathbb{R}^n \to \mathbb{R}^k$ and $B : \mathbb{R}^n \to \mathbb{R}^k$, such that, for any $j \in [k] : \mathsf{Var}(A_j), \mathsf{Var}(B_j) \leq 1$.*
Then, there exist functions $A^{(1)} : \mathbb{R}^n \to \mathbb{R}^k$ and $B^{(1)} : \mathbb{R}^n \to \mathbb{R}^k$ such that statements 1-4 below hold.

1. *$A^{(1)}$ and $B^{(1)}$ have degree at most d.*

2. *For any $i \in [k]$, it holds that $\mathsf{Var}(A_i^{(1)}) \leq \mathsf{Var}(A_i) \leq 1$ and $\mathsf{Var}(B_i^{(1)}) \leq \mathsf{Var}(B_i) \leq 1$.*

3. *$\left\| \mathcal{R}(A^{(1)}) - A^{(1)} \right\|_2 \leq \left\| \mathcal{R}(A) - A \right\|_2 + \delta$ and $\left\| \mathcal{R}(B^{(1)}) - B^{(1)} \right\|_2 \leq \left\| \mathcal{R}(B) - B \right\|_2 + \delta$*

4. *For every $i, j \in [k]$,*

$$\left| \left\langle A_i^{(1)}, B_j^{(1)} \right\rangle_{\mu^{\otimes n}} - \left\langle A_i, B_j \right\rangle_{\mu^{\otimes n}} \right| \leq \frac{\delta}{\sqrt{k}} \qquad \textbf{(Correlated Discrete Hypercube)}$$

$$\left| \left\langle A_i^{(1)}, B_j^{(1)} \right\rangle_{\mathcal{G}_\rho^{\otimes n}} - \left\langle A_i, B_j \right\rangle_{\mathcal{G}_\rho^{\otimes n}} \right| \leq \frac{\delta}{\sqrt{k}} \qquad \textbf{(Correlated Gaussian)}$$

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

5 Reduction from General Polynomials to Multilinear Polynomials

In this section, we present a simple technique to convert low-degree (non-multilinear) polynomials into multilinear polynomials, without hurting the correlation, albeit increasing the number of variables slightly. This step is of a similar nature as Lemma 4.1. In particular, note that the conditions 1, 2, 3, 5 in the following lemma are also present in Lemma 4.1. This idea also appears in [DMN17a, DMN17b]. Since the exact statement we desire is slightly different, we provide a proof for completeness in Appendix C.

Lemma 5.1 (Multi-linearization Lemma). *Let $\rho \in [0,1], \delta > 0, d, k \in \mathbb{Z}_{\geq 0}$ be any given constant parameters. 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 $j \in [k] : \mathsf{Var}(A_j), \mathsf{Var}(B_j) \leq 1$.
Then, there exist functions $A^{(1)} : \mathbb{R}^{nt} \to \mathbb{R}^k$ and $B^{(1)} : \mathbb{R}^{nt} \to \mathbb{R}^k$ such that the following hold

1. *$A^{(1)}$ and $B^{(1)}$ are multilinear with degree d.*

2. *For any $j \in [k]$, it holds that $\mathsf{Var}(A_j^{(1)}) \leq \mathsf{Var}(A_j) \leq 1$ and $\mathsf{Var}(B_j^{(1)}) \leq \mathsf{Var}(B_j) \leq 1$.*

3. *$\left\| \mathcal{R}(A^{(1)}) - A^{(1)} \right\|_2 \leq \left\| \mathcal{R}(A) - A \right\|_2 + \delta$ and $\left\| \mathcal{R}(B^{(1)}) - B^{(1)} \right\|_2 \leq \left\| \mathcal{R}(B) - B \right\|_2 + \delta$*

4. *For any $\ell \in [nt]$ and $j \in [k]$, it holds that $\mathrm{Inf}_\ell(A_j^{(1)}) \leq \delta$ and $\mathrm{Inf}_\ell(B_j^{(1)}) \leq \delta$.*

5. *For every $i, j \in [k]$,*

$$\left| \left\langle A_i^{(1)}, B_j^{(1)} \right\rangle_{\mathcal{G}_\rho^{\otimes nt}} - \left\langle A_i, B_j \right\rangle_{\mathcal{G}_\rho^{\otimes n}} \right| \leq \frac{\delta}{\sqrt{k}}$$

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

6 Non-Interactive Simulation from Correlated Gaussian Sources

In this section, we show our main theorem regarding non-interactive simulation from Correlated Gaussian sources. That is, we show Theorem 1.4 (restated below as Theorem 6.1), and Theorem 1.1 (which follows immediately as Corollary 6.2).

Theorem 6.1. *Given parameters $k \geq 2$, $\rho \in [0,1]$ and $\varepsilon > 0$, there exists an explicitly computable $n_0 = n_0(\rho, k, \varepsilon)$ such that the following holds:*
For any N, and any $A : \mathbb{R}^N \to \Delta_k$ and $B : \mathbb{R}^N \to \Delta_k$, there exist functions $\widetilde{A} : \mathbb{R}^{n_0} \to \Delta_k$ and $\widetilde{B} : \mathbb{R}^{n_0} \to \Delta_k$ such that,

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

Moreover, there exists $d_0 = d_0(\rho, k, \varepsilon)$ for which there are degree-d_0 polynomials $A_0 : \mathbb{R}^{n_0} \to \mathbb{R}^k$ and $B_0 : \mathbb{R}^{n_0} \to \mathbb{R}^k$, such that, $\widetilde{A}(\boldsymbol{a}) = \mathcal{R}\left(A_0\left(\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2}\right)\right)$ and $\widetilde{B}(\boldsymbol{b}) = \mathcal{R}\left(B_0\left(\frac{\boldsymbol{b}}{\|\boldsymbol{b}\|_2}\right)\right)$.
In particular, one may take $n_0 = \exp\left(\mathrm{poly}\left(k, \frac{1}{\varepsilon}, \frac{1}{1-\rho}\right)\right)$ and $d_0 = \mathrm{poly}\left(k, \frac{1}{\varepsilon}, \frac{1}{1-\rho}\right)$.[7]

In fact, 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 simultaneously with the same random seed, if $A = B$, then the transformation gives $\widetilde{A} = \widetilde{B}$ as well.

Before proving the theorem, we remark that it immediately implies the desired statement needed to prove a dimension bound on ε-approximate noise stable function (i.e. Theorem 1.1).

Corollary 6.2. *Given parameters $k \geq 2$, $\rho \in [0,1]$ and $\varepsilon > 0$, there exists an explicitly computable $n_0 = n_0(\rho, k, \varepsilon)$ such that the following holds:*
Let $f : \mathbb{R}^N \to [k]$. Then, there exists a function $\widetilde{f} : \mathbb{R}^{n_0} \to \Delta_k$ such that

1. $\left\| \mathbb{E}[f] - \mathbb{E}[\widetilde{f}] \right\|_1 \leq \varepsilon.$

2. $\mathbb{E}\left[\left\langle \widetilde{f}, U_\rho \widetilde{f} \right\rangle \right] \geq \mathbb{E}[\langle f, U_\rho f \rangle] - \varepsilon.$

Moreover, there exists $d_0 = d_0(\rho, k, \varepsilon)$ for which there is a degree-d_0 polynomial $g : \mathbb{R}^{n_0} \to \mathbb{R}^k$, such that, $\widetilde{f}(\boldsymbol{a}) = \mathcal{R}\left(g\left(\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_2}\right)\right)$. In particular, one may take $n_0 = \exp\left(\mathrm{poly}\left(k, \frac{1}{\varepsilon}, \frac{1}{1-\rho}\right)\right)$ and $d_0 = \mathrm{poly}\left(k, \frac{1}{\varepsilon}, \frac{1}{1-\rho}\right)$.

Proof. We invoke Theorem 6.1 with both A and B as f and with parameter $\varepsilon/2$, thereby obtaining functions $\widetilde{A}$, $\widetilde{B}$ which map $\mathbb{R}^{n_0} \to \Delta_k$. Note that since $A = B = f$, we also have $\widetilde{A} = \widetilde{B} = \widetilde{f}$. We get both our desired goals by observing that both $\left\| \mathbb{E}[f] - \mathbb{E}[\widetilde{f}] \right\|_1$ and $\left| \mathbb{E}\left[\left\langle \widetilde{f}, U_\rho \widetilde{f} \right\rangle \right] - \mathbb{E}[\langle f, U_\rho f \rangle] \right|$ are upper bounded by at most twice the total variation distance between the distributions $(A, B)_{\mathbf{X}, \mathbf{Y}}$ and $(\widetilde{A}, \widetilde{B})_{\boldsymbol{a}, \boldsymbol{b}}$. $\square$

---

[7]the details of the exact value of n_0 could be inferred from combining the bounds across various lemmas used. We skip it for brevity, and instead stress on the qualitative nature of the bound.
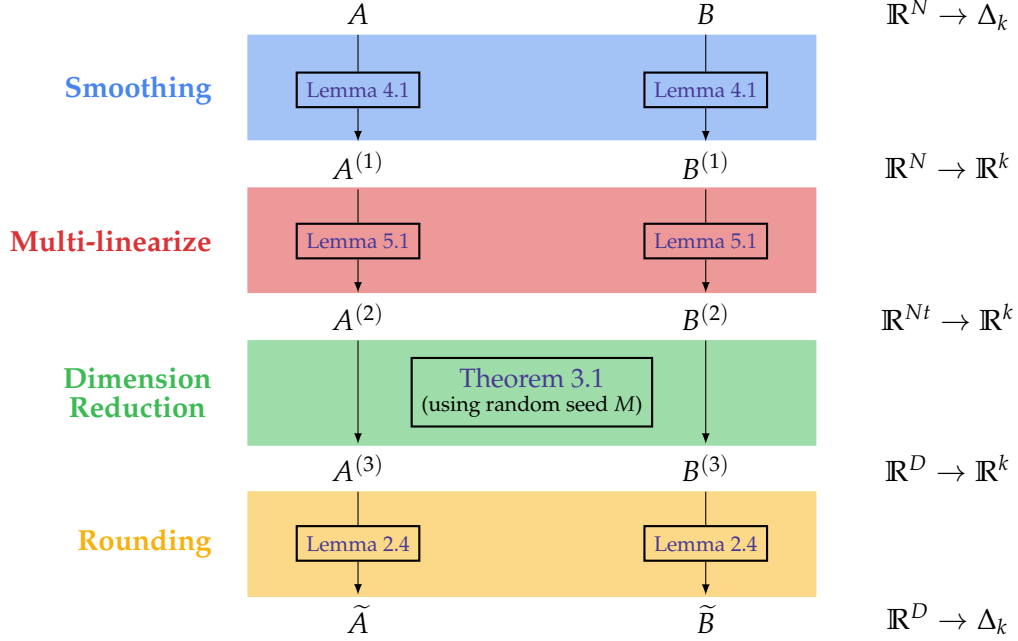
Figure 3: Transformations for Non-interactive simulation from Correlated Gaussian Sources

The rest of this section is dedicated to proving Theorem 6.1. We first provide the main intuition behind the proof. Starting with functions $A : \mathbb{R}^N \to \Delta_k$ and $B : \mathbb{R}^N \to \Delta_k$, we would have liked to directly apply our dimension reduction. That would have entailed having $\widetilde{A}(\boldsymbol{a}) = A(M\boldsymbol{a}/\|\boldsymbol{a}\|_2)$ and $\widetilde{B}(\boldsymbol{b}) = B(M\boldsymbol{b}/\|\boldsymbol{b}\|_2)$, where $(\boldsymbol{a}, \boldsymbol{b}) \sim \mathcal{G}_\rho^{\otimes n_0}$ and M is a $N \times n_0$ matrix with entries sampled randomly from $\mathcal{N}(0, 1)$. This already gives us that the range of $\widetilde{A}$ and $\widetilde{B}$ is Δ_k, since that was the range of A and B as well. Thus, if our dimension reduction were to approximately preserve correlations, i.e. $\langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes N}} \approx \langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes n_0}}$ for all $i, j \in [k]$ with high probability over M, we would have been done! However our actual dimension reduction (Theorem 3.1) works only for low-degree multilinear polynomials A and B. To get around this, we first apply the Smoothing (Lemma 4.1) and Multilinearization (Lemma 5.1) transformations that make A and B both low-degree and multilinear, and then subsequently apply our dimension reduction (Theorem 3.1). Unfortunately, this creates a new problem, that after these transformations, the range is no longer Δ_k, but is instead $\mathbb{R}^k$. Nevertheless, we do have that these transformations ensure that the functions still output something "close" to the simplex Δ_k. This allows us to use the standard rounding operation to get the range as Δ_k again (using Lemma 2.4). An overview of the transformations done is presented in Figure 3.

Proof of Theorem 6.1. For any $i, j \in [k]$, we focus on the quantity $\langle A_i, B_j \rangle_{\mathcal{G}_\rho^{\otimes n}}$ which is the probability of the event that [*Alice outputs i and Bob outputs j*]. Through several steps, we modify Alice's and Bob's strategies, while approximately preserving this quantity for every i, j. Note that if we preserve the probability that Alice outputs i and Bob outputs j for every i, j upto an additive ε/k^2, this implies that we would preserve the joint distribution of Alice and Bob's outputs up to ℓ_1-distance of ε.

We transform A and B through each of the following steps, as illustrated in Figure 3. At each step, we approximately preserve the correlation $\langle A_i, B_j \rangle$ for every $i, j \in [k]$. Additionally, in each step $\|\mathcal{R}(A) - A\|_2$ and $\|\mathcal{R}(B) - B\|_2$ also doesn't increase significantly. Note that, to begin with the range of A and B is Δ_k and hence $\|\mathcal{R}(A) - A\|_2 = \|\mathcal{R}(B) - B\|_2 = 0$.

1. **Smoothing**: We apply Lemma 4.1 with parameter δ, setting $d = d(\rho, k, \delta)$ as required, on A and B to get the smoothened versions $A^{(1)} : \mathbb{R}^N \to \mathbb{R}^k$ and $B^{(1)} : \mathbb{R}^N \to \mathbb{R}^k$. This guarantees that $A^{(1)}$

18

and $B^{(1)}$ have degree at most d. Moreover, we have that for every $i, j \in [k]$,

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

Additionally,

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

Similarly, we also have that,

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

2. **Multilinearization**: We apply Lemma 5.1 with parameter δ, setting $t = t(d, k, \delta)$ as required, on $A^{(1)}$ and $B^{(1)}$ to get the multilinearized versions $A^{(2)} : \mathbb{R}^{Nt} \to \mathbb{R}^k$ and $B^{(2)} : \mathbb{R}^{Nt} \to \mathbb{R}^k$. This guarantees that both $A^{(2)}$ and $B^{(2)}$ are multilinear and have degree at most d, albeit over a slightly larger number of variables. We get,

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

Additionally,

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

Similarly, we also have that,

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

3. **Dimension reduction**: We apply Theorem 3.1 with parameter δ/k^2, setting $D = D(d, \rho, \delta/k^2)$ as required, on individual coordinates of $A^{(2)}$ and $B^{(2)}$ to obtain functions $A^{(3)} : \mathbb{R}^D \to \mathbb{R}^k$ and $B^{(3)} : \mathbb{R}^D \to \mathbb{R}^k$. Taking a union bound, we have that with probability at least $1 - \delta$, it holds for every $i, j \in [k]$ that,

$$\left| \left\langle A_i^{(3)}, B_j^{(3)} \right\rangle_{\mathcal{G}_\rho^{\otimes D}} - \left\langle A_i^{(2)}, B_j^{(2)} \right\rangle_{\mathcal{G}_\rho^{\otimes Nt}} \right| \leq \delta \tag{4}$$

From Proposition 3.2, we have that with probability $1 - 4\delta$,

$$\left\| \mathcal{R}(A^{(3)}) - A^{(3)} \right\|_2 \leq \sqrt{\left\| \mathcal{R}(A^{(2)}) - A^{(2)} \right\|_2} \leq \sqrt{2\delta}$$

$$\left\| \mathcal{R}(B^{(3)}) - B^{(3)} \right\|_2 \leq \sqrt{\left\| \mathcal{R}(B^{(1)}) - B^{(1)} \right\|_2} \leq \sqrt{2\delta}$$

Note that this is the only randomized procedure in the entire transformation. This reduction succeeds in obtaining the three constraints above with probability at least $1 - 5\delta$.

4. **Rounding to Δ_k**: We obtain our final functions as $\widetilde{A} = \mathcal{R}(A^{(3)})$ and $\widetilde{B} = \mathcal{R}(B^{(3)})$. Note that $\left\| \mathcal{R}(A^{(3)}) - A^{(3)} \right\|_2, \left\| \mathcal{R}(B^{(3)}) - B^{(3)} \right\|_2 \leq \sqrt{2\delta}$, and hence we have for any $i, j \in [k]$ that

$$\left\| \widetilde{A}_i - A_i^{(3)} \right\|_2 \leq \sqrt{2\delta} \qquad \text{and} \qquad \left\| \widetilde{B}_j - B_j^{(3)} \right\|_2 \leq \sqrt{2\delta} .$$

Hence we can invoke Lemma 2.4, to conclude that for $\widetilde{A} = \mathcal{R}(A^{(3)})$ and $\widetilde{B} = \mathcal{R}(B^{(3)})$,

$$\left| \left\langle \widetilde{A}_i, \widetilde{B}_j \right\rangle_{\mathcal{G}_\rho^{\otimes D}} - \left\langle A_i^{(3)}, B_j^{(3)} \right\rangle_{\mathcal{G}_\rho^{\otimes D}} \right| \leq 2\sqrt{\delta}. \tag{5}$$

Finally, we choose $n_0 = D$ and $d_0 = d$ as obtained above. Note that we started with functions $A : \mathbb{R}^N \to \Delta_k$ and $B : \mathbb{R}^N \to \Delta_k$ and we 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 $i, j \in [k]$, we have by combining Equations (2) to (5) that,

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

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,

$$d_{\mathrm{TV}}((A(\mathbf{X}), B(\mathbf{Y}))_{\mathbf{X}, \mathbf{Y}}, (\widetilde{A}(\boldsymbol{a}), \widetilde{B}(\boldsymbol{b}))_{\boldsymbol{a}, \boldsymbol{b}}) \leq \varepsilon ,$$

where recall that $(\mathbf{X}, \mathbf{Y}) \sim \mathcal{G}_\rho^{\otimes N}$ and $(\boldsymbol{a}, \boldsymbol{b}) \sim \mathcal{G}_\rho^{\otimes n_0}$. It is easy to see that the parameters work out to be

$$d_0 = d = \widetilde{O}\left(\frac{k^{4.5}}{\varepsilon^2(1-\rho)} \right) ,$$

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

$\square$

7   Non-Interactive Simulation from Arbitrary Discrete Sources

In this section we prove our main theorem regarding non-interactive simulation from arbitrary discrete sources. That is, we prove Theorem 1.5 (restated below as Theorem 7.1).

