Non interactive simulation of correlated distributions is decidable

Non interactive simulation of correlated distributions is decidable

Anindya De[*] Elchanan Mossel[†] Joe Neeman[‡]

Abstract

A basic problem in information theory is the following: Let $\mathbf{P} = (\mathbf{X}, \mathbf{Y})$ be an arbitrary distribution where the marginals $\mathbf{X}$ and $\mathbf{Y}$ are (potentially) correlated. Let Alice and Bob be two players where Alice gets samples $\{x_i\}_{i\geq 1}$ and Bob gets samples $\{y_i\}_{i\geq 1}$ and for all i, $(x_i, y_i) \sim \mathbf{P}$. What joint distributions $\mathbf{Q}$ can be simulated by Alice and Bob without any interaction?

Classical works in information theory by Gács-Körner and Wyner answer this question when at least one of $\mathbf{P}$ or $\mathbf{Q}$ is the distribution Eq (Eq is defined as uniform over the points $(0,0)$ and $(1,1)$). However, other than this special case, the answer to this question is understood in very few cases. Recently, Ghazi, Kamath and Sudan showed that this problem is decidable for $\mathbf{Q}$ supported on $\{0,1\} \times \{0,1\}$. We extend their result to $\mathbf{Q}$ supported on any finite alphabet. Moreover, we show that If $\mathbf{Q}$ can be simulated, our algorithm also provides a (non-interactive) simulation protocol.

We rely on recent results in Gaussian geometry (by the authors) as well as a new *smoothing argument* inspired by the method of *boosting* from learning theory and potential function arguments from complexity theory and additive combinatorics.

1 Introduction

The starting point of this paper is a rather basic problem in information theory and communication complexity, known as the problem of *non-interactive simulation of joint distributions*: Consider two non-communicating players Alice and Bob. Suppose that we give Alice and Bob the sequences $\{\mathbf{X}_1\}_{i=1}^{\infty}$ and $\{\mathbf{Y}_i\}_{i=1}^{\infty}$ respectively, where the pairs $(\mathbf{X}_i, \mathbf{Y}_i)$ are independently drawn from some joint distribution $\mathbf{P}$. Without communicating with each other, which joint distributions $\mathbf{Q}$ can Alice and Bob jointly simulate? Note that both $\mathbf{P}$ and $\mathbf{Q}$ are fully known to the players and thus, there is no *uncertainty* in the problem definition.

To state the problem more precisely, suppose that $\mathbf{P}$ is a distribution on $\mathcal{Z} \times \mathcal{Z}$ and that $\mathbf{Q}$ is a distribution on $\mathcal{W} \times \mathcal{W}$. A *non-interactive strategy* for Alice and Bob simply denotes a triple (n, f, g) such that $f, g : \mathcal{Z}^n \to \mathcal{W}$, and for which $(f(\mathbf{X}^n), g(\mathbf{Y}^n))$ has distribution $\mathbf{Q}$ whenever $(\mathbf{X}_i, \mathbf{Y}_i)$ are drawn independently from $\mathbf{P}$ (here, $\mathbf{X}^n$ denotes $\mathbf{X}_1, \ldots, \mathbf{X}_n$). The main question that we consider in this paper is how to decide whether a non-interactive strategy exists for a given input distribution $\mathbf{P}$ and a given target distribution $\mathbf{Q}$. Note that not every pair of input and target distributions admits a non-interactive strategy. The most obvious example of this is the case where the two coordinates of $\mathbf{P}$ are independent; in this case, one can obviously only simulate distributions $\mathbf{Q}$ whose coordinates are also independent.

Witsenhausen [Wit75] introduced the problem of non-interactive simulation, and he studied the case where $\mathbf{Q}$ is a Gaussian measure on $\mathbb{R}^2$. In this case, he showed that $\mathbf{Q}$ can be approximately simulated by $\mathbf{P}$ if and only if the absolute value of the correlation between the components of $\mathbf{Q}$ is at most the so-called "maximal correlation coefficient" (which we will define later) of $\mathbf{P}$. In this case, Witsenhausen showed that for any $\delta > 0$, Alice and Bob can simulate $\mathbf{Q}$ up to ℓ_1 error δ with $n = \mathrm{poly}(|\mathbf{P}|, \log(1/\delta))$. Here $|\mathbf{P}|$ is the bit complexity of representing $\mathbf{P}$. Further, he gave an explicit algorithm to compute f and g in time $\mathsf{poly}(n)$.