Theorem 7.1. *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 $n_0 = n_0(\mu, k, \varepsilon)$ such that the following holds:*
   Let $A : \mathcal{Z}^N \to \Delta_k$ and $B : \mathcal{Z}^N \to \Delta_k$. Then there exist functions $\widetilde{A} : \mathcal{Z}^{n_0} \to \Delta_k$ and $\widetilde{B} : \mathcal{Z}^{n_0} \to \Delta_k$ such that,

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

In particular, n_0 is an explicit function upper bounded by $\exp\left(\mathrm{poly}\left(k, \frac{1}{\varepsilon}, \frac{1}{1-\rho}, \log\left(\frac{1}{\alpha}\right) \right) \right)$[8], where $\alpha = \alpha(\mu)$ is the smallest atom in μ and $\rho = \rho(\mu)$ is the maximal correlation of μ.

Note, that this theorem is, in a way, a generalization of Theorem 6.1, where $\mathcal{Z}$ was $\mathbb{R}$ and the distribution μ was $\mathcal{G}_\rho$. On the other hand, this theorem is only for the case when $\mathcal{Z}$ is a finite set, so in this sense it is incomparable to Theorem 6.1.

Proof Overview: The proof works by a reduction to Theorem 6.1. This reduction is done along the same framework as introduced in [GKS16b]. We first apply a Smoothing operation (Lemma 4.1) similar to the Gaussian case, to make the functions A and B have *low-degree*. Next, we will apply a Regularity Lemma (Lemma D.3), to identify a constant sized subset of coordinates, such that for a random fixing of these coordinates, the restricted function is low influential on the remaining coordinates. This allows us to apply the invariance principle (Lemma E.1) to replace the coordinates of Alice and Bob on the remaining coordinates by ρ-correlated Gaussians. We now use Theorem 6.1 to reduce the number of coordinates of ρ-correlated Gaussians needed. Finally, we wish to get the strategies to use samples from μ instead of $\mathcal{G}_\rho$, by simulating the correlated multivariate Gaussians using a bounded number of samples of μ. However,

---

[8]the details of the exact value of n_0 could be inferred from combining the bounds across various lemmas used. We skip it for brevity, and instead stress on the qualitative nature of the bound.
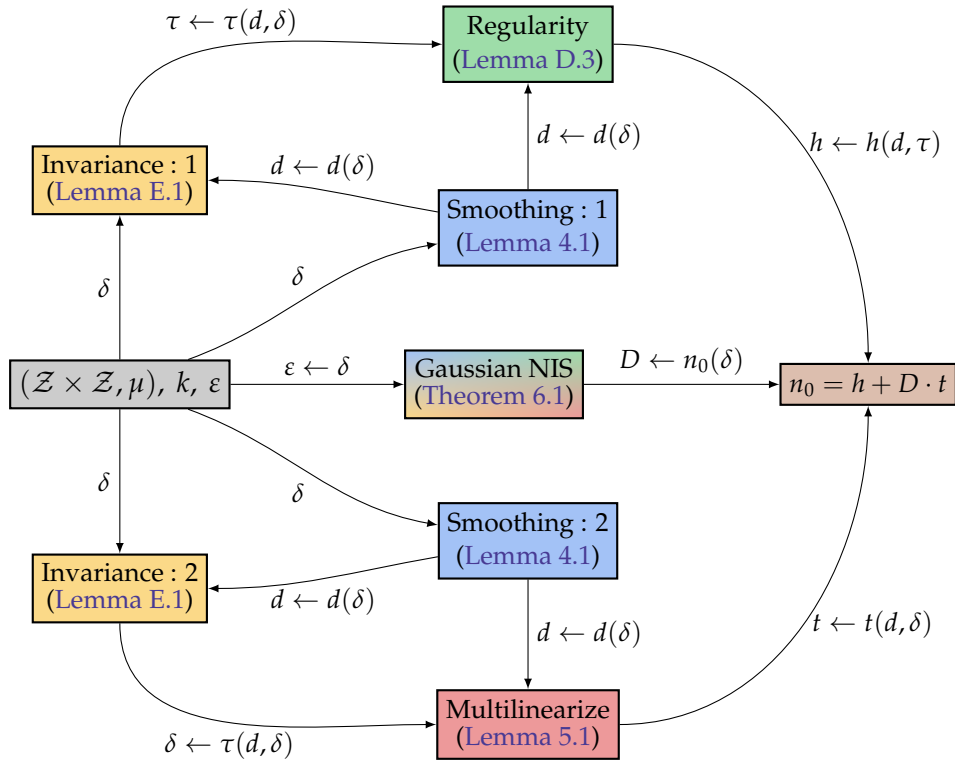
Figure 4: Dependency of parameters in the proof of Theorem 7.1

after the transformation of Theorem 6.1, the resulting function might be none of low-degree, multilinear or low-influential. To get around this we apply Smoothing (Lemma 4.1) to make it low-degree and Multilinearization (Lemma 5.1) to make it multilinear and low-influential, after which we can apply the invariance principle (Lemma E.1). An overview of the transformations done is presented in Figure 5.

Proof of Theorem 7.1. As in the proof of Theorem 6.1, for any $i, j \in [k]$, we focus on the quantity $\left\langle A_i, B_j \right\rangle_{\mu^{\otimes n}}$ which is the probability of the event that [*Alice outputs i and Bob outputs j*]. Through the several steps we modify Alice's and Bob's strategy, while preserving this quantity approximately for every i, j. If we preserve the probability that Alice outputs i and Bob outputs j for every i, j upto an additive ε/k^2, it implies that we preserve the joint distribution of Alice and Bob's outputs up to an ℓ_1-distance of ε.

Choice of parameters & bound on n_0. As described in the overview, we are going to invoke several of the lemmas we have developed in our proof. We now describe the choice of parameters for which we invoke these lemmas, thereby obtaining our final explicit bound on n_0. We recommend consulting Figure 4 to follow the exact chain of dependencies among the parameters (the dependencies on μ and k are suppressed in the figure for clarity).

Let δ be a running parameter, that we finalize at the end (in terms of μ, k and ε). We wish to invoke the Smoothing operation over $\mathcal{Z}^N$ (Lemma 4.1) with parameter δ. This dictates a value of $d = d(\rho, k, \delta)$, which is the degree of the polynomials obtained after smoothing. We will invoke the Invariance Principle (Lemma E.1) with parameters δ and d as obtained just now. This dictates a value of $\tau = \tau(\mu, k, d, \delta)$, which is the bound on the influence needed in order to apply the invariance principle. We will invoke the Regularity Lemma (Lemma D.3) with parameters d and τ as obtained above. This dictates a value of $h = h(\mu, k, d, \tau)$, which is the bound on number of *head* coordinates that need to be fixed to get all influences less than τ on the remaining coordinates, for a random restriction of the head coordinates.

21

We will apply Theorem 6.1 with the error parameter ε as δ. This dictates a value of $D = n_0(k, \rho, \delta)$, which the number of coordinates of correlated Gaussians needed after dimension reduction. In order to go back from correlated Gaussians to samples from $(\mathcal{Z} \times \mathcal{Z}; \mu)$, we will again apply the Smoothing operation, this time over Gaussian space (Lemma 4.1) with parameter δ, which again dictates a value of $d = d(\rho, k, \delta)$. We will again apply the Invariance principle (Lemma E.1) with parameters δ and d as obtained just now. This dictates a value of $\tau = \tau(\mu, k, d, \delta)$, which is the bound on the influence needed in order to apply the invariance principle. In order to make the influences small, we will apply the Multilinearization operation (Lemma 5.1) with parameters $\delta \leftarrow \tau$ and d as obtained above. This dictates a value of $t = t(k, d, \delta)$, which is the blow up incurred while making the polynomials multilinear and to make them have all influences smaller than δ.

Finally $n_0 = h + D \cdot t$. Recall that h is the number of *head* coordinates in the Regularity Lemma. D is the number of coordinates obtained after applying Gaussian NIS (Theorem 6.1). Finally t is the blow up incurred while going back from correlated Gaussian space to $(\mathcal{Z} \times \mathcal{Z}; \mu)$. We will eventually choose $\delta = \varepsilon^2/k^4$. It can be inferred by going through all the parameters carefully that $n_0(\mu, k, \varepsilon)$ is an explicit function that can be upper bounded as $\exp\left(\text{poly}\left(k, \frac{1}{\varepsilon}, \frac{1}{1-\rho}, \log\left(\frac{1}{\alpha}\right)\right)\right)$. We skip this the details of this calculation for brevity.

Analysis of the transformations. We now turn to the analysis of the above transformation. We wish to show that $\langle A_i, B_j \rangle_{\mu^{\otimes N}} \approx \langle \widetilde{A}_i, \widetilde{B}_j \rangle_{\mathcal{G}_\rho^{\otimes n_0}}$ for every $i, j \in [k]$. We transform A and B through each of the following steps, as illustrated in Figure 5. At each step, we approximately preserve the correlation $\langle A_i, B_j \rangle$ for every $i, j \in [k]$.

1. **Smoothing (over hypercube)**: We apply Lemma 4.1 on A and B to get the low-degree versions $A^{(1)} : \mathcal{Z}^N \to \mathbb{R}^k$ and $B^{(1)} : \mathcal{Z}^N \to \mathbb{R}^k$. This guarantees that $A^{(1)}$ and $B^{(1)}$ have degree at most d. Moreover, we have that for every $i, j \in [k]$,

$$\left| \left\langle A_i^{(1)}, B_j^{(1)} \right\rangle_{\mu^{\otimes N}} - \langle A_i, B_j \rangle_{\mu^{\otimes N}} \right| \leq \delta$$

Additionally,

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

Similarly, we also have that,

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

Using Lemma 2.4, we can conclude that for every $i, j \in [k]$,

$$\left| \left\langle \mathcal{R}_i(A^{(1)}), \mathcal{R}_j(B^{(1)}) \right\rangle_{\mu^{\otimes N}} - \langle A_i, B_j \rangle_{\mu^{\otimes N}} \right| \leq 3\delta \tag{6}$$

2. **Regularity Lemma**: We apply Lemma D.3 to identify a subset $H \subseteq [n]$ with $|H| = h$, such that, for a random restriction $(\boldsymbol{x}_H, \boldsymbol{y}_H) \sim \mu^{\otimes h}$, it holds with probability at least $1 - \tau$, that the restricted functions $(A_i^{(1)})^{\boldsymbol{x}_H} : \mathcal{Z}^{N-h} \to \mathbb{R}^k$ and $(B_i^{(1)})^{\boldsymbol{y}_H} : \mathcal{Z}^{N-h} \to \mathbb{R}^k$ have all individual influences smaller than τ, for any $i \in [k]$. We call a restriction $(\boldsymbol{x}_H, \boldsymbol{y}_H)$ as "*good*" in this case, and "*bad*" otherwise. Note that any such restriction ("good" or "bad") has degree at most d.

For the rest of the steps, we will focus on a *good* $(\boldsymbol{x}_H, \boldsymbol{y}_H)$. For convenience, define $N_1 = N - h$.

3. **Invariance Principle (from $\mathcal{Z}$ to $\mathbb{R}$)**: For a *good* $(\boldsymbol{x}_H, \boldsymbol{y}_H)$, we apply Lemma E.1 on $(A^{(1)})^{\boldsymbol{x}_H}$ and $(B^{(1)})^{\boldsymbol{y}_H}$ to get functions $(A^{(2)})^{\boldsymbol{x}_H} : \mathbb{R}^{N_2} \to \Delta_k$ and $(B^{(2)})^{\boldsymbol{y}_H} : \mathbb{R}^{N_2} \to \Delta_k$ (where $N_2 = N_1 \cdot (q-1)$), such that for every $i, j \in [k]$,

$$\left| \left\langle (A_i^{(2)})^{\boldsymbol{x}_H}, (B_j^{(2)})^{\boldsymbol{y}_H} \right\rangle_{\mathcal{G}_\rho^{\otimes N_2}} - \left\langle \mathcal{R}_i\left((A^{(1)})^{\boldsymbol{x}_H}\right), \mathcal{R}_j\left((B^{(1)})^{\boldsymbol{y}_H}\right) \right\rangle_{\mu^{\otimes N_1}} \right| \leq \delta \tag{7}$$
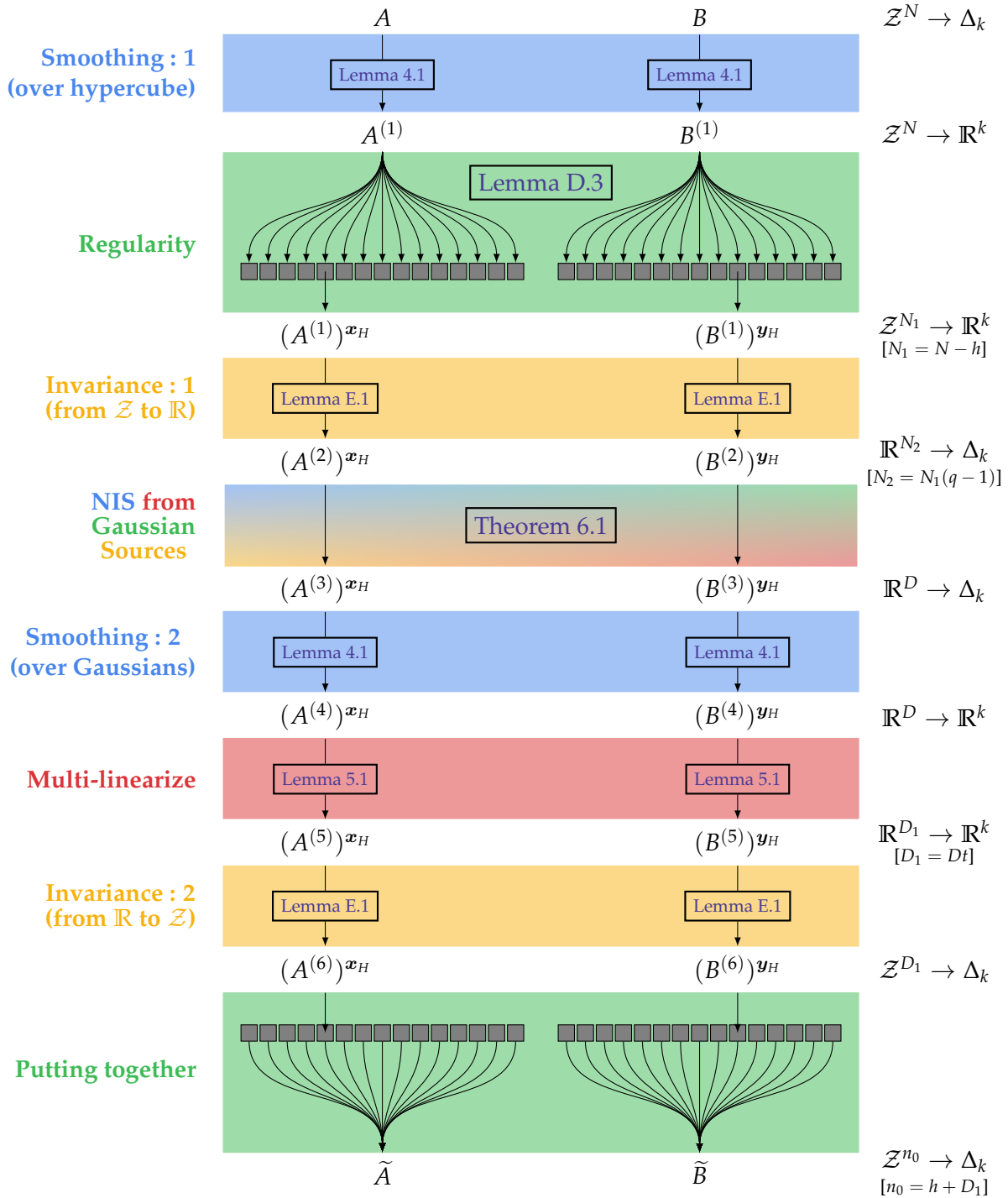
Figure 5: Transformations for Non-interactive simulation from Arbitrary Discrete Sources

Note that strictly speaking Lemma E.1, as stated, gives us functions mapping to $\mathbb{R}^k$ and not Δ_k. However, we consider their rounded versions, which exactly gives us the statement above.

4. **Dimension Reduction**: We apply Theorem 3.1 on $(A^{(2)})^{\boldsymbol{x}_H}$ and $(B^{(2)})^{\boldsymbol{y}_H}$, to get functions $(A^{(3)})^{\boldsymbol{x}_H} : \mathbb{R}^D \to \Delta_k$ and $(B^{(3)})^{\boldsymbol{y}_H} : \mathbb{R}^D \to \Delta_k$, such that, for every $i, j \in [k]$,

$$\left| \left\langle (A_i^{(3)})^{\boldsymbol{x}_H}, (B_j^{(3)})^{\boldsymbol{y}_H} \right\rangle_{\mathcal{G}_\rho^{\otimes D}} - \left\langle (A_i^{(2)})^{\boldsymbol{x}_H}, (B_j^{(2)})^{\boldsymbol{y}_H} \right\rangle_{\mathcal{G}_\rho^{\otimes N_2}} \right| \leq \delta \qquad (8)$$

Note that this is the only randomized step in the entire transformation. This reduction succeeds with probability at least $1 - 4\delta$ for every *good* $(\boldsymbol{x}_H, \boldsymbol{y}_H)$. For a fixed choice of the random seed, we call a *good* $(\boldsymbol{x}_H, \boldsymbol{y}_H)$ as "*lucky*"[9] if the reduction succeeds for that $(\boldsymbol{x}_H, \boldsymbol{y}_H)$. In expectation over the choice of random seeds, a $(1 - 4\delta)$ fraction of the *good* $(\boldsymbol{x}_H, \boldsymbol{y}_H)$ are going to be *lucky*. We can hence choose a choice of random seed for which indeed a $(1 - 4\delta)$ fraction of the *good* $(\boldsymbol{x}_H, \boldsymbol{y}_H)$ are *lucky*.

On the other hand, Regularity Lemma ensures that with probability $1 - \delta$, a sampled $(\boldsymbol{x}_H, \boldsymbol{y}_H)$ will be *good*. This gives that for the said choice of random seed in Theorem 6.1, with probability at least $1 - 5\delta$, the sampled $(\boldsymbol{x}_H, \boldsymbol{y}_H)$ is both *good* and *lucky*. For the rest of the steps, we will focus on a *good* and *lucky* $(\boldsymbol{x}_H, \boldsymbol{y}_H)$.

5. **Smoothing (over Gaussian space)**: We again apply Lemma 4.1 on $(A^{(3)})^{\boldsymbol{x}_H}$ and $(B^{(3)})^{\boldsymbol{y}_H}$ to get the low-degree versions $(A^{(4)})^{\boldsymbol{x}_H} : \mathbb{R}^D \to \mathbb{R}^k$ and $(B^{(4)})^{\boldsymbol{y}_H} : \mathbb{R}^D \to \mathbb{R}^k$. This guarantees that $(A^{(4)})^{\boldsymbol{x}_H}$ and $(B^{(4)})^{\boldsymbol{y}_H}$ have degree at most d. Moreover, we have that for every $i, j \in [k]$,

$$\left| \left\langle (A_i^{(4)})^{\boldsymbol{x}_H}, (B_j^{(4)})^{\boldsymbol{y}_H} \right\rangle_{\mathcal{G}_\rho^{\otimes D}} - \left\langle (A_i^{(3)})^{\boldsymbol{x}_H}, (B_j^{(3)})^{\boldsymbol{y}_H} \right\rangle_{\mathcal{G}_\rho^{\otimes D}} \right| \leq \delta \qquad (9)$$

Additionally,

$$\left\| \mathcal{R}(A^{(4)}) - A^{(4)} \right\|_2 \leq \delta \qquad \text{and} \qquad \left\| \mathcal{R}(B^{(4)}) - B^{(4)} \right\|_2 \leq \delta$$

6. **Multilinearization**: We apply Lemma 5.1 on $(A^{(4)})^{\boldsymbol{x}_H}$ and $(B^{(4)})^{\boldsymbol{y}_H}$ to get the multilinearized and low-influential versions $(A^{(5)})^{\boldsymbol{x}_H} : \mathbb{R}^{D_1} \to \mathbb{R}^k$ and $(B^{(5)})^{\boldsymbol{y}_H} : \mathbb{R}^{D_1} \to \mathbb{R}^k$ (where $D_1 = Dt$). Thus, we have for every $i, j \in [k]$,

$$\left| \left\langle (A_i^{(5)})^{\boldsymbol{x}_H}, (B_j^{(5)})^{\boldsymbol{y}_H} \right\rangle_{\mathcal{G}_\rho^{\otimes D_1}} - \left\langle (A_i^{(4)})^{\boldsymbol{x}_H}, (B_j^{(4)})^{\boldsymbol{y}_H} \right\rangle_{\mathcal{G}_\rho^{\otimes D}} \right| \leq \delta$$

Additionally, combining with Lemma 2.5,

$$\left\| \mathcal{R}(A^{(5)}) - A^{(5)} \right\|_2 \leq \left\| \mathcal{R}(A^{(4)}) - A^{(4)} \right\|_2 + \left\| A^{(5)} - A^{(4)} \right\|_2 \leq 2\delta$$

Similarly,

$$\left\| \mathcal{R}(B^{(5)}) - B^{(5)} \right\|_2 \leq 2\delta$$

Combining all this with Lemma 2.4 we get that,

$$\left| \left\langle \mathcal{R}_i\left((A^{(5)})^{\boldsymbol{x}_H} \right), \mathcal{R}_j\left((B^{(5)})^{\boldsymbol{y}_H} \right) \right\rangle_{\mathcal{G}_\rho^{\otimes D_1}} - \left\langle (A_i^{(4)})^{\boldsymbol{x}_H}, (B_j^{(4)})^{\boldsymbol{y}_H} \right\rangle_{\mathcal{G}_\rho^{\otimes D}} \right| \leq 5\delta \qquad (10)$$

Additionally, note that we also have that $\mathrm{Inf}_\ell((A_i^{(5)}))^{\boldsymbol{x}_H} \leq \tau$ and $\mathrm{Inf}_\ell((B_i^{(5)}))^{\boldsymbol{y}_H} \leq \tau$ for all $i \in [k]$ and $\ell \in [n]$. This is helpful for us to apply the invariance principle next.

---

[9] rather poor choice of terminology, given that most $(\boldsymbol{x}_H, \boldsymbol{y}_H)$ end up being *lucky*!

7. **Invariance Principle (from $\mathbb{R}$ to $\mathcal{Z}$):** We apply Lemma E.1 on $(A^{(5)})^{x_H}$ and $(B^{(5)})^{y_H}$ to get functions $(A^{(6)})^{x_H} : \mathcal{Z}^{D_1} \to \Delta_k$ and $(B^{(6)})^{y_H} : \mathcal{Z}^{D_1} \to \Delta_k$, such that for every $i, j \in [k]$,

$$\left| \left\langle (A_i^{(6)})^{x_H}, (B_j^{(6)})^{y_H} \right\rangle_{\mu^{\otimes D_1}} - \left\langle \mathcal{R}_i \left((A^{(5)})^{x_H} \right), \mathcal{R}_j \left((B^{(5)})^{y_H} \right) \right\rangle_{\mathcal{G}_\rho^{\otimes D_1}} \right| \leq \delta \tag{11}$$

Note again that strictly speaking Lemma E.1 gives us functions mapping to $\mathbb{R}^k$ and not Δ_k. However, we consider their rounded versions, which exactly gives us the statement above.

Putting it together. We now show how to put together Equations (6) to (11) to get our final conclusion. We now define our final functions $\widetilde{A} : \mathcal{Z}^{n_0} \to \Delta_k$ and $\widetilde{B} : \mathcal{Z}^{n_0} \to \Delta_k$ as follows. Firstly, we interpret the $n_0 = h + D_1$ coordinates of x as two parts: *head* coordinates $x_H \in \mathcal{Z}^h$ and the remaining coordinates $x_R \in \mathcal{Z}^{D_1}$. Similarly for y.

$$\widetilde{A}(x) = \widetilde{A}(x_H, x_R) = (A^{(6)})^{x_H}(x_R) \qquad \text{and} \qquad \widetilde{B}(y) = \widetilde{B}(y_H, y_R) = (B^{(6)})^{y_H}(y_R) \,.$$

We now show that for all $i, j \in [k]$, it holds that,

$$\left| \left\langle \widetilde{A}_i, \widetilde{B}_j \right\rangle_{\mu^{\otimes n_0}} - \left\langle A_i, B_j \right\rangle_{\mu^{\otimes N}} \right| \leq O(\delta)$$

We note that,

$$\left\langle \widetilde{A}_i, \widetilde{B}_j \right\rangle_{\mu^{\otimes n_0}}$$

$$= \mathop{\mathbb{E}}_{(x_H, y_H) \sim \mu^{\otimes h}} \left\langle (A_i^{(6)})^{x_H}, (B_j^{(6)})^{y_H} \right\rangle_{\mu^{\otimes D_1}}$$

$$= \Pr \left[\begin{array}{c} (x_H, y_H) \text{ is} \\ (good \text{ and } lucky) \end{array} \right] \cdot \mathop{\mathbb{E}}_{\substack{(x_H, y_H) \sim \mu^{\otimes h} \\ \mid good \text{ \& } lucky}} \left\langle (A_i^{(6)})^{x_H}, (B_j^{(6)})^{y_H} \right\rangle_{\mu^{\otimes D_1}}$$

$$+ \Pr \left[\begin{array}{c} (x_H, y_H) \text{ is not} \\ (good \text{ and } lucky) \end{array} \right] \cdot \mathop{\mathbb{E}}_{\substack{(x_H, y_H) \sim \mu^{\otimes h} \\ \mid \text{not } (good \text{ \& } lucky)}} \left\langle (A_i^{(6)})^{x_H}, (B_j^{(6)})^{y_H} \right\rangle_{\mu^{\otimes D_1}}$$

$$= \Pr \left[\begin{array}{c} (x_H, y_H) \text{ is} \\ (good \text{ and } lucky) \end{array} \right] \cdot \mathop{\mathbb{E}}_{\substack{(x_H, y_H) \sim \mu^{\otimes h} \\ \mid good \text{ \& } lucky}} \left\langle (A_i^{(6)})^{x_H}, (B_j^{(6)})^{y_H} \right\rangle_{\mu^{\otimes D_1}} \pm O(\delta)$$

$$\dots \left(\text{since } \Pr \left[\begin{array}{c} (x_H, y_H) \text{ is not} \\ (good \text{ and } lucky) \end{array} \right] \leq O(\delta) \right)$$

$$= \Pr \left[\begin{array}{c} (x_H, y_H) \text{ is} \\ (good \text{ and } lucky) \end{array} \right] \cdot \mathop{\mathbb{E}}_{\substack{(x_H, y_H) \sim \mu^{\otimes h} \\ \mid good \text{ \& } lucky}} \left\langle \mathcal{R}_i \left((A^{(1)})^{x_H} \right), \mathcal{R}_j \left((B^{(1)})^{y_H} \right) \right\rangle_{\mu^{\otimes N_1}} \pm O(\delta)$$

$$\dots \left(\begin{array}{c} \text{combining Equations (7) to (11)} \\ \text{which hold for } good \text{ and } lucky \ (x_H, y_H) \end{array} \right)$$

$$= \mathop{\mathbb{E}}_{(x_H, y_H) \sim \mu^{\otimes h}} \left\langle \mathcal{R}_i \left((A^{(1)})^{x_H} \right), \mathcal{R}_j \left((B^{(1)})^{y_H} \right) \right\rangle_{\mu^{\otimes N_1}} \pm O(\delta)$$

$$\dots \left(\text{since } \Pr \left[\begin{array}{c} (x_H, y_H) \text{ is} \\ (good \text{ and } lucky) \end{array} \right] \in [1 - O(\delta), 1] \right)$$

$$= \left\langle \mathcal{R}_i(A^{(1)}), \mathcal{R}_j(B^{(1)}) \right\rangle_{\mu^{\otimes N}} \pm O(\delta)$$

$$= \left\langle A_i, B_j \right\rangle_{\mu^{\otimes N}} \pm O(\delta) \qquad \dots \text{(using Equation (6))}$$

Thus, more strongly, if we instantiate δ as $O(\varepsilon/k^2)$, then we get that

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

$\square$

8 Acknowledgments

The authors are extremely grateful to Madhu Sudan for very many helpful conversations 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.

References

[AC93] Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography. part i: secret sharing. *IEEE Transactions on Information Theory*, 39(4), 1993. 8

[AC98] 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. 8

[AL06] 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. 8

[Bei12] Salman Beigi. A new quantum data processing inequality. *CoRR*, abs/1210.1689, 2012. 8

[BG15] Salman Beigi and Amin Gohari. On the duality of additivity and tensorization. *arXiv preprint arXiv:1502.00827*, 2015. 8

[BGI14] 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. 8

[BHH+08] 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, 2008. 8

[BM11] Andrej Bogdanov and Elchanan Mossel. On extracting common random bits from correlated sources. *Information Theory, IEEE Transactions on*, 57(10):6351–6355, 2011. 8

[Bor85] Christer Borell. Geometric bounds on the ornstein-uhlenbeck velocity process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 70(1):1–13, 1985. 1, 4, 5

[BR11] 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. 8

[BS94] Gilles Brassard and Louis Salvail. Secret-key reconciliation by public discussion. In *advances in CryptologyEUROCRYPT93*, pages 410–423. Springer, 1994. 8

[BS15] Mark Braverman and Jon Schneider. Information complexity is computable. *arXiv preprint arXiv:1502.02971*, 2015. 8

[CDS08]   Eric Chitambar, Runyao Duan, and Yaoyun Shi. Tripartite entanglement transformations and tensor rank. *Physical review letters*, 101(14):140502, 2008. 8

[CGMS14]   Clement Cannonne, Venkat Guruswami, Raghu Meka, and Madhu Sudan. Communication with imperfectly shared randomness. *ITCS*, 2014. 8

[CN00]   Imre Csiszár and Prakash Narayan. Common randomness and secret key generation with a helper. *Information Theory, IEEE Transactions on*, 46(2):344–366, 2000. 8

[DB13]   Payam Delgosha and Salman Beigi. Impossibility of local state transformation via hyper-contractivity. *CoRR*, abs/1307.2747, 2013. 8

[DB14]   Payam Delgosha and Salman Beigi. Impossibility of local state transformation via hyper-contractivity. *Communications in Mathematical Physics*, 332(1):449–476, 2014. 8

[DMN17a]   Anindya De, Elchanan Mossel, and Joe Neeman. Noise stability is computable and low dimensional. *Computational Complexity Conference, CCC*, 2017. 2, 5, 8, 16

[DMN17b]   Anindya De, Elchanan Mossel, and Joe Neeman. Non interactive simulation of correlated distributions is decidable. *CoRR*, abs/1701.01485, 2017. 5, 8, 16

[DS14]   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. 2

[DSTW10]   Ilias Diakonikolas, Rocco A Servedio, Li-Yang Tan, and Andrew Wan. A regularity lemma, and low-weight approximators, for low-degree polynomial threshold functions. In *Computational Complexity (CCC), 2010 IEEE 25th Annual Conference on*, pages 211–222. IEEE, 2010. 9, 48

[Geb41]   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. 3, 11

[GK73]   Peter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973. 3

[GKS16a]   Badih Ghazi, Pritish Kamath, and Madhu Sudan. Communication complexity of permutation-invariant functions. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1902–1921, 2016. 8

[GKS16b]   Badih Ghazi, Pritish Kamath, and Madhu Sudan. Decidability of non-interactive simulation of joint distributions. In *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, 2016. 4, 5, 8, 9, 16, 20, 48, 49

[GW95]   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. 3

[Hir35]   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. 3, 11

[HMN16]   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. 1

[Hol09]   Thomas Holenstein. Parallel repetition: Simplification and the no-signaling case. *Theory of Computing*, 5(1):141–172, 2009. 8

[IM12]   Marcus Isaksson and Elchanan Mossel. Maximally stable gaussian partitions with discrete applications. *Israel Journal of Mathematics*, 189(1):347–396, 2012. 1, 9, 50

[KA16]   Sudeep Kamath and Venkat Anantharam. On non-interactive simulation of joint distributions. *IEEE Trans. Information Theory*, 62(6):3419–3435, 2016. 3, 4, 8

[KKM+11]   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. 8

[KKMO07]   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. 1

[Lov79]   László Lovász. On the shannon capacity of a graph. *Information Theory, IEEE Transactions on*, 25(1):1–7, 1979. 8

[Mau93]   Ueli M Maurer. Secret key agreement by public discussion from common information. *Information Theory, IEEE Transactions on*, 39(3):733–742, 1993. 8

[MO04]   Elchanan Mossel and Ryan O'Donnell. Coin flipping from a cosmic source: On error correction of truly random bits. *arXiv preprint math/0406504*, 2004. 8

[MOO05]   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. 1, 50

[MOR+06]   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. 8

[Mos10]   Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010. 1, 9, 41, 42, 50

[Nie99]   Michael A Nielsen. Conditions for a class of entanglement transformations. *Physical Review Letters*, 83(2):436, 1999. 8

[ope]   OpenQIProblemsWiki - All the Bell Inequalities. http://qig.itp.uni-hannover.de/qiproblems/1. Accessed: 2016-07-12. 8

[Rao11]   Anup Rao. Parallel repetition in projection games and a concentration bound. *SIAM J. Comput.*, 40(6):1871–1891, 2011. 8

[Raz98]   Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998. 8

[Raz11]   Ran Raz. A counterexample to strong parallel repetition. *SIAM J. Comput.*, 40(3):771–777, 2011. 8

[Rén59]   Alfréd Rényi. On measures of dependence. *Acta mathematica hungarica*, 10(3-4):441–451, 1959. 3, 11

[RW05]   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. 8

[Sha56]   Claude E Shannon. The zero error capacity of a noisy channel. *Information Theory, IRE Transactions on*, 2(3):8–19, 1956. 8

[Wit75]   Hans S Witsenhausen. On sequences of pairs of dependent random variables. *SIAM Journal on Applied Mathematics*, 28(1):100–113, 1975. 3, 4, 5

[Wol07]   Pawel Wolff. Hypercontractivity of simple random variables. *Studia Mathematica*, 180(3):219–236, 2007. 40

[Wyn75]   Aaron D. Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, 1975. 3

A   Proofs of Mean and Variance Bounds in Dimension Reduction

In this section, we provide the proof of Lemma 3.3. This is the main new technical component introduced in this paper. Even though the calculations might seem cumbersome, they involve mostly elementary steps. The proof breaks down into three modular steps, which we describe first.

Recall that we are given degree d multilinear polynomials $A : \mathbb{R}^N \to \mathbb{R}$ and $B : \mathbb{R}^N \to \mathbb{R}$. For a matrix M sampled from $\mathcal{N}(0,1)^{\otimes(N \times D)}$, we defined functions $A_M : \mathbb{R}^D \to \mathbb{R}$ and $B_M : \mathbb{R}^D \to \mathbb{R}$ as

$$A_M(\boldsymbol{a}) \;=\; A\left(\frac{M\boldsymbol{a}}{\|\boldsymbol{a}\|_2}\right) \qquad \text{and} \qquad B_M(\boldsymbol{b}) = B\left(\frac{M\boldsymbol{b}}{\|\boldsymbol{b}\|_2}\right)$$

and we defined their correlation as $F(M) \stackrel{\text{def}}{=} \langle A_M, B_M \rangle_{\mathcal{G}_\rho^{\otimes D}}$. Lemma 3.3 proves bounds on the mean and variance of $F(M)$, which we restate below for convenience.

Lemma 3.3. (Bound on Mean & Variance). *Given parameters d and δ, there exists $D := D(d, \delta)$ such that the following holds: For $M \sim \mathcal{N}(0,1)^{\otimes(N \times D)}$,*

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

$$\mathop{\text{Var}}_M (F(M)) \;\leq\; \delta \qquad \textbf{(Variance bound)}$$

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

The proof of the above Lemma proceeds in three main steps.

1. In Appendix A.1, we first prove a *meta-lemma* (Lemma A.1) that will help us prove both the mean and variance bounds; indeed this meta-lemma is at the heart of why Theorem 3.1 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.

2. In Appendix A.2, we prove strong enough bounds on the mean and co-variances of degree-d multilinear monomials, under the above transformation of replacing $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^N$ (inputs to A and B) by $\frac{M\boldsymbol{a}}{\|\boldsymbol{a}\|_2}$ and $\frac{M\boldsymbol{b}}{\|\boldsymbol{b}\|_2}$ respectively.