Remark 1. *Note that as stated here, Witsenhausen's result can only handle finitely supported* $\mathbf{P}$. *However, the result is actually more general and can handle a much larger class of distributions. However, stating the resulting bound on* n *requires more notation. So, for sake of simplicity, we only state the result for the case of finitely supported* $\mathbf{P}$.

Various other questions of this flavor have been explored in information theory. We discuss two examples here. Let us use Eq to denote the distribution uniform on the two points $(0,0)$ and $(1,1)$.

1. In their seminal paper, Gács and Körner [GK73] studied non-interactive simulation in the case $\mathbf{Q} = \mathsf{Eq}$. In this case, they obtained a simple and complete characterization of all $\mathbf{P}$ such that it is possible to non-interactively simulate $\mathbf{Q}$ from $\mathbf{P}$. They also studied the notion of *simulation capacity*: roughly, how many samples from $\mathbf{P}$ are needed to produce each sample from $\mathbf{Q}$? They showed that the simulation capacity is equal to another quantity, which is now known as the *Gács-Körner common information of* $\mathbf{P}$.

[*] Northwestern University

[†] MIT. Supported by ONR grant N00014-16-1-2227 and NSF CCF-1665252 and DMS-1737944

[‡] UT Austin

2. Around the same time, Wyner [Wyn75] considered the complementary problem where $\mathbf{P} = \mathsf{Eq}$ and $\mathbf{Q}$ is arbitrary. In other words, Alice and Bob have access to shared random bits and they want to simulate $\mathbf{Q}$. In this case it is always possible to approximately simulate $\mathbf{Q}$; Wyner studied the simulation capacity, and showed that it is equal to what is now known as the *Wyner common information of* $\mathbf{Q}$.

Outside of these cases, the problem becomes significantly more complicated (see, for example, [KA16] and the references therein). Before we study the general problem in more detail, let us briefly explain the motivation for studying the problem of non-interactive simulation.

Motivation While the problem of non-interactive simulation has been studied in information theory for a long time, the interest in computer science is somewhat more recent. Probably the earliest work in this direction in computer science are the works of [MO05, MOR$^+$06] who studied the problem of non-interactive correlation distillation (which is a special case of the problem of non-interactive simulation). More recently, several papers [BGI14, CGMS15, GKS16a] re-examined a very basic assumption in communication complexity and more broadly in distributed computing: most models in these areas assume that there is a source of common randomness available to the different parties. How does the quality and nature of randomness affect the difficulty of accomplishing a computational task? Questions of similar flavor have also been asked in other "distributed settings" such as cryptography [AC93, AC98, CN00]. In this context, very recently, Ghazi, Kamath and Sudan [GKS16b] raised the general question of non-interactive simulation. Note that if two parties share a source of randomness distributed as $\mathbf{P}$ and they can non-interactively simulate $\mathbf{Q}$, then it means that any computational task which can be simulated with $\mathbf{Q}$ can also be simulated with $\mathbf{P}$ without any further overhead in communication. We discuss the work of [GKS16b] in more detail later on.

Another motivation for studying this problem (which was also raised in [GKS16b]) is its connection to the so-called "tensor power" problems, which involve the limiting behavior of certain quantities as the underlying dimensionality tends to infinity. To be less abstract, we will mention two examples from [GKS16b]: the Shannon capacity of a graph, and the MIP^* class.

For the first example, let G be a graph and let $G^{\otimes n}$ be the n^{th} tensor power of G and I_n be the size of the largest independent set in $G^{\otimes n}$. The Shannon capacity of G, denoted by $\Theta(G)$ is defined to be $\lim_{n\to\infty} I_n^{1/n}$. How fast does $I_n^{1/n}$ converge to $\Theta(G)$? Is the quantity $\Theta(G)$ (approximately) computable? Despite having received significant attention [Lov79, AL06, Hae79] in literature, both these questions remain open.

In the second example, MIP^* is defined as the class of multiprover interactive games with (arbitrary) quantum entanglement. How does the value of the game vary as the entanglement increases? In particular, obtaining explicit convergence results would imply that this class is decidable [DLTW08, Reg16, KKM$^+$11], an outstanding problem in the theory of multiprover games.

Non-interactive simulation also fits the "tensor power" motif. Namely, simulating a sample from $\mathbf{Q}$ potentially requires multiple copies of $\mathbf{P}$ and an explicit upper bound on the number of copies of $\mathbf{P}$ required to simulate $\mathbf{Q}$ automatically implies that one can find a simulation protocol (provided one exists) by exhaustively trying out all possible simulation strategies (which is finite, if $\mathbf{P}$ and $\mathbf{Q}$ are supported on finite number of elements). Indeed, our main theorem (Theorem 2) follows by proving such an upper bound on the number of copies of $\mathbf{P}$ required for non-interactive simulation (Theorem 3). We hope that the techniques introduced in this paper will be of use in some of these other problems as well.