3. In Appendix A.3, we finally use the above bounds on mean and co-variances of degree-d multilinear monomials in order to prove Lemma 3.3.

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

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 A.1. *Given parameter d and δ, there exists an explicitly computable $D := D(d, \delta)$ such that the following holds:*

Let $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_d)$ be a multivariate Gaussian distribution such that,

- *each $\boldsymbol{u}_i, \boldsymbol{v}_i \in \mathbb{R}^D$ are marginally distributed as D-dimensional standard Gaussians, i.e. γ_D.*

- *for each $j \in [D]$, the joint distribution of the j-th coordinates, i.e., $(u_{1,j}, \ldots, u_{d,j}, v_{1,j}, \ldots, v_{d,j})$, is independent across different values of j.*

Then,

$$\left| \underset{\{\boldsymbol{u}_i, \boldsymbol{v}_i\}_i}{\mathbb{E}} \left[\prod_{i=1}^{d} \frac{\langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle}{\|\boldsymbol{u}_i\|_2 \|\boldsymbol{v}_i\|_2} \right] - \prod_{i=1}^{d} \underset{\{\boldsymbol{u}_i, \boldsymbol{v}_i\}_i}{\mathbb{E}} \left[\frac{\langle \boldsymbol{u}_i, \boldsymbol{v}_i \rangle}{D} \right] \right| \leq \delta.$$

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

We point out that there are two steps taking place in Lemma A.1:
(i) the replacement of $\|\boldsymbol{u}_i\|_2$ (and $\|\boldsymbol{v}_i\|_2$) by $\sqrt{D}$ (around which it is tightly concentrated), and
(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 show the following lemma which gives us useful bounds on the mean and variance of products of negative powers of the ℓ_2-norm of a standard Gaussian vector.

Lemma A.2. *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 a D-dimensional standard Gaussian (i.e., γ_D), and let $d_1, d_2, \ldots, d_\ell$ be non-negative integers with $d := \sum_{i=1}^{\ell} d_i$. Then,*

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

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

Remark A.3. *It is conceivable that the bounds in Lemma A.2 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 A.4. *There exists an absolute constant C such that for sufficiently large $d, D \in \mathbb{Z}_{>0}$, such that $D > Cd^2$, we have that for $\boldsymbol{w} \sim \gamma_D$,*

$$\left| \underset{\boldsymbol{w}}{\mathbb{E}} \left[\frac{1}{\|\boldsymbol{w}\|_2^d} \right] - \frac{1}{D^{d/2}} \right| \leq C \cdot \left(\frac{d^2}{D^{\frac{d}{2}+1}} \right), \tag{12}$$

$$\underset{\boldsymbol{w}}{\mathrm{Var}} \left[\frac{1}{\|\boldsymbol{w}\|_2^d} \right] \leq 8C \cdot \left(\frac{d^2}{D^{d+1}} \right). \tag{13}$$

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

$$f_D(x) = \frac{x^{D-1} \cdot e^{-\frac{x^2}{2}}}{2^{\frac{D}{2}-1} \cdot \Gamma(\frac{D}{2})},$$

for every $x \geq 0$ (and where $\Gamma(\cdot)$ denotes the Gamma function). Thus, we have that

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{w}} \left[\frac{1}{\|\boldsymbol{w}\|^d} \right] &= \int_0^\infty \frac{1}{x^d} \cdot f_D(x) dx \\
&= \int_0^\infty \frac{x^{D-d-1} \cdot e^{-\frac{x^2}{2}}}{2^{\frac{D}{2}-1} \cdot \Gamma(\frac{D}{2})} dx \\
&= \frac{2^{\frac{D-d-1}{2}} \cdot \Gamma\left(\frac{D-d}{2}\right)}{2^{\frac{D}{2}-1} \cdot \Gamma(\frac{D}{2})} \\
&= \frac{1}{D^{d/2}} \cdot \left(1 \pm O\left(\frac{d^2}{D} \right) \right),
\end{aligned}
$$

where the last equality follows from the following 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 (12), for the explicit constant C that can be derived from the Stirling's approximation. Now, Equation (13) immediately follows as:

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

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. $D > Cd^2$. $\qquad\square$

We now show how to generalize the above to prove Lemma A.2.

Proof of Lemma A.2. More specifically, we will show that,

$$\left| \mathbb{E} \left[\prod_{i=1}^{\ell} \frac{1}{\|\boldsymbol{w}_i\|_2^{d_i}} \right] - \frac{1}{D^{d/2}} \right| \leq C \cdot \ell^3 \cdot \left(\frac{d^2}{D^{\frac{d}{2}+1}} \right) \tag{14}$$

$$\operatorname{Var} \left[\prod_{i=1}^{\ell} \frac{1}{\|\boldsymbol{w}_i\|_2^{d_i}} \right] \leq 8C \cdot \ell^3 \cdot \left(\frac{d^2}{D^{d+1}} \right) \tag{15}$$

where C is the absolute constant (as obtained in Proposition A.4). 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 A.4. For the inductive step, we assume that the bound in Equations (14) and (15) 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 (14). The main idea that we use to prove this inductively is: for any two random variables X and Y, it holds that $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathrm{Cov}[X, Y]$, and that the covariance satisfies $|\mathrm{Cov}[X, Y]| \leq \sqrt{\mathrm{Var}[X] \cdot \mathrm{Var}[Y]}$ (by Cauchy-Schwarz inequality). Thus, we get,

$$
\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^{2\ell} \frac{1}{\|w_i\|_2^{d_i}}\right] &= \mathbb{E}\left[\prod_{i=1}^{\ell} \frac{1}{\|w_i\|_2^{d_i}} \cdot \prod_{i=\ell+1}^{2\ell} \frac{1}{\|w_i\|_2^{d_i}}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{\ell} \frac{1}{\|w_i\|_2^{d_i}}\right] \cdot \mathbb{E}\left[\prod_{i=\ell+1}^{2\ell} \frac{1}{\|w_i\|_2^{d_i}}\right] + \mathrm{Cov}\left[\prod_{i=1}^{\ell} \frac{1}{\|w_i\|_2^{d_i}}, \prod_{i=\ell+1}^{2\ell} \frac{1}{\|w_i\|_2^{d_i}}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{\ell} \frac{1}{\|w_i\|_2^{d_i}}\right] \cdot \mathbb{E}\left[\prod_{i=\ell+1}^{2\ell} \frac{1}{\|w_i\|_2^{d_i}}\right] \pm \sqrt{\mathrm{Var}\left[\prod_{i=1}^{\ell} \frac{1}{\|w_i\|_2^{d_i}}\right] \cdot \mathrm{Var}\left[\prod_{i=\ell+1}^{2\ell} \frac{1}{\|w_i\|_2^{d_i}}\right]}. \quad (16)
\end{aligned}
$$

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

$$
\mathbb{E}\left[\prod_{i=1}^{\ell} \frac{1}{\|w_i\|_2^{d_i}}\right] = \frac{1}{D^{s_1/2}}\left(1 \pm C \cdot \ell^3 \cdot \left(\frac{s_1^2}{D}\right)\right) \quad (17)
$$

$$
\mathbb{E}\left[\prod_{i=\ell+1}^{2\ell} \frac{1}{\|w_i\|_2^{d_i}}\right] = \frac{1}{D^{s_2/2}}\left(1 \pm C \cdot \ell^3 \cdot \left(\frac{s_2^2}{D}\right)\right) \quad (18)
$$

and

$$
\mathrm{Var}\left[\prod_{i=1}^{\ell} \frac{1}{\|w_i\|_2^{d_i}}\right] \leq \frac{1}{D^{s_1}} \cdot 8C \cdot \ell^3 \cdot \left(\frac{s_1^2}{D}\right) \quad (19)
$$

$$
\mathrm{Var}\left[\prod_{i=\ell+1}^{2\ell} \frac{1}{\|w_i\|_2^{d_i}}\right] \leq \frac{1}{D^{s_2}} \cdot 8C \cdot \ell^3 \cdot \left(\frac{s_2^2}{D}\right) \quad (20)
$$

Plugging Equations (17) to (20) in Equation (16), it is not hard to see that,

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

This completes the proof of Equation (14). Now, Equation (15) follows easily as,

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

$\square$

Interchanging Product and Expectation

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

Lemma A.5. *Let $(u_1, \ldots, u_d, v_1, \ldots, v_d)$ be a multivariate Gaussian distribution such that,*

- *each of $u_i, v_i \in \mathbb{R}^D$ is marginally distributed as a D-dimensional standard Gaussian, i.e., γ_D.*

- *for each $j \in [D]$, the joint distribution of the j-th coordinates, i.e., $(u_{1,j}, \ldots, u_{d,j}, v_{1,j}, \ldots, v_{d,j})$, is independent across different values of j.*

Then,

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

Remark A.6. *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 A.5 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 A.7. *Let $w \in \mathbb{R}^\ell$ 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,*

$$\left| \mathbb{E} \left[\prod_{i=1}^\ell w_i^{d_i} \right] \right| \leq (2d)^{3d}. \tag{21}$$

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

$$\left| \mathbb{E} \left[\prod_{i=1}^\ell w_i^{d_i} \right] \right| \leq 2^{\ell-1} (\ell d)^d. \tag{22}$$

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 $w \sim \gamma_1$,

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

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 (22) 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 that we use to prove the inductive step is: for any two random variables X and Y, it holds that $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y] + \mathrm{Cov}[X, Y]$, and that the covariance satisfies $|\mathrm{Cov}[X, Y]| \leq \sqrt{\mathrm{Var}[X] \cdot \mathrm{Var}[Y]}$ (by Cauchy-Schwarz inequality). Additionally, we use that $\mathrm{Var}[X] \leq \mathbb{E}[X^2]$. Thus, we get,

$$\left| \mathbb{E} \left[\prod_{i=1}^{2\ell} w_i^{d_i} \right] \right| = \left| \mathbb{E} \left[\prod_{i=1}^\ell w_i^{d_i} \cdot \prod_{i=\ell+1}^{2\ell} w_i^{d_i} \right] \right|$$

$$= \left| \mathbb{E}\left[\prod_{i=1}^{\ell} w_i^{d_i}\right] \cdot \mathbb{E}\left[\prod_{i=\ell+1}^{2\ell} w_i^{d_i}\right] + \mathsf{Cov}\left[\prod_{i=1}^{\ell} w_i^{d_i}, \prod_{i=\ell+1}^{2\ell} w_i^{d_i}\right]\right|$$

$$\leq \left| \mathbb{E}\left[\prod_{i=1}^{\ell} w_i^{d_i}\right] \cdot \mathbb{E}\left[\prod_{i=\ell+1}^{2\ell} w_i^{d_i}\right]\right| + \sqrt{\mathsf{Var}\left[\prod_{i=1}^{\ell} w_i^{d_i}\right] \cdot \mathsf{Var}\left[\prod_{i=\ell+1}^{2\ell} w_i^{d_i}\right]}$$

$$\leq \left| \mathbb{E}\left[\prod_{i=1}^{\ell} w_i^{d_i}\right] \cdot \mathbb{E}\left[\prod_{i=\ell+1}^{2\ell} w_i^{d_i}\right]\right| + \sqrt{\mathbb{E}\left[\prod_{i=1}^{\ell} w_i^{2d_i}\right] \cdot \mathbb{E}\left[\prod_{i=\ell+1}^{2\ell} w_i^{2d_i}\right]}$$

$$\leq 2^{\ell-1}(\ell s_1)^{s_1} \cdot 2^{\ell-1}(\ell s_2)^{s_2} + \sqrt{2^{\ell-1}(2\ell s_1)^{2s_1} \cdot 2^{\ell-1}(2\ell s_2)^{2s_2}}$$

$$\leq 2^{2\ell-1}(2\ell d)^d,$$

where, the second last inequality uses the inductive assumption regarding product of ℓ terms. The last inequality follows from $s_1^{s_1} \cdot s_2^{s_2} \leq d^{s_1} \cdot d^{s_2} = d^d$. $\qquad\square$

Using the above proposition, we are now able to prove Lemma A.5.

Proof of Lemma A.5. 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

$$\mathbb{E}\left[\prod_{i=1}^{d}\langle u_i, v_i\rangle\right] = \mathbb{E}\left[\prod_{i=1}^{d}\sum_{k=1}^{D} u_{i,k}v_{i,k}\right]$$

$$= \mathbb{E}\left[\sum_{c\in[D]^d}\prod_{i=1}^{d} u_{i,c_i}v_{i,c_i}\right]$$

$$= \sum_{c\in[D]^d}\mathbb{E}\left[\prod_{i=1}^{d} u_{i,c_i}v_{i,c_i}\right]$$

$$= \sum_{c\in S}\mathbb{E}\left[\prod_{i=1}^{d} u_{i,c_i}v_{i,c_i}\right] + \sum_{c\in\overline{S}}\mathbb{E}\left[\prod_{i=1}^{d} u_{i,c_i}v_{i,c_i}\right]$$

$$= \sum_{c\in S}\prod_{i=1}^{d}\mathbb{E}[u_{i,c_i}v_{i,c_i}] + \sum_{c\in\overline{S}}\mathbb{E}\left[\prod_{i=1}^{d} u_{i,c_i}v_{i,c_i}\right], \qquad (23)$$

where the last equality follows from the assumption that the distribution of $(u_{1,j}, \ldots, u_{d,j}, v_{1,j}, \ldots, v_{d,j})$ is independent across $j \in [D]$. On the other hand, we have that

$$\prod_{i=1}^{d}\mathbb{E}[\langle u_i, v_i\rangle] = \prod_{i=1}^{d}\mathbb{E}\left[\sum_{k=1}^{D} u_{i,k}v_{i,k}\right]$$

$$= \sum_{c\in[D]^d}\prod_{i=1}^{d}\mathbb{E}[u_{i,c_i}v_{i,c_i}]$$

$$= \sum_{c\in S}\prod_{i=1}^{d}\mathbb{E}[u_{i,c_i}v_{i,c_i}] + \sum_{c\in\overline{S}}\prod_{i=1}^{d}\mathbb{E}[u_{i,c_i}v_{i,c_i}] \qquad (24)$$

Combining Equations (23) and (24), we get

$$\left|\mathbb{E}\left[\prod_{i=1}^{d}\langle u_i, v_i\rangle\right] - \prod_{i=1}^{d}\mathbb{E}[\langle u_i, v_i\rangle]\right| = \left|\sum_{c\in\overline{S}}\left(\mathbb{E}\left[\prod_{i=1}^{d} u_{i,c_i}v_{i,c_i}\right] - \prod_{i=1}^{d}\mathbb{E}[u_{i,c_i}v_{i,c_i}]\right)\right|$$

$$\leq |\overline{S}| \cdot \max_{c \in \overline{S}} \left| \mathbb{E}\left[\prod_{i=1}^{d} u_{i,c_i} v_{i,c_i} \right] - \prod_{i=1}^{d} \mathbb{E}[u_{i,c_i} v_{i,c_i}] \right|$$

$$\leq d^2 \cdot D^{d-1} \cdot \left((2d)^{3d} + 1 \right)$$

$$\leq d^{O(d)} \cdot D^{d-1},$$

where the second last inequality follows from the fact that $|\overline{S}| \leq d^2 \cdot D^{d-1}$ and from Proposition A.7. □

Putting things together to prove Lemma A.1

Proof of Lemma A.1. We prove this lemma in two steps. We show the following two bounds,

$$\left| \mathop{\mathbb{E}}_{\{u_i, v_i\}_i} \left[\prod_{i=1}^{d} \frac{\langle u_i, v_i \rangle}{\|u_i\|_2 \|v_i\|_2} \right] - \mathop{\mathbb{E}}_{\{u_i, v_i\}_i} \left[\prod_{i=1}^{d} \frac{\langle u_i, v_i \rangle}{D} \right] \right| \leq O\left(\frac{d^{2.5}}{\sqrt{D}} + \frac{d^{O(d)}}{D} \right). \tag{25}$$

$$\left| \mathop{\mathbb{E}}_{\{u_i, v_i\}_i} \left[\prod_{i=1}^{d} \frac{\langle u_i, v_i \rangle}{D} \right] - \prod_{i=1}^{d} \mathop{\mathbb{E}}_{\{a_i, b_i\}_i} \left[\frac{\langle u_i, v_i \rangle}{D} \right] \right| \leq \frac{d^{O(d)}}{D}. \tag{26}$$

Lemma A.1 now follows immediately by the triangle inequality, for an explicit choice of D that is upper bounded by $\frac{d^{O(d)}}{\delta^2}$.

Note that Equation (26) is simply a restatement of Lemma A.5. To prove Equation (25), we define the random variables

$$W := \prod_{i=1}^{d} \langle u_i, v_i \rangle \qquad \text{and} \qquad Z := \prod_{i=1}^{d} \frac{1}{\|u_i\|_2 \|v_i\|_2} - \frac{1}{D^d}.$$

Note that Equation (25) is equivalent to showing bounds on $|\mathbb{E}[W \cdot Z]|$. In order to do so, we show the following four bounds:

1. $|\mathbb{E}[W]| = D^d + d^{O(d)} \cdot D^{d-1}$.

2. $\mathsf{Var}[W] = D^{2d} + d^{O(d)} \cdot D^{2d-1}$.

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

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

We now prove each of these four bounds. We start by proving Item 1. By Lemma A.5, we have that

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

$$= \left| \prod_{i=1}^{d} \sum_{j=1}^{D} \mathbb{E}[u_{i,j} v_{i,j}] \right| + d^{O(d)} \cdot D^{d-1}$$

$$\leq D^d + d^{O(d)} \cdot D^{d-1}$$

To prove Item 2, we can again apply Lemma A.5 as follows:

$$\mathsf{Var}[W] \leq \mathbb{E}[W^2]$$

$$= \mathbb{E}\left[\prod_{i=1}^{d} \langle u_i, v_i \rangle^2 \right]$$

$$\leq D^{2d} + d^{O(d)} \cdot D^{2d-1}$$

Finally, we note that Items 3 and 4 follows exactly from Lemma A.2.

To put this together, recall the definition of covariance $\mathrm{Cov}[W, Z] := \mathbb{E}[W \cdot Z] - \mathbb{E}[W] \cdot \mathbb{E}[Z]$, and that by Cauchy-Schwarz inequality, $|\mathrm{Cov}[W, Z]| \leq \sqrt{\mathrm{Var}[W] \cdot \mathrm{Var}[Z]}$. Thus,

$$
\begin{aligned}
\left| \mathbb{E}[W \cdot Z] \right| = \left| \mathbb{E}[W] \cdot \mathbb{E}[Z] + \mathrm{Cov}(W, Z) \right| \\
\leq \left| \mathbb{E}[W] \cdot \mathbb{E}[Z] \right| + \left| \mathrm{Cov}(W, Z) \right| \\
\leq \left| \mathbb{E}[W] \right| \cdot \left| \mathbb{E}[Z] \right| + \sqrt{\mathrm{Var}[W] \cdot \mathrm{Var}[Z]} \\
\leq O\left(\frac{d^{2.5}}{\sqrt{D}} + \frac{d^{O(d)}}{D} \right)
\end{aligned}
$$

where the last line combines Items 1 to 4. $\qquad\square$

A.2 Mean & Variance Bounds for Multilinear Monomials

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

- For vectors $a, b \in \mathbb{R}^D$, we will use $\widetilde{a}$ and $\widetilde{b}$ to denote the normalized vectors $\frac{a}{\|a\|_2}$ and $\frac{b}{\|b\|_2}$ respectively.

- We will use $\mathbf{U} \in \mathbb{R}^N$ to denote $M\widetilde{a}$ and similarly $\mathbf{V} \in \mathbb{R}^N$ to denote $M\widetilde{b}$. We will also have independent variables a' and b', for which we use $\mathbf{U}' = M\widetilde{a}'$ and $\mathbf{V}' = M\widetilde{b}'$.

- U_i denotes the i-th coordinate of U. Similarly, m_i denotes the vector corresponding to the i-th row of M. Note that $U_i = \langle m_i, \widetilde{a} \rangle$. For $S \subseteq [N]$, let $\mathbf{U}_S$ denote $\prod_{i \in S} U_i = \prod_{i \in S} \langle m_i, \widetilde{a} \rangle$. Similarly for $\mathbf{V}_S$.

- We will take expectations over random variables M, a, b, a', b'. It will be understood that we are sampling $M \sim \mathcal{N}(0,1)^{\otimes(N \times D)}$. Also, (a, b) and (a', b') are independently sampled from $\mathcal{G}_\rho^{\otimes D}$.

Lemma A.8 (Mean bounds for monomials). *Given parameter d and δ, there exists an explicitly computable $D := D(d, \delta)$ such that the following holds: For any subsets $S, T \subseteq [N]$ satisfying $|S|, |T| \leq d$, it holds that,*

$$
\begin{aligned}
\text{if } S \neq T : \quad & \mathbb{E}_M \mathbb{E}_{a,b} \mathbf{U}_S \mathbf{V}_T = 0. \\
\text{if } S = T : \quad & \left| \mathbb{E}_M \mathbb{E}_{a,b} \mathbf{U}_S \mathbf{V}_T - \rho^{|S|} \right| \leq \delta.
\end{aligned}
$$

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

Proof. We have that

$$
\begin{aligned}
\mathbb{E}_M \mathbb{E}_{a,b} \mathbf{U}_S \mathbf{V}_T &= \mathbb{E}_M \mathbb{E}_{a,b} \left[\prod_{i \in S} U_i \cdot \prod_{i \in T} V_i \right] \\
&= \mathbb{E}_{a,b} \mathbb{E}_M \left[\prod_{i \in S \cap T} U_i V_i \cdot \prod_{i \in S \setminus T} U_i \cdot \prod_{i \in T \setminus S} V_i \right] \\
&= \mathbb{E}_{a,b} \mathbb{E}_M \left[\prod_{i \in S \cap T} \langle m_i, \widetilde{a} \rangle \langle m_i, \widetilde{b} \rangle \cdot \prod_{i \in S \setminus T} \langle m_i, \widetilde{a} \rangle \cdot \prod_{i \in T \setminus S} \langle m_i, \widetilde{b} \rangle \right] \\
&= \mathbb{E}_{a,b} \left[\prod_{i \in S \cap T} \mathbb{E}_{m_i} \langle m_i, \widetilde{a} \rangle \langle m_i, \widetilde{b} \rangle \cdot \prod_{i \in S \setminus T} \mathbb{E}_{m_i} \langle m_i, \widetilde{a} \rangle \cdot \prod_{i \in T \setminus S} \mathbb{E}_{m_i} \langle m_i, \widetilde{b} \rangle \right], \qquad (27)
\end{aligned}
$$

36

where the last equality follows from the independence of the m_i's.

If $S \neq T$, then either $S \setminus T$ is non-empty in which case $\prod_{i \in S \setminus T} \mathbb{E}_{m_i}[\langle m_i, \tilde{a} \rangle] = 0$ or $T \setminus S$ is non-empty in which case $\prod_{i \in T \setminus S} \mathbb{E}_{m_i}[\langle m_i, \tilde{b} \rangle] = 0$. This is because for any fixed vector a and for each $i \in [N]$, the random variable $\langle m_i, \tilde{a} \rangle$ has zero-mean (and similarly for $\langle m_i, \tilde{b} \rangle$). The first part of the lemma now follows from Equation (27).

If $S = T$, Equation (27) becomes

$$
\begin{aligned}
\mathbb{E}_M \mathbb{E}_{a,b} U_S V_T &= \mathbb{E}_{a,b} \left[\prod_{i \in S} \mathbb{E}_{m_i} \frac{\langle m_i, a \rangle}{\|a\|_2} \frac{\langle m_i, b \rangle}{\|b\|_2} \right] \\
&= \mathbb{E}_{a,b} \left[\prod_{i \in S} \frac{\langle a, b \rangle}{\|a\|_2 \|b\|_2} \right] \qquad \left[\text{since } \mathbb{E}_{m_i} m_i \cdot m_i^T = I_{D \times D}. \right] \\
&= \prod_{i \in S} \left[\frac{\mathbb{E}_{a,b} \langle a, b \rangle}{D} \right] \pm \delta \\
&= \rho^{|S|} \pm \delta,
\end{aligned}
$$

where the penultimate equality above follows from Lemma A.1 for an explicit choice of D that is upper bounded by $\frac{d^{O(d)}}{\delta^2}$. □

Lemma A.9 (Covariance bounds for monomials). *Given parameters d and δ, there exists an explicitly computable $D := D(d, \delta)$ such that the following holds: For any subsets $S, T, S', T' \subseteq [N]$ satisfying $|S|, |T|, |S'|, |T'| \leq d$, it holds that,*

if $S \triangle T \triangle S' \triangle T' \neq \emptyset$: $\left| \mathbb{E}_M \mathbb{E}_{a,b} \mathbb{E}_{a',b'} [U_S V_T U'_{S'} V'_{T'}] - \left(\mathbb{E}_M \mathbb{E}_{a,b} [U_S V_T] \right) \cdot \left(\mathbb{E}_M \mathbb{E}_{a',b'} [U'_{S'} V'_{T'}] \right) \right| = 0$,

if $S \triangle T \triangle S' \triangle T' = \emptyset$: $\left| \mathbb{E}_M \mathbb{E}_{a,b} \mathbb{E}_{a',b'} [U_S V_T U'_{S'} V'_{T'}] - \left(\mathbb{E}_M \mathbb{E}_{a,b} [U_S V_T] \right) \cdot \left(\mathbb{E}_M \mathbb{E}_{a',b'} [U'_{S'} V'_{T'}] \right) \right| \leq \delta$.

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 $D = \frac{d^{O(d)}}{\delta^2}$.

In order to prove Lemma A.9, we need the following lemma.

Lemma A.10. *Let m be distributed as $\mathcal{N}(0, 1)^{\otimes D}$. Then,*

$$
\mathbb{E}_{a,b,a',b'} \left[\left(\mathbb{E}_m [\langle m, \tilde{a} \rangle \langle m, \tilde{b} \rangle \langle m, \tilde{a}' \rangle \langle m, \tilde{b}' \rangle] - \mathbb{E}_m [\langle m, \tilde{a} \rangle \langle m, \tilde{b} \rangle] \cdot \mathbb{E}_m [\langle m, \tilde{a}' \rangle \langle m, \tilde{b}' \rangle] \right)^2 \right] \leq O\left(\frac{1}{D^2} \right)
$$

and

$$
\mathbb{E}_{a,a'} \left[\left(\mathbb{E}_m [\langle m, \tilde{a} \rangle \langle m, \tilde{a}' \rangle] - \mathbb{E}_m [\langle m, \tilde{a} \rangle] \cdot \mathbb{E}_m [\langle m, \tilde{a}' \rangle] \right)^2 \right] \leq O\left(\frac{1}{D} \right).
$$

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

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

where we use that for any $i \in [D]$, it holds that $\mathbb{E}_m[m_i^4] = 3$ and $\mathbb{E}_m[m_i^2] = 1$. Thus,

$$
\begin{aligned}
\mathbb{E}_{a,b,a',b'} \left[T(a,b,a',b')^2 \right] &= \mathbb{E}_{a,b,a',b'} \left[\left[\langle \tilde{a}, \tilde{a}' \rangle \langle \tilde{b}, \tilde{b}' \rangle + \langle \tilde{a}, \tilde{b}' \rangle \langle \tilde{a}', \tilde{b} \rangle \right]^2 \right] \\
&\leq 2 \cdot \mathbb{E}_{a,b,a',b'} \left[\langle \tilde{a}, \tilde{a}' \rangle^2 \langle \tilde{b}, \tilde{b}' \rangle^2 \right] + 2 \cdot \mathbb{E}_{a,b,a',b'} \left[\langle \tilde{a}, \tilde{b}' \rangle^2 \langle \tilde{a}', \tilde{b} \rangle^2 \right] \\
&\leq O\left(\frac{1}{D^2} \right),
\end{aligned}
$$

where the last step follows by two applications of Lemma A.1 (with $d = 4$). This completes the proof of the first part of the lemma. The second part of the lemma similarly follows from Lemma A.1 (with $d = 2$) along with the fact that $\mathbb{E}_m[\langle m, \tilde{a} \rangle] = 0$. $\qquad \square$

Proof of Lemma A.9. Let $\mathbb{1}(E)$ denote the $0/1$ indicator function of an event E. We have that

$$
\begin{aligned}
\mathbb{E}_M \mathbb{E}_{a,b} \mathbb{E}_{a',b'} \left[\mathbf{U}_S \mathbf{V}_T \mathbf{U}'_{S'} \mathbf{V}'_{T'} \right] &= \mathbb{E}_M \mathbb{E}_{a,b} \mathbb{E}_{a',b'} \left[\prod_{i \in S \cup T \cup S' \cup T'} U_i^{\mathbb{1}(i \in S)} V_i^{\mathbb{1}(i \in T)} U_i'^{\mathbb{1}(i \in S')} V_i'^{\mathbb{1}(i \in T')} \right] \\
&= \mathbb{E}_{a,b} \mathbb{E}_{a',b'} \left[\prod_{i \in S \cup T \cup S' \cup T'} \mathbb{E}_{m_i} \left[U_i^{\mathbb{1}(i \in S)} V_i^{\mathbb{1}(i \in T)} U_i'^{\mathbb{1}(i \in S')} V_i'^{\mathbb{1}(i \in T')} \right] \right]. \quad (28)
\end{aligned}
$$

On the other hand, we have that

$$
\mathbb{E}_M \mathbb{E}_{a,b} \left[\mathbf{U}_S \mathbf{V}_T \right] = \mathbb{E}_{a,b} \mathbb{E}_M \left[\prod_{i \in S \cup T} U_i^{\mathbb{1}(i \in S)} V_i^{\mathbb{1}(i \in T)} \right] = \mathbb{E}_{a,b} \left[\prod_{i \in S \cup T} \mathbb{E}_{m_i} \left[U_i^{\mathbb{1}(i \in S)} V_i^{\mathbb{1}(i \in T)} \right] \right], \quad (29)
$$

and similarly

$$
\mathbb{E}_M \mathbb{E}_{a',b'} \left[\mathbf{U}'_{S'} \mathbf{V}'_{T'} \right] = \mathbb{E}_{a',b'} \left[\prod_{i \in S' \cup T'} \mathbb{E}_{m_i} \left[U_i'^{\mathbb{1}(i \in S')} V_i'^{\mathbb{1}(i \in T')} \right] \right]. \quad (30)
$$

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 (28) is equal to 0, and that at least one of the expectations in Equations (29) and (30) 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

$$
g_i(a,b,a',b') := \mathbb{E}_{m_i} \left[U_i^{\mathbb{1}(i \in S)} V_i^{\mathbb{1}(i \in T)} U_i'^{\mathbb{1}(i \in S')} V_i'^{\mathbb{1}(i \in T')} \right] \quad (31)
$$

$$
h_i(a,b) := \mathbb{E}_{m_i} \left[U_i^{\mathbb{1}(i \in S)} V_i^{\mathbb{1}(i \in T)} \right]. \quad (32)
$$

$$
h_i'(a',b') := \mathbb{E}_{m_i} \left[U_i'^{\mathbb{1}(i \in S')} V_i'^{\mathbb{1}(i \in T')} \right]. \quad (33)
$$

Combining Equations (28) to (30) along with the definitions in Equations (31) to (33), we get

$$
\begin{aligned}
&\left| \mathbb{E}_M \mathbb{E}_{a,b} \mathbb{E}_{a',b'} \left[\mathbf{U}_S \mathbf{V}_T \mathbf{U}'_{S'} \mathbf{V}'_{T'} \right] - \mathbb{E}_M \mathbb{E}_{a,b} \left[\mathbf{U}_S \mathbf{V}_T \right] \cdot \mathbb{E}_M \mathbb{E}_{a',b'} \left[\mathbf{U}'_{S'} \mathbf{V}'_{T'} \right] \right| \\
&= \left| \mathbb{E}_{a,b} \mathbb{E}_{a',b'} \left[\prod_{i=1}^{4d} g_i(a,b,a',b') - \prod_{i=1}^{4d} h_i(a,b) \cdot h_i'(a',b') \right] \right| \\
&= \left| \mathbb{E}_{a,b} \mathbb{E}_{a',b'} \left[\sum_{j=1}^{4d} \left[\prod_{i=1}^{j-1} h_i(a,b) \cdot h_i'(a',b') \prod_{i=j}^{4d} g_i(a,b,a',b') - \prod_{i=1}^{j} h_i(a,b) \cdot h_i'(a',b') \prod_{i=j+1}^{4d} g_i(a,b,a',b') \right] \right] \right|
\end{aligned}
$$

38

$$\leq \sum_{j=1}^{4d} \left| \mathop{\mathbb{E}}_{a,b} \mathop{\mathbb{E}}_{a',b'} \left[\prod_{i=1}^{j-1} h_i(a,b) \cdot h_i'(a',b') \prod_{i=j+1}^{4d} g_i(a,b,a',b') \cdot \left[g_j(a,b,a',b') - h_j(a,b) \cdot h_j'(a',b') \right] \right] \right|$$

$$\leq 4 \cdot d \cdot \sqrt{\tau \cdot \kappa},$$

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

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

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

Lemma A.10 implies that $\kappa \leq O(1/D)$. We now show that $\tau \leq 2^{O(d)}$. Note that for any $i \in [D]$, it holds that,

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

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

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

A.3 Mean & Variance Bounds for Multilinear Polynomials

We are now ready to prove Lemma 3.3. Recall that,

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

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

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

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

Proof of Lemma 3.3.   We start out by proving the bound on $\left| \mathbb{E}_M F(M) - \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes N}} \right|$. To this end, we will use Lemma A.8 with parameters d and δ. Thus, for a choice of $D = d^{O(d)}/\delta^2$, we have that,

$$
\left| \mathbb{E}_M F(M) - \langle A, B \rangle_{\mathcal{G}_\rho^{\otimes N}} \right|
$$

$$
= \left| \mathbb{E}_M \underset{\boldsymbol{a,b}}{\mathbb{E}} [A(\mathbf{U}) \cdot B(\mathbf{V})] - \underset{\mathbf{X,Y} \sim \mathcal{G}_\rho^{\otimes N}}{\mathbb{E}} [A(\mathbf{X}) \cdot B(\mathbf{Y})] \right|
$$

$$
= \left| \sum_{S,T \subseteq [N]} \widehat{A}_S \widehat{B}_T \cdot \left(\underset{M}{\mathbb{E}} \underset{\boldsymbol{a,b}}{\mathbb{E}} [\mathbf{U}_S \cdot \mathbf{V}_T] - \mathbb{E}_{\mathbf{X,Y} \sim \mathcal{G}_\rho^{\otimes N}} [\mathbf{X}_S \cdot \mathbf{Y}_T] \right) \right|
$$

$$
= \left| \sum_{S \subseteq [N]} \widehat{A}_S \widehat{B}_S \cdot \left(\underset{M}{\mathbb{E}} \underset{\boldsymbol{a,b}}{\mathbb{E}} [\mathbf{U}_S \cdot \mathbf{V}_S] - \rho^{|S|} \right) \right| \qquad \dots \text{ (terms corresponding to } S \neq T \text{ are 0.)}
$$

$$
\leq \sum_{S \subseteq [N]} \left| \widehat{A}_S \widehat{B}_S \right| \cdot \delta \qquad \dots\dots \text{ (using Lemma A.8)}
$$

$$
\leq \|A\|_2 \cdot \|B\|_2 \cdot \delta \qquad \dots\dots \text{ (Cauchy-Schwarz inequality)}
$$

$$
\leq \delta \qquad \dots\dots (\|A\|_2, \|B\|_2 \leq 1)
$$

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

$$
\underset{M}{\mathbb{E}} \left(\underset{\boldsymbol{a,b}}{\mathbb{E}} A(\mathbf{U}) \cdot B(\mathbf{V}) \right)^2 - \left(\underset{M}{\mathbb{E}} \underset{\boldsymbol{a,b}}{\mathbb{E}} A(\mathbf{U}) \cdot B(\mathbf{V}) \right)^2
$$

$$
= \left| \underset{M}{\mathbb{E}} \underset{\boldsymbol{a,b}}{\mathbb{E}} \underset{\boldsymbol{a',b'}}{\mathbb{E}} [A(\mathbf{U})B(\mathbf{V})A(\mathbf{U'})B(\mathbf{V'})] - \left(\underset{M}{\mathbb{E}} \underset{\boldsymbol{a,b}}{\mathbb{E}} [A(\mathbf{U})B(\mathbf{V})] \right) \cdot \left(\underset{M}{\mathbb{E}} \underset{\boldsymbol{a',b'}}{\mathbb{E}} [A(\mathbf{U'})B(\mathbf{V'})] \right) \right|
$$

$$
\leq \sum_{\substack{S,T \subseteq [N] \\ S',T' \subseteq [N]}} \left| \widehat{A}_S \widehat{B}_T \widehat{A}_{S'} \widehat{B}_{T'} \right| \cdot \left| \underset{M}{\mathbb{E}} \underset{\boldsymbol{a,b}}{\mathbb{E}} \underset{\boldsymbol{a',b'}}{\mathbb{E}} [\mathbf{U}_S \mathbf{V}_T \mathbf{U}'_{S'} \mathbf{V}'_{T'}] - \left(\underset{M}{\mathbb{E}} \underset{\boldsymbol{a,b}}{\mathbb{E}} [\mathbf{U}_S \mathbf{V}_T] \right) \cdot \left(\underset{M}{\mathbb{E}} \underset{\boldsymbol{a',b'}}{\mathbb{E}} [\mathbf{U}'_{S'} \mathbf{V}'_{T'}] \right) \right|
$$

$$
\leq \frac{\delta}{9^d} \cdot \sum_{\substack{S,T,S',T' \subseteq [N] \\ S \triangle T \triangle S' \triangle T' = \varnothing}} \left| \widehat{A}_S \widehat{B}_T \widehat{A}_{S'} \widehat{B}_{T'} \right| .
$$

To finish the proof, we will show that,

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

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 [Wol07] for degree-d polynomials over the boolean hypercube imply that,

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

We now finish the proof as follows,

$$
\sum_{\substack{S,T,S',T' \subseteq [N] \\ S \triangle T \triangle S' \triangle T' = \varnothing}} \left| \widehat{A}_S \widehat{B}_T \widehat{A}_{S'} \widehat{B}_{T'} \right| = \underset{x}{\mathbb{E}} \left[f(x)^2 g(x)^2 \right]
$$

$$\leq \left(\mathop{\mathbb{E}}_{x} \left[f(x)^4 \right] \right)^{1/2} \cdot \left(\mathop{\mathbb{E}}_{x} \left[g(x)^4 \right] \right)^{1/2}$$

$$\leq 9^d \cdot \left(\mathop{\mathbb{E}}_{x} \left[f(x)^2 \right] \right) \cdot \left(\mathop{\mathbb{E}}_{x} \left[g(x)^2 \right] \right)$$

$$= 9^d \cdot \left(\sum_S \widehat{A}_S^2 \right) \cdot \left(\sum_S \widehat{B}_S^2 \right)$$

$$= 9^d \cdot \|A\|_2^2 \cdot \|B\|_2^2 .$$

Thus, overall we get that, $\mathsf{Var}_M(F(M)) \leq \delta$.