Connection to Gaussian noise stability Having explained the motivation, we now again turn to the problem of non-interactive simulation. To understand the problem outside of the cases where at least one of $\mathbf{P}$ or $\mathbf{Q}$ is Eq, we look at the following important example. Let $\mathbf{G}_{\rho,2}$ be the centered Gaussian measure on $\mathbb{R}^2$, where each coordinate has unit variance and the correlation between the coordinates is $\rho > 0$. Consider the setting where $\mathbf{P} = \mathbf{G}_{\rho,2}$. For $\kappa > 0$, let us define Eq_κ to be the distribution on $\{0,1\} \times \{0,1\}$ where each marginal is unbiased and their correlation is κ (i.e. $\mathsf{Eq} = \mathsf{Eq}_1$). For what values of κ can Alice and Bob non-interactive simulate Eq_κ? The influential result of Borell [Bor85] on Gaussian noise stability can be interpreted as saying that Alice and Bob can simulate Eq_κ precisely when $\kappa \leq \frac{2}{\pi}\sin^{-1}(\rho)$. In fact, the simulation strategy is also quite simple: Let δ be such that $(1-\delta)^2 \cdot \frac{2}{\pi}\sin^{-1}(\rho) = \kappa$. Let $(\mathbf{X}_1, \mathbf{Y}_1) \sim \mathbf{P}$. To non-interactive simulate Eq_κ, Alice (resp. Bob) outputs the sign of $\mathbf{X}_1$ (resp. $\mathbf{Y}_1$) with probability $1-\delta$ and a random unbiased bit with probability δ. Note that the private randomness involved in this protocol can be simulated by Alice and Bob non-interactively: Alice can use a threshold on $\mathbf{X}_2$ to make her choice, and then output $\mathrm{sign}(\mathbf{X}_3)$ in the event that she decides to output a random unbiased bit; Bob can use $\mathbf{Y}_4$ and $\mathbf{Y}_5$ in the same way.

What happens if we consider the ternary analogue of Eq_κ? Namely, what is the maximum $\kappa > 0$ such that Alice and Bob can simulate a distribution on $\{0,1,2\} \times \{0,1,2\}$ where their individual marginals are uniform on $\{0,1,2\}$ and their probability of agreement is κ? To begin with, the analogue of Borell's result in this setting is not known: we do not know the optimal strategy for Alice and Bob in order to maximize the probability of their agreement. This issue was partially addressed in a recent work of the authors [DMN17]: while [DMN17] does not solve the complete simulation problem, [DMN17] shows that they can approximately compute a strategy that maximizes the agreement probability, to an arbitrarily small error.

Note that the above result is not sufficient to settle the simulation problem for arbitrary $\mathbf{Q}$ over ternary and larger alphabet even for $\mathbf{P} = \mathbf{G}_{\rho,2}$ (unlike the case of the binary alphabet). The reason is as follows: any distribution over $\{0,1\} \times \{0,1\}$ is completely specified just by the individual marginals and probability of agreement of the two coordinates. However, this is not the case for ternary or larger alphabets. A simple example is

$$\mathbf{D}_1 = \text{Uniform on}\{(0,0),(0,1),(1,1),(1,2),(2,0),(2,2)\}.$$
$$\mathbf{D}_2 = \text{Uniform on}\{(0,0),(0,2),(1,1),(1,0),(2,0),(2,1)\}.$$

Note that $\mathbf{D}_1$ and $\mathbf{D}_2$ are distinct distributions which both satisfy:

(i) The individual marginals are uniform over $\{0,1,2\}$.

(ii) The two coordinates agree with probability $1/2$.

In this work, we extend to framework of [DMN17] to answer the non-interactive simulation problem for arbitrary $\mathbf{P}$ and $\mathbf{Q}$. Specifically, we show that if $\mathbf{Q}$ can be non-interactively simulated from $\mathbf{P}$ then one can compute, for every $\delta > 0$, a δ-approximate simulation protocol. By a δ-approximate simulation protocol, we mean the output distribution is δ-close to $\mathbf{Q}$ in total variation distance. Here is an equivalent formulation, in which $|\mathbf{P}|$ denotes