This completes the proof of Lemma 3.3 for an explicit choice of D that is upper bounded by $d^{O(d)}/\delta^2$. $\square$

B   Proof of Main Smoothing Lemma

In order to prove Lemma 4.1, we consider the definition of *low-degree truncation*.

Definition B.1 (Low-degree truncation). *We define this for functions in $L^2(\mathcal{Z}^n, \mu^{\otimes n})$ and also for those in $L^2(\mathbb{R}^n, \gamma_n)$.*

Discrete Hypercube: *Suppose $A \in L^2(\mathcal{Z}^n, \mu^{\otimes n})$ is given by the Fourier expansion $A(\boldsymbol{x}) = \sum\limits_{\boldsymbol{\sigma} \in \mathbb{Z}_q^n} \widehat{A}_{\boldsymbol{\sigma}} \mathcal{X}_{\boldsymbol{\sigma}}(\boldsymbol{x})$.*

The degree-d truncation of A is defined as the function $A^{\leq d} \in L^2(\mathcal{Z}^n, \mu^{\otimes n})$ given by

$$A^{\leq d}(\boldsymbol{x}) := \sum_{\substack{\boldsymbol{\sigma} \in \mathbb{Z}_q^n \\ |\boldsymbol{\sigma}| \leq d}} \widehat{A}_{\boldsymbol{\sigma}} \mathcal{X}_{\boldsymbol{\sigma}}(\boldsymbol{x}).$$

That is, $A^{\leq d}$ is obtained by retaining only the terms with degree at most d in the Fourier expansion of A, where recall that for $\boldsymbol{\sigma} \in \mathbb{Z}_q^n$, its degree is defined as $|\boldsymbol{\sigma}| = \{i \in [n] : \boldsymbol{\sigma}_i \neq 0\}$.

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

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

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 $\boldsymbol{\sigma} \in \mathbb{Z}_{\geq 0}^n$, its degree is defined as $|\boldsymbol{\sigma}| = \sum_{i=1}^n \sigma_i$.

For convenience, in either case, 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.

To prove the discrete part of Lemma 4.1, we will use a lemma from [Mos10, Lemma 6.1], which is proved using Efron-Stein decompositions. To state this lemma, we first introduce the Bonami-Beckner operator.

Definition B.2 (Bonami-Beckner operator). *For any $v \in [0,1]$, the Bonami-Beckner operator T_v on a probability space $(\mathcal{Z}, \mu)$ is given by its action on any $f : \mathcal{Z} \to \mathbb{R}$, as follows,*

$$(T_v f)(x) = \mathbb{E}[f(Y)|X = x]$$

where the conditional distribution of Y given $X = x$ is $\nu\delta_x + (1 - \nu)\mu$ where δ_x is the delta measure on x. In other words, given $X = x$, we obtain Y by either setting it to x with probability ν or independently sampling from μ with probability $(1 - \nu)$.

For the product space $(\mathcal{Z}^n, \mu^{\otimes n})$, we define the Bonami-Beckner operator T_ν as, $T_\nu = \otimes_{i=1}^n T_\nu^{(i)}$, where $T_\nu^{(i)}$ is the Bonami-Beckner operator on the i-th coordinate $(\mathcal{Z}, \mu)$.

We now state a specialized version of Mossel's lemma, which suffices for our application.

Lemma B.3 ([Mos10]). *Let $(\mathcal{Z} \times \mathcal{Z}, \mu)$ be finite joint probability space, such that $\rho(\mathcal{Z}, \mathcal{Z}; \mu) = \rho$ for some $\rho \in [0,1]$. Let $P \in L^2(\mathcal{Z}^n, \mu_A^{\otimes n})$ and $Q \in L^2(\mathcal{Z}^n, \mu_B^{\otimes n})$ be multi-linear polynomials. Let $\varepsilon > 0$ and ν be chosen sufficiently close to 1 so that,*

$$\nu \geq (1 - \varepsilon)^{\log\rho/(\log\varepsilon + \log\rho)}$$

Then:

$$\left| \langle P, Q \rangle_{\mu^{\otimes n}} - \langle T_\nu P, T_\nu Q \rangle_{\mu^{\otimes n}} \right| \leq \varepsilon \cdot \sqrt{\mathsf{Var}[P]\,\mathsf{Var}[Q]}$$

In particular, there exists an absolute constant C such that it suffices to take

$$\nu \overset{\text{def}}{=} 1 - C\frac{(1 - \rho)\varepsilon}{\log(1/\varepsilon)}$$

To prove the Gaussian version of Lemma 4.1, we will need the analog of the above lemma for correlated Gaussian spaces which can be proved in a similar way, using Hermite expansions instead of Efron-Stein decompositions. Here, we use the Ornstein-Uhlenbeck operator U_ν instead of the Bonami-Beckner operator T_ν. In particular, the following lemma holds.

Lemma B.4. *Consider the correlated Gaussian space $\mathcal{G}_\rho^{\otimes n}$ for some $\rho \in [0,1]$. Let $P \in L^2(\mathbb{R}^n, \gamma_n)$ and $Q \in L^2(\mathbb{R}^n, \gamma_n)$. Let $\varepsilon > 0$ and ν be chosen sufficiently close to 1 so that,*

$$\nu \geq (1 - \varepsilon)^{\log\rho/(\log\varepsilon + \log\rho)}$$

Then:

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

In particular, there exists an absolute constant C such that it suffices to take

$$\nu \overset{\text{def}}{=} 1 - C\frac{(1 - \rho)\varepsilon}{\log(1/\varepsilon)}$$

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

$$P(\mathbf{X}) = \sum_{\boldsymbol{\sigma} \in \mathbb{Z}_{\geq 0}^n} \widehat{P}(\boldsymbol{\sigma}) H_{\boldsymbol{\sigma}}(\mathbf{X}) \qquad \text{and} \qquad Q(\mathbf{Y}) = \sum_{\boldsymbol{\sigma} \in \mathbb{Z}_{\geq 0}^n} \widehat{Q}(\boldsymbol{\sigma}) H_{\boldsymbol{\sigma}}(\mathbf{Y}).$$

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

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

Note that our choice of ν gives us that, $\rho^d \left(1 - \nu^{2d}\right) \leq \varepsilon$ for all $d \in \mathbb{N}$. Thus, we get that,

$$\left| \langle P, Q \rangle_{\mathcal{G}_\rho^{\otimes n}} - \langle U_\nu P, U_\nu Q \rangle_{\mathcal{G}_\rho^{\otimes n}} \right|$$
$$= \left| \sum_{\boldsymbol{\sigma} \in \mathbb{Z}_{\geq 0}^n} \rho^{|\boldsymbol{\sigma}|} \cdot \widehat{P}(\boldsymbol{\sigma}) \widehat{Q}(\boldsymbol{\sigma}) \cdot \left(1 - \nu^{2|\boldsymbol{\sigma}|}\right) \right|$$

$$\leq \sum_{\boldsymbol{\sigma} \in \mathbb{Z}_{\geq 0}^n} |\widehat{P}(\boldsymbol{\sigma})\widehat{Q}(\boldsymbol{\sigma})| \cdot \rho^{|\boldsymbol{\sigma}|}\left(1 - \nu^{2|\boldsymbol{\sigma}|}\right)$$

$$\leq \varepsilon \cdot \sum_{\boldsymbol{\sigma} \in \mathbb{Z}_{\geq 0}^n} |\widehat{P}(\boldsymbol{\sigma})\widehat{Q}(\boldsymbol{\sigma})| \qquad \dots (\text{since, } \rho^d\left(1 - \nu^{2d}\right) \leq \varepsilon \text{ for all } d \in \mathbb{N})$$

$$\leq \varepsilon \cdot \sqrt{\mathrm{Var}[P]\,\mathrm{Var}[Q]} \qquad \dots (\text{Cauchy-Schwarz inequality})$$

$\square$

We are now ready to prove our main smoothing lemma (Lemma 4.1).

Proof of Lemma 4.1. We prove the lemma for the case of correlated discrete hypercubes, i.e. $(\mathcal{Z} \times \mathcal{Z}, \mu)$. The proof for the correlated Gaussian case follows similarly.

   We obtain $A^{(1)}$ and $B^{(1)}$ in two steps. In the first step we apply some suitable amount of noise to the functions such that the functions have decaying Fourier tails. In the second step, we truncate the Fourier coefficients corresponding to terms larger than degree d.

Noising step. In this step, we obtain intermediate functions $\overline{A} : \mathcal{Z}^n \to \mathbb{R}^k$ and $\overline{B} : \mathcal{Z}^n \to \mathbb{R}^k$ such that,

1. $\overline{A}$ and $\overline{B}$ have decaying Fourier tails. In particular, for any $j \in [k]$: $\left\|\overline{A}_j^{>d}\right\|_2, \left\|\overline{B}_j^{>d}\right\|_2 \leq \frac{\delta}{2\sqrt{k}}$.

2. $\mathrm{Var}(\overline{A}_j) \leq \mathrm{Var}(A_j)$ and $\mathrm{Var}(\overline{B}_j) \leq \mathrm{Var}(B_j)$, for any $j \in [k]$.

3. $\left\|\mathcal{R}(\overline{A}) - \overline{A}\right\|_2 \leq \left\|\mathcal{R}(A) - A\right\|_2$ and $\left\|\mathcal{R}(\overline{B}) - \overline{B}\right\|_2 \leq \left\|\mathcal{R}(B) - B\right\|_2$.

4. For every $i, j \in [k]$: $\left|\langle \overline{A}_i, \overline{B}_j \rangle_{\mu^{\otimes n}} - \langle A_i, B_j \rangle_{\mu^{\otimes n}}\right| \leq \frac{\delta}{2\sqrt{k}}$.

   Firstly, note that we have $\mathrm{Var}[A_j], \mathrm{Var}[B_j] \leq 1$ for any $j \in [k]$. Given parameter δ, we first choose ε and ν in Lemma B.3, 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 $\nu^{2d} \leq \frac{\delta}{2\sqrt{k}}$, that is, $d = O\left(\frac{\log(k/\delta)}{\log(1/\nu)}\right) = O\left(\frac{\sqrt{k}\log^2(k/\delta)}{\delta(1-\rho)}\right)$.

   Let $\overline{A} = T_\nu A$ and $\overline{B} = T_\nu B$. We get the above statements as follows,

1. $\left\|\overline{A}_j^{>d}\right\|_2 = \sum_{\substack{\boldsymbol{\sigma} \in \mathbb{Z}_q^n \\ |\boldsymbol{\sigma}| > d}} \nu^{2|\boldsymbol{\sigma}|} \cdot \widehat{A_j}(\boldsymbol{\sigma})^2 \leq \nu^{2d} \cdot \mathrm{Var}(A_j) \leq \frac{\delta}{2\sqrt{k}}$. Similarly, $\left\|\overline{B}_j^{>d}\right\|_2 \leq \frac{\delta}{2\sqrt{k}}$.

2. $\mathrm{Var}(\overline{A}_j) = \sum_{\boldsymbol{\sigma} \in \mathbb{Z}_q^n} \nu^{2|\boldsymbol{\sigma}|} \cdot \widehat{A_j}(\boldsymbol{\sigma})^2 \leq \mathrm{Var}(A_j)$. Similarly, $\mathrm{Var}(\overline{B}) \leq \mathrm{Var}(B)$.

3. Observe that $\left\|\mathcal{R}(v) - v\right\|_2$ is the Euclidean distance of $v \in \mathbb{R}^k$ from the simplex Δ_k, which is a convex body. Hence $\left\|\mathcal{R}(v) - v\right\|_2^2$ is convex function in v. Thus, we have that,

$$\left\|\mathcal{R}(\overline{A}) - \overline{A}\right\|_2^2$$
$$= \mathbb{E}_{\boldsymbol{x} \sim \mu_A^{\otimes n}} \left\|\mathcal{R}(\overline{A}(\boldsymbol{x})) - \overline{A}(x)\right\|_2^2$$
$$= \mathbb{E}_{\boldsymbol{x} \sim \mu_A^{\otimes n}} \left\|\mathcal{R}\left(\mathbb{E}_{\boldsymbol{x}' \sim T_\nu(\boldsymbol{x})} A(\boldsymbol{x}')\right) - \mathbb{E}_{\boldsymbol{x}' \sim T_\nu(\boldsymbol{x})} A(\boldsymbol{x}')\right\|_2^2$$
$$\leq \mathbb{E}_{\boldsymbol{x} \sim \mu_A^{\otimes n}} \mathbb{E}_{\boldsymbol{x}' \sim T_\nu(\boldsymbol{x})} \left\|\mathcal{R}(A(\boldsymbol{x}')) - A(\boldsymbol{x}')\right\|_2^2 \qquad \dots (\text{using convexity of } \left\|\mathcal{R}(v) - v\right\|_2^2)$$
$$= \mathbb{E}_{\boldsymbol{x}' \sim \mu_A^{\otimes n}} \left\|\mathcal{R}(A(\boldsymbol{x}')) - A(\boldsymbol{x}')\right\|_2^2$$

$$= \left\| \mathcal{R}(A) - A \right\|_2^2 .$$

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

4. For every $i, j \in [k]$, we simply have from Lemma B.3 that $\left| \langle \overline{A}_i, \overline{B}_j \rangle_{\mu^{\otimes n}} - \langle A_i, B_j \rangle_{\mu^{\otimes n}} \right| \leq \varepsilon = \frac{\delta}{2\sqrt{k}}$.

Low-degree truncation step. In this step, we obtain the final $A^{(1)}$ and $B^{(1)}$ such that,

1. $A^{(1)}$ and $B^{(1)}$ have degree at most d.

2. $\mathsf{Var}(A^{(1)}) \leq \mathsf{Var}(\overline{A})$ and $\mathsf{Var}(B^{(1)}) \leq \mathsf{Var}(\overline{B})$.

3. $\left\| \mathcal{R}(A^{(1)}) - A^{(1)} \right\|_2 \leq \left\| \mathcal{R}(\overline{A}) - \overline{A} \right\|_2 + \delta/2$ and $\left\| \mathcal{R}(B^{(1)}) - B^{(1)} \right\|_2 \leq \left\| \mathcal{R}(\overline{B}) - \overline{B} \right\|_2 + \delta/2$

4. For every $i, j \in [k] : \left| \langle A_i^{(1)}, B_j^{(1)} \rangle_{\mu^{\otimes n}} - \langle \overline{A}_i, \overline{B}_j \rangle_{\mu^{\otimes n}} \right| \leq \frac{\delta}{2\sqrt{k}}$

It is easy to see that combining statements 1-4 above, with statements 1-4 in the Noising step, we get all the desired conditions in Lemma 4.1.

In this step, we let $A^{(1)} = \overline{A}^{\leq d}$ and $B^{(1)} = \overline{B}^{\leq d}$ (i.e. degree-d truncation on every coordinate $j \in [k]$). We get the above statements as follows,

1. By definition of degree-d truncation, we have that $A^{(1)}$ and $B^{(1)}$ have degree at most d.

2. $\mathsf{Var}(A_j^{(1)}) = \sum_{\substack{\sigma \in \mathbb{Z}_q^n \\ |\sigma| \leq d}} \nu^{2|\sigma|} \cdot \widehat{A}_j^{(1)}(\sigma)^2 \leq \mathsf{Var}(\overline{A}_j)$. Similarly, $\mathsf{Var}(B_j^{(1)}) \leq \mathsf{Var}(\overline{B}_j)$.

3. We have that,

$$\begin{aligned} \left\| \mathcal{R}(A^{(1)}) - A^{(1)} \right\|_2 &\leq \left\| \mathcal{R}(\overline{A}) - \overline{A} \right\|_2 + \left\| \overline{A} - A^{(1)} \right\|_2 && \text{(Lemma 2.5)} \\ &= \left\| \mathcal{R}(\overline{A}) - \overline{A} \right\|_2 + \left\| \overline{A}^{>d} \right\|_2 \\ &\leq \left\| \mathcal{R}(\overline{A}) - \overline{A} \right\|_2 + \sqrt{k} \cdot \frac{\delta}{2\sqrt{k}} \\ &\leq \left\| \mathcal{R}(\overline{A}) - \overline{A} \right\|_2 + \delta/2 . \end{aligned}$$

Similarly for $B^{(1)}$.

4. We have that $\left\| A_i^{(1)} - \overline{A}_i \right\|_2 \leq \frac{\delta}{2\sqrt{k}}$ and $\left\| B_j^{(1)} - \overline{B}_j \right\|_2 \leq \frac{\delta}{2\sqrt{k}}$. Hence, using Lemma 2.4, we get that for every $i, j \in [k]$

$$\left| \langle A_i^{(1)}, B_j^{(1)} \rangle_{\mu^{\otimes n}} - \langle \overline{A}_i, \overline{B}_j \rangle_{\mu^{\otimes n}} \right| \leq \frac{\delta}{\sqrt{k}}$$

This completes the proof of Lemma 4.1 for the case of correlated discrete hypercubes. The proof for the case of correlated Gaussians follows in almost the same way. The only change is that we use U_ν operator instead of T_ν operator, and use Lemma B.4 instead of Lemma B.3. $\qquad\square$

C   Proof of Multi-linearization Lemma

In order to prove the lemma, we consider the definition of a *multi-linear truncation*.

Definition C.1 (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.

The proof of Lemma 5.1 will proceed by simply applying the transformation given in following lemma to each coordinate of A and B. The following lemma shows that low-degree polynomials over $\mathbb{R}^n$ can be converted to multilinear polynomials without hurting the correlation. This is done by slightly increasing the number of variables. In addition, we also get that these new polynomials have small individual influences.

Lemma C.2. *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 $\overline{\mathbf{X}} := \left\{ X^{(i)}_j : (i,j) \in [n] \times [t] \right\}$ and $\overline{\mathbf{Y}} := \left\{ Y^{(i)}_j : (i,j) \in [n] \times [t] \right\}$ respectively, as,

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

where $X^{(i)} = \left(X^{(i)}_1 + \cdots + X^{(i)}_t \right) / \sqrt{t}$ and $Y^{(i)} = \left(Y^{(i)}_1 + \cdots + Y^{(i)}_t \right) / \sqrt{t}$. Note that, intuitively this doesn't change the "structure" of A and B. In particular, it is easy to see that,

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

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 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. *For all $(i,j) \in [n] \times [t]$, it holds that $\mathsf{Inf}_{X^{(i)}_j}\left(\overline{A}^{\mathrm{ml}} \right) \leq \delta$ and $\mathsf{Inf}_{Y^{(i)}_j}\left(\overline{B}^{\mathrm{ml}} \right) \leq \delta.$*

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

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

In order to prove Lemma C.2, we will need the following multinomial theorem for Hermite polynomials. It can be proved quite easily using the generating function for Hermite polynomials.

Fact C.3 (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 C.2. Before we prove the theorem, we will first understand the effect of the transformation from X to $\overline{X}$ for a single Hermite polynomial. Instantiating β_i's in Fact C.3 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,

$$\left\| H^{\mathrm{nml}} \right\|_2^2 = \sum_{\substack{d_1, \ldots, d_t \in \mathbb{Z}_{\geq 0} \\ d_1 + \cdots + d_t = d \\ \exists i \ d_i \geq 2}} \left(\frac{1}{t^{d/2}} \cdot \sqrt{\frac{d!}{d_1! \cdots d_t!}} \right)^2 \leq O\left(\frac{d^2 t^{d-1}}{d!} \right) \cdot \frac{d!}{t^d} \leq O\left(\frac{d^2}{t} \right) \quad (34)$$

More generally, if we consider a term $\overline{H}_{\boldsymbol{\sigma}}(\overline{\mathbf{X}}) = H_{\sigma_1}(X^{(1)}) \cdot H_{\sigma_2}(X^{(2)}) \cdots H_{\sigma_N}(X^{(N)})$, where each $X^{(i)} = \left(X_1^{(i)} + \cdots + X_t^{(i)} \right) / \sqrt{t}$. Let's write $\overline{H}_{\boldsymbol{\sigma}}(\overline{\mathbf{X}}) = \overline{H}_{\boldsymbol{\sigma}}^{\mathrm{ml}}(\overline{\mathbf{X}}) + \overline{H}_{\boldsymbol{\sigma}}^{\mathrm{nml}}(\overline{\mathbf{X}})$, 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 $\left\| \overline{H}_{\boldsymbol{\sigma}}^{\mathrm{nml}} \right\|_2$, which can be done as follows,

$$\left\| \overline{H}_{\boldsymbol{\sigma}}^{\mathrm{nml}} \right\|_2^2 = \left\| \prod_{i=1}^{n} (H_{\sigma_i}^{\mathrm{ml}} + H_{\sigma_i}^{\mathrm{nml}}) - \prod_{i=1}^{n} H_{\sigma_i}^{\mathrm{ml}} \right\|_2^2$$

$$\leq \prod_{i=1}^{n} \left(1 + O\left(\frac{\sigma_i^2}{t} \right) \right) - 1 \qquad \text{(from Equation (34))}$$

$$\leq O\left(\frac{|\boldsymbol{\sigma}|^2}{t} \right) \qquad \text{(since, } t \gg |\boldsymbol{\sigma}|^2\text{)}$$

Thus, $\quad \left\| \overline{H}_{\boldsymbol{\sigma}}^{\mathrm{nml}} \right\|_2^2 < \delta^2 / 4.$ $\qquad$ (for $t = \Theta(d^2/\delta^2)$) $\qquad$ (35)

We are now ready to prove the parts of our Lemma C.2.

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 $\mathsf{Var}(\overline{A}^{\mathrm{ml}})$ 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 $A(\mathbf{X}) = \sum_{\boldsymbol{\sigma} \in \mathbb{Z}_{\geq 0}^N} \widehat{A}(\boldsymbol{\sigma}) \cdot H_{\boldsymbol{\sigma}}(\mathbf{X})$. Note that, $\overline{A}^{\mathrm{nml}}(\overline{\mathbf{X}}) = \sum_{\boldsymbol{\sigma} \in \mathbb{Z}_{\geq 0}^N} \widehat{A}(\boldsymbol{\sigma}) \cdot \overline{H}_{\boldsymbol{\sigma}}^{\mathrm{nml}}(\overline{\mathbf{X}})$, where

recall that $\overline{H}_{\sigma}^{\mathrm{nml}}$ is the non-multilinear part of $\overline{H}_{\sigma}(\mathbf{X}) = H_{\sigma_1}(X^{(1)}) \cdot H_{\sigma_2}(X^{(2)}) \cdots H_{\sigma_n}(X^{(N)})$, where each $X^{(i)} = \left(X_1^{(i)} + \cdots + X_t^{(i)}\right)/\sqrt{t}$.

From Equation 35, we have that for any $\sigma \in \mathbb{Z}_{\geq 0}^N$, it holds that $\left\|\overline{H}_{\sigma}^{\mathrm{nml}}(\overline{X})\right\|_2^2 < \delta^2/4$. And hence we get that,

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

Note that, here we use that $\overline{H}_{\sigma}(\overline{X})$ 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. We prove that $\mathrm{Inf}_{X_j^{(i)}}\left(\overline{A}^{\mathrm{ml}}\right) \leq d/t < \delta^2/d < \delta$. The case $\mathrm{Inf}_{Y_j^{(i)}}\left(\overline{B}^{\mathrm{ml}}\right)$ will follow similarly.

   For simplicity, let's first consider the case of a univariate polynomial $P \in L^2(\mathbb{R}, \gamma_1)$ of degree-d, such that $\|P\|_2 \leq 1$. We will show that for the function $\overline{P}(X_1, \cdots, X_t) := P\left(\frac{X_1 + \cdots + X_t}{\sqrt{t}}\right)$, it holds that $\mathrm{Inf}_{X_i}\left(\overline{P}^{\mathrm{ml}}\right) \leq \mathrm{Var}(P) \cdot (d/t)$. This follows from a few simple observations:

   (a) By symmetry, $\mathrm{Inf}_{X_i}\left(\overline{P}^{\mathrm{ml}}\right)$ is the same for all i.

   (b) For degree-d multilinear polynomials, $\sum_{i \in [t]} \mathrm{Inf}_{X_i}\left(\overline{P}^{\mathrm{ml}}\right) \leq d\,\mathrm{Var}\left(\overline{P}^{\mathrm{ml}}\right)$.

   (c) $\mathrm{Var}\left(\overline{P}^{\mathrm{ml}}\right) \leq \mathrm{Var}\left(\overline{P}\right) = \mathrm{Var}(P)$

   Thus, combining the above, we get that $\mathrm{Inf}_{X_i}\left(\overline{P}^{\mathrm{ml}}\right) \leq \mathrm{Var}(P) \cdot (d/t)$.

   In the more general case n-variate case, we observe that,

$$
\begin{aligned}
\mathrm{Inf}_{X_j^{(i)}}\left(\overline{A}^{\mathrm{ml}}\right) &= \underset{\overline{X}\setminus\{X_j^{(i)}\}}{\mathbb{E}} \underset{X_j^{(i)}}{\mathrm{Var}}\left(\overline{A}_{\overline{X}\setminus\{X_j^{(i)}\}}^{\mathrm{ml}}\left(X_j^{(i)}\right)\right) \\
&= \underset{\mathbf{X}^{(-i)}}{\mathbb{E}} \underset{\mathbf{X}_{-j}^{(i)}}{\mathbb{E}} \underset{X_j^{(i)}}{\mathrm{Var}}\left(\overline{A}_{\overline{X}\setminus\{X_j^{(i)}\}}^{\mathrm{ml}}\left(X_j^{(i)}\right)\right) \quad \text{(for simplicity, } \mathbf{X}^{(-i)} := \overline{\mathbf{X}} \setminus \left\{X_j^{(i)} : j \in [t]\right\}) \\
&= \underset{\mathbf{X}^{(-i)}}{\mathbb{E}} \mathrm{Inf}_{X_j^{(i)}}\left(\overline{A}_{X^{(-i)}}^{\mathrm{ml}}(X^{(i)})\right) \\
&= \underset{\mathbf{X}^{(-i)}}{\mathbb{E}} \mathrm{Var}\left(\overline{A}_{X^{(-i)}}^{\mathrm{ml}}\right) \cdot (d/t) \\
&\leq (d/t) \cdot \mathrm{Inf}_i(A) \\
&\leq d/t \\
&< \delta.
\end{aligned}
$$

   where, in last two inequalities, we use that $\mathrm{Inf}_i(A) \leq \mathrm{Var}(A) \leq 1$ and that $t = \Theta(d^2/\delta^2)$.

5. 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 2.4, 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 $\left\|\overline{B}^{\mathrm{ml}}\right\|_2 \leq \|\overline{B}\|_2 \leq 1$ and $\left\|\overline{A}^{\mathrm{ml}}\right\|_2 \leq \|\overline{A}\|_2 \leq 1$.

□

We are now able to prove Lemma 5.1.

Proof of Lemma 5.1. We apply the transformation in Lemma C.2, 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 function $A^{(1)} : \mathbb{R}^{nt} \to \mathbb{R}^k$ and $B^{(1)} : \mathbb{R}^n \to \mathbb{R}^k$. Namely, for any $j \in [k]$, we set $A_j^{(1)}(\overline{\mathbf{X}}) = \overline{A}_j^{\text{ml}}(\overline{\mathbf{X}})$ and $B_j^{(1)}(\overline{\mathbf{Y}}) = \overline{B}_j^{\text{ml}}(\overline{\mathbf{Y}})$ as described in Lemma C.2.

It is easy to see that parts 1, 2, 4, 5 follow immediately from the conditions satisfied in Lemma C.2. For part 3, we note that we have that $\left\| \overline{A}_j^{\text{ml}} - \overline{A}_j \right\|_2 \leq \delta/\sqrt{k}$ for every $j \in [k]$. Thus, combining these for all $j \in [k]$, we get that $\left\| \overline{A}^{\text{ml}} - \overline{A} \right\|_2 \leq \delta$. Now, using Lemma 2.5, we immediately get that,

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

Finally, we observe that $\left\| \mathcal{R}(\overline{A}) - \overline{A} \right\|_2 = \left\| \mathcal{R}(A) - A \right\|_2$, to conclude that

$$\left\| \mathcal{R}(A^{(1)}) - A^{(1)} \right\|_2 \leq \left\| \mathcal{R}(A) - A \right\|_2 + \delta \, .$$

Similarly, $\left\| \mathcal{R}(B^{(1)}) - B^{(1)} \right\|_2 \leq \left\| \mathcal{R}(B) - B \right\|_2 + \delta$. This concludes the proof.

□

D Regularity Lemma for low-degree functions

In this section we state and prove a regularity lemma that we need for proving Theorem 1.5, i.e. non-interactive simulation from discrete sources. Our regularity lemma follows immediately from the version stated in [GKS16b], which was inspired from [DSTW10].

We begin by first recalling the basic notions of influences and partial restrictions of functions over product spaces.

Definition D.1 (Influence). *For every coordinate $i \in [n]$, $\text{Inf}_i(f)$ is the i-th influence of f, and $\text{Inf}(f)$ is the total influence, which are defined as*

$$\text{Inf}_i(f) \stackrel{\text{def}}{=} \mathbb{E}_{\boldsymbol{x}_{-i}} \left[\underset{x_i}{\text{Var}} \left[f(\boldsymbol{x}) \right] \right] \qquad \text{Inf}(f) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \text{Inf}_i(f)$$

The basic properties of influence are summarized in the following fact.

Fact D.2. *For any function $f \in L^2(\mathcal{Z}^n, \mu_A^{\otimes n})$, we have the following:*

(i) $\text{Inf}_i(f) = \sum_{\boldsymbol{\sigma}: \sigma_i \neq \mathbf{0}} \widehat{f}(\boldsymbol{\sigma})^2$ *and hence, for all i, $\text{Inf}_i(f) \leq \text{Var}(f)$*

(ii) $\text{Inf}(f) = \sum_{\boldsymbol{\sigma}} |\boldsymbol{\sigma}| \cdot \widehat{f}(\boldsymbol{\sigma})^2$

(iii) *If $\deg(f) = d$, then $\text{Inf}(f) \leq d \cdot \text{Var}[f]$.*

In the regularity lemma, we deal with restrictions of polynomials. For any subset $H \subseteq [n]$, we will use $\boldsymbol{x}_H$ to denote the tuple of variables in $\boldsymbol{x}$ with indices in H. For any function $P \in L^2(\mathcal{Z}^n, \mu^{\otimes n})$, and any $\xi \in \mathcal{Z}^H$, we will use P^ξ to denote the function obtained by restriction of $\boldsymbol{x}_H$ to ξ, that is, $P^\xi(\boldsymbol{x}_T) = P(\boldsymbol{x}_H \leftarrow \xi, \boldsymbol{x}_T)$ (where $T = [n] \setminus H$); whenever we use such terminology, the subset H will be clear from context. We will use the phrase "ξ fixes H over $\mathcal{A}$" to mean such a restriction. We will use $\{\boldsymbol{\sigma}_H\}$ to denote all degree sequences in $\mathbb{Z}_q^H$, and similarly $\{\boldsymbol{\sigma}_T\}$ to denote all degree sequences in $\mathbb{Z}_q^T$. We use $\boldsymbol{\sigma}_H \circ \boldsymbol{\sigma}_T$ to denote $\boldsymbol{\sigma} \in \mathbb{Z}_q^n$ such that $\sigma_i = (\sigma_H)_i$ if $i \in H$ or $(\sigma_T)_i$ if $i \in T$.

We now state our main Joint Regularity Lemma.

Lemma D.3 (Joint Regularity Lemma). *Let $(\mathcal{Z} \times \mathcal{Z}, \mu)$ be a joint probability space. Let $k, d \in \mathbb{N}$ and $\tau > 0$ be any given constant parameters. There exists $h \stackrel{\text{def}}{=} h((\mathcal{Z} \times \mathcal{Z}, \mu), k, d, \tau)$ such that the following holds:*

For all degree-d polynomials $P : \mathcal{Z}^n \to \mathbb{R}^k$ and $Q : \mathcal{Z}^n \to \mathbb{R}^k$ such that, $\text{Var}[P_j] \leq 1$ and $\text{Var}[Q_j] \leq 1$ for all $j \in [k]$, there exists a subset of indices $H \subseteq [n]$ with $|H| \leq h$ such that with probability at least $(1 - \tau)$ over the assignment $(\xi_A, \xi_B) \sim \mu^{\otimes h}$, the following holds for any $j \in [k]$ (where we denote $T = [n] \setminus H$),

- *the restriction $P_j^{\xi_A}(x_T)$ is such that for all $i \in T$, it holds that $\text{Inf}_i(P_j^{\xi_A}(x_T)) \leq \tau$,*

- *the restriction $Q_j^{\xi_B}(y_T)$ is such that for all $i \in T$, it holds that $\text{Inf}_i(Q_j^{\xi_B}(y_T)) \leq \tau$.*

In particular, one may take $h = \frac{dk^2}{\tau} \cdot \left(\frac{C_4(\alpha)}{\alpha} \log \frac{C_4(\alpha) \cdot k}{\alpha \cdot d \cdot \tau} \right)^{O(d)}$ which is a constant that depends on k, d, τ and $\alpha \stackrel{\text{def}}{=} \alpha(\mu)$, which is the minimum non-zero probability in μ.

The proof of the Joint Regularity lemma follows quite easily by applying the Regularity Lemma for degree-d polynomials (cf. Lemma 5.2 in [GKS16b]).

Lemma D.4 (Regularity Lemma for degree-d functions). *Let $(\mathcal{Z}, \mu_A)$ be a probability space. Let $d \in \mathbb{N}$ and $\tau > 0$ be any given constant parameters. There exists $h \stackrel{\text{def}}{=} h((\mathcal{Z}, \mu_A), d, \tau)$ such that the following holds:*

For all degree-d multilinear polynomials $P \in L^2(\mathcal{Z}^n, \mu_A^{\otimes n})$ with $\text{Var}[P] \leq 1$, there exists a subset of indices $H_0 \subseteq [n]$ with $|H_0| \leq h$, such that for any superset $H \supseteq H_0$, the restrictions of P obtained by evaluating the coordinates in H according to distribution μ_A, satisfies the following (where we denote $T = [n] \setminus H$):

$$\Pr_{\xi \sim \mu_A^{\otimes |H|}} \left[\forall i \in T : \text{Inf}_i(P^\xi(x_T)) \leq \tau \right] \geq 1 - \tau$$

In other words, with probability at least $1 - \tau$ over the random restriction $\xi \sim \mu_A^{\otimes |H|}$, the restricted function $P^\xi(x_T)$ is such that $\text{Inf}_i(P^\xi(x_T)) \leq \tau$ for all $i \in T$.

In particular, one may take $h = \frac{d}{\tau} \cdot \left(\frac{C_4(\alpha)}{\alpha} \log \frac{C_4(\alpha)}{\alpha \cdot d \cdot \tau} \right)^{O(d)}$ which is a constant that depends on d, τ and $\alpha \stackrel{\text{def}}{=} \alpha(\mu_A)$.

Lemma D.3 follows quite easily from the above lemma.

Proof of Lemma D.3. We invoke the Regularity Lemma for individual degree-d polynomials Lemma D.4, namely P_j's and Q_j's and then applying a union bound.

In particular, given our parameters d and τ, we invoke Lemma D.4 with parameters d and $\tau/2k$ for each P_j and Q_j. Suppose we get the set $H_A^{(j)}$ (resp. $H_B^{(j)}$) when applying the regularity lemma on P_j (resp. Q_j). We let $H = \bigcup_{j \in [k]} \left(H_A^{(j)} \cup H_B^{(j)} \right)$. Note that, $|H| \leq 2k \cdot \frac{2dk}{\tau} \cdot \left(\frac{C_4(\alpha)}{\alpha} \log \frac{C_4(\alpha) \cdot k}{\alpha \cdot d \cdot \tau} \right)^{O(d)}$.

Lemma D.4 gives us that for this H, it holds for any P_j and Q_j that,

$$\Pr_{\xi_A \sim \mu_A^{\otimes |H|}} \left[\forall i \in T : \text{Inf}_i \left(P_j^{\xi_A}(x_T) \right) \leq \tau/2k \right] \geq 1 - (\tau/2k)$$

$$\Pr_{\xi_B \sim \mu_B^{\otimes |H|}} \left[\forall i \in T : \text{Inf}_i \left(Q_j^{\xi_B}(y_T) \right) \leq \tau/2k \right] \geq 1 - (\tau/2k)$$

Taking a union bound over $2k$ such statements in total, we get the conclusion we desire.

$\square$

E Invariance Principle

In this section, we prove an invariance principle statement tailor-made for our application, that follows readily as a special case of known invariance principle statements [MOO05, Mos10, IM12]. In particular, we desire a statement as follows.

Lemma E.1. *Let $(\mathcal{Z} \times \mathcal{Z}, \mu)$ be a finite joint probability space, such that $|\mathcal{Z}| = q$ and $\alpha := \alpha(\mu) > 0$ is the minimum probability of any atom in μ. Given parameters $k, d \in \mathbb{N}$ and $\delta > 0$, there exists $\tau = \tau((\mathcal{Z} \times \mathcal{Z}, \mu), k, d, \delta)$ such that the following holds:*
Let $A : \mathcal{Z}^n \to \mathbb{R}^k$ and $B : \mathcal{Z}^n \to \mathbb{R}^k$ be degree-d multilinear polynomials, such that, $\mathrm{Var}(A_j), \mathrm{Var}(B_j) \leq 1$ and $\mathrm{Inf}_\ell(A_j), \mathrm{Inf}_\ell(B_j) \leq \tau$ for all $\ell \in [n]$ and $j \in [k]$. Then, there exist degree-d multilinear polynomials $\widetilde{A} : \mathbb{R}^{n \cdot (q-1)} \to \mathbb{R}^k$ and $\widetilde{B} : \mathbb{R}^{n \cdot (q-1)} \to \mathbb{R}^k$, such that, for all $i, j \in [k]$, it holds that,

$$\left| \left\langle \mathcal{R}_i(\widetilde{A}), \mathcal{R}_j(\widetilde{B}) \right\rangle_{\mathcal{G}_\rho^{\otimes n(q-1)}} - \left\langle \mathcal{R}_i(A), \mathcal{R}_j(B) \right\rangle_{\mu^{\otimes n}} \right| \leq \delta , \tag{36}$$

where $\rho = \rho(\mathcal{Z}, \mathcal{Z}; \mu)$ is the maximal correlation of μ. In particular, one may take $\tau = O\left(\frac{\delta^{1.5} \cdot \alpha^{d/2}}{2^{O(d)} \cdot 2^{O(k)}} \right)$.

Additionally, the theorem also works in reverse, namely, given degree-d multilinear polynomials $\widetilde{A} : \mathbb{R}^n \to \mathbb{R}^k$ and $\widetilde{B} : \mathbb{R}^n \to \mathbb{R}^k$, such that $\mathrm{Var}(\widetilde{A}_j), \mathrm{Var}(\widetilde{B}_j) \leq 1$ and $\mathrm{Inf}_\ell(\widetilde{A}_j), \mathrm{Inf}_\ell(\widetilde{B}_j) \leq \tau$ for all $\ell \in [n]$ and $j \in [k]$, there exist $A : \mathcal{Z}^n \to \mathbb{R}^k$ and $B : \mathcal{Z}^n \to \mathbb{R}^k$, such that Equation (36) holds.

The proof is pretty standard, nevertheless we provide a proof for completeness. We use the vector-valued invariance principle from [IM12], which builds on [MOO05, Mos10]. We state a version that is more tailored to our application, modified from [IM12, Theorem 3.4].

Lemma E.2 (Invariance Principle (cf. [IM12])). *Let $(\Omega^n, \mu^{\otimes n})$ be a finite probability space, such that $\alpha > 0$ is the minimum probability of any atom in μ. Let $\mathcal{F} = (\mathcal{F}_1, \ldots, \mathcal{F}_n)$ be an independent sequence of orthonormal ensembles such that $\mathcal{F}_\ell$ is a basis for functions $\Omega \to \mathbb{R}$, for any $\ell \in [n]$. Let P be a K-dimensional multilinear polynomial such that for every $j \in [K]$, it holds that $\mathrm{Var}(P_j) \leq 1$, $\deg(P_j) \leq d$ and $\mathrm{Inf}_i(P_j) \leq \tau$ for every $i \in [n]$. Finally, let $\Psi : \mathbb{R}^K \to \mathbb{R}$ be Lipschitz continuous with Lipschitz constant L. Then,*

$$\left| \mathbb{E}_\mathcal{F} \Psi(P(\mathcal{F})) - \mathbb{E}_\mathcal{G} \Psi(P(\mathcal{G})) \right| \leq D_K \cdot L \cdot \left(d(8/\sqrt{\alpha})^d \sqrt{\tau} \right)^{1/3} ,$$

where $\mathcal{G}$ is an independent sequence of Gaussian ensembles with same covariance structure as $\mathcal{F}$ and $D_K = 2^{O(K)}$.[10]

Proof of Lemma E.1. We will apply Lemma E.2 on the space $(\Omega^n, \mu^{\otimes n}) = (\mathcal{Z}^n \times \mathcal{Z}^n, \mu^{\otimes n})$. We consider the independent sequence of orthonormal ensembles $\mathcal{F}$ given by $\mathcal{F}_\ell = \left\{ \mathcal{X}_i^{(\ell)}(x) \mathcal{Y}_j^{(\ell)}(y) : i, j \in \{0, \ldots, q-1\} \right\}$ for any $\ell \in [n]$ (where, recall that $|\mathcal{Z}| = q$). Although as defined $\mathcal{F}_\ell$ has q^2 elements, the polynomials we consider will only depend on the subset of characters $\left\{ \mathcal{X}_1^{(\ell)}, \ldots, \mathcal{X}_{q-1}^{(\ell)}, \mathcal{Y}_1^{(\ell)}, \ldots, \mathcal{Y}_{q-1}^{(\ell)} : \ell \in [n] \right\}$. Also, observe that we can choose the characters $\left\{ \mathcal{X}_i^{(\ell)} \right\}$ and $\left\{ \mathcal{Y}_i^{(\ell)} \right\}$ such that $\left\langle \mathcal{X}_i^{(\ell)}, \mathcal{Y}_j^{(\ell)} \right\rangle_\mu = \rho_i \cdot \mathbf{1}_{\{i=j\}}$, where $\rho_1 = \rho(\mathcal{Z}, \mathcal{Z}; \mu)$ is the maximal correlation, and $\rho_i \leq \rho$ for all $1 \leq i \leq q-1$.

Given degree-d multilinear polynomials $A : \mathcal{Z}^n \to \mathbb{R}^k$ and $B : \mathcal{Z}^n \to \mathbb{R}^k$, we consider the polynomial $P : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}^{2k}$ given by $P = (P_1, \ldots, P_{2k})$, where for $j \leq k$, we take $P_j(x, y) = A_j(x)$ and for $j > k$, we take $P_j(x, y) = B_{j-k}(y)$. In other words, P is a *concatenation* of A and B. It is clear that P is also a degree-d multilinear polynomial in $\mathcal{F}$. Also, since A and B are such that $\mathrm{Var}(A_j), \mathrm{Var}(B_j) \leq 1$ and $\mathrm{Inf}_\ell(A_j), \mathrm{Inf}_\ell(B_j) \leq \tau$ for all $\ell \in [n]$ and $j \in [k]$, we have that for all $j \in [2k]$, it holds that, $\mathrm{Var}(P_j) \leq 1$ and $\mathrm{Inf}_\ell(P_j) \leq \tau$. Thus, P satisfies all the conditions needed to prove Lemma E.2.

---

[10]Note that the lemma stated in [IM12] does not state the explicit bound of $D_K = 2^{O(K)}$. However, it is possible to infer this bound from their proof.

We will use the test function $\Psi : \mathbb{R}^{2k} \to \mathbb{R}$, given by $\Psi(\boldsymbol{u}, \boldsymbol{v}) := \mathcal{R}_i(\boldsymbol{u}) \cdot \mathcal{R}_j(\boldsymbol{v})$, for any given $i, j \in [k]$. It is easy to show that $\mathcal{R}(\cdot)$ is a contraction map (since it is rounding to Δ_k which is a convex body), and hence any coordinate $\mathcal{R}_i(\cdot)$ has Lipschitz constant of at most 1. This gives us that our test function $\Psi(\boldsymbol{u}, \boldsymbol{v}) = \mathcal{R}_i(\boldsymbol{u}) \cdot \mathcal{R}_j(\boldsymbol{v})$ also has a Lipschitz constant of at most 1.

We can interpret the invariance principle, as substituting $\left\{ \mathcal{X}_1^{(\ell)}, \ldots, \mathcal{X}_{q-1}^{(\ell)}, \mathcal{Y}_1^{(\ell)}, \ldots, \mathcal{Y}_{q-1}^{(\ell)} : \ell \in [n] \right\}$ by correlated multivariate Gaussians $(\mathbf{g}^{(\ell)}, \mathbf{h}^{(\ell)}) = (g_1^{(\ell)}, \cdots, g_{q-1}^{(\ell)}, h_1^{(\ell)}, \ldots, h_{q-1}^{(\ell)})$, such that $\mathbb{E} \, g_i^{(\ell)} h_j^{(\ell)} = \rho_i \cdot \mathbf{1}_{\{i=j\}}$ and $\mathbb{E} \, g_i^{(\ell)} g_j^{(\ell)} = \mathbb{E} \, h_i^{(\ell)} h_j^{(\ell)} = \mathbf{1}_{\{i=j\}}$. Thus the invariance principle is taking $P : \mathcal{Z}^n \times \mathcal{Z}^n \to \mathbb{R}^{2k}$ and producing $P' : \mathbb{R}^{(q-1)n} \times \mathbb{R}^{(q-1)n} \to \mathbb{R}^{2k}$. Note that the first k coordinates of P are polynomials over the subset $\left\{ \mathcal{X}_1^{(\ell)}, \ldots, \mathcal{X}_{q-1}^{(\ell)} : \ell \in [n] \right\}$ and the latter k coordinates are polynomials over $\left\{ \mathcal{Y}_1^{(\ell)}, \ldots, \mathcal{Y}_{q-1}^{(\ell)} : \ell \in [n] \right\}$. Hence we can interpret the first k coordinates as $A' : \mathbb{R}^{(q-1)n} \to \mathbb{R}^k$ and latter k coordinates as $B' : \mathbb{R}^{(q-1)n} \to \mathbb{R}^k$, and Lemma E.2 gives us that,

$$\left| \langle \mathcal{R}_i(A'), \mathcal{R}_j(B') \rangle_{\mathcal{G}^{\otimes n}} - \langle \mathcal{R}_i(A), \mathcal{R}_j(B) \rangle_{\mu^{\otimes n}} \right| \;\leq\; 2^{O(k)} \cdot \left(d(8/\sqrt{\alpha})^d \sqrt{\tau} \right)^{1/3} \;\leq\; \delta$$

for a choice of $\tau \leq O\left(\frac{\delta^{1.5} \cdot \alpha^{d/2}}{2^{O(d)} \cdot 2^{O(k)}} \right)$.

We are still not done though! We want $\widetilde{A}$ and $\widetilde{B}$ which act on coorrelated inputs sampled from $\mathcal{G}_\rho^{\otimes (q-1)n}$, that is, $(\mathbf{g}^{(\ell)}, \mathbf{h}^{(\ell)}) = (g_1^{(\ell)}, \cdots, g_{q-1}^{(\ell)}, h_1^{(\ell)}, \ldots, h_{q-1}^{(\ell)})$, such that $\mathbb{E} \, g_i^{(\ell)} h_j^{(\ell)} = \rho \cdot \mathbf{1}_{\{i=j\}}$ and $\mathbb{E} \, g_i^{(\ell)} g_j^{(\ell)} = \mathbb{E} \, h_i^{(\ell)} h_j^{(\ell)} = \mathbf{1}_{\{i=j\}}$. However the correlation pattern obtained in $\mathcal{G}$ is not exactly this. But note that each $\rho_i \leq \rho$. Thus, given $(g_i^{(\ell)}, h_i^{(\ell)})$ with correlation ρ, we can simply apply $U_{\rho_i/\rho}$ operator on $h_i^{(\ell)}$ to bring down the correlation from ρ to ρ_i. Applying this appropriate operation for every $\ell \in [n]$ and $i \in [q]$, we get our desired $\widetilde{A} : \mathbb{R}^{n \cdot (q-1)} \to \mathbb{R}^k$ and $\widetilde{B} : \mathbb{R}^{n \cdot (q-1)} \to \mathbb{R}^k$.

The reverse part of the theorem follows similarly as well. Here, we get polynomials A and B, that depend only on $\left\{ \mathcal{X}_1^{(\ell)} \right\}$ and $\left\{ \mathcal{Y}_1^{(\ell)} \right\}$ respectively. $\qquad \square$

F Decidability of Non-Interactive Simulation

In this section, we prove Theorem 1.6 showing the decidability of the GAP-NIS problem.

Proof of Theorem 1.6. If we were in the YES case of GAP-NIS$((\mathcal{Z} \times \mathcal{Z}, \mu), V, k, \varepsilon)$, then we have that there exists an N and functions $A : \mathcal{Z}^N \to \Delta_k$ and $B : \mathcal{Z}^N \to \Delta_k$, such that the distribution $\nu' = (A(\boldsymbol{x}), B(\boldsymbol{y}))_{(\boldsymbol{x}, \boldsymbol{y}) \sim \mu^{\otimes N}}$ is such that $d_{\text{TV}}(\nu', \nu) \leq \varepsilon$ for some $\nu \in V$. Using Theorem 7.1, with parameter $\varepsilon/3$, we get that there exists functions $\widetilde{A} : \mathcal{Z}^{n_0} \to \Delta_k$ and $\widetilde{B} : \mathcal{Z}^{n_0} \to \Delta_k$ such that the distribution $\nu'' = (\widetilde{A}(\boldsymbol{x}), \widetilde{B}(\boldsymbol{y}))_{(\boldsymbol{x}, \boldsymbol{y}) \sim \mu^{\otimes n_0}}$ is such that $d_{\text{TV}}(\nu'', \nu') \leq \varepsilon/3$. Hence, $d_{\text{TV}}(\nu'', \nu) \leq 4\varepsilon/3$ for some $\nu \in V$.

In the NO case of GAP-NIS$((\mathcal{Z} \times \mathcal{Z}, \mu), V, k, \varepsilon)$, we have that for all N, in particular for $N = n_0$, and for all functions $A : \mathcal{Z}^{n_0} \to \Delta_k$ and $B : \mathcal{Z}^{n_0} \to \Delta_k$ it holds that the distribution $\nu'' = (A(\boldsymbol{x}), B(\boldsymbol{y}))_{(\boldsymbol{x}, \boldsymbol{y}) \sim \mu^{\otimes n_0}}$ satisfies $d_{\text{TV}}(\nu'', \nu) > 2\varepsilon$ for all $\nu \in V$.

This naturally gives us a brute force algorithm: Analyze all possible functions $\widetilde{A} : \mathcal{Z}^{n_0} \to \Delta_k$ and $\widetilde{B} : \mathcal{Z}^{n_0} \to \Delta_k$ to check if there exist functions $\widetilde{A}$ and $\widetilde{B}$ with distribution $\nu'' = (\widetilde{A}(\boldsymbol{x}), \widetilde{B}(\boldsymbol{y}))_{(\boldsymbol{x}, \boldsymbol{y}) \sim \mu^{\otimes n_0}}$ satisfying $d_{\text{TV}}(\nu'', \nu) \leq 4\varepsilon/3$ for some $\nu \in V$.

For purposes of our algorithm we can replace the range Δ_k by any $(\varepsilon/3)$-cover C, that is, a set of discrete points in Δ_k such that any point in Δ_k is within an ℓ_1 distance of $\varepsilon/3$ from some point in C. Note that we could choose such a C of size at most $(1/\varepsilon)^{\widetilde{O}(k)}$. This ensures that if indeed such a desired $\widetilde{A}$ and $\widetilde{B}$ exist, then we will find functions $\widetilde{A}' : \mathcal{Z}^{n_0} \to C$ and $\widetilde{B}' : \mathcal{Z}^{n_0} \to C$ such that distribution $\nu''' = (\widetilde{A}'(\boldsymbol{x}), \widetilde{B}'(\boldsymbol{y}))_{(\boldsymbol{x}, \boldsymbol{y}) \sim \mu^{\otimes n_0}}$ satisfying $d_{\text{TV}}(\nu''', \nu) \leq 5\varepsilon/3 < 2\varepsilon$ for some $\nu \in V$. In the YES case, we will find such functions, whereas in the NO case, $\widetilde{A}'$ and $\widetilde{B}'$ as above simply don't exist.

The number of pair of functions $(\widetilde{A}, \widetilde{B})$ to brute force over is $|C|^{O(|\mathcal{Z}|^{n_0})}$, which gives us an upper bound on the running time as

$$\exp\exp\exp\left(\mathrm{poly}\left(k, \frac{1}{\varepsilon}, \frac{1}{1-\rho_0}, \log\left(\frac{1}{\alpha}\right)\right)\right) .$$

$\square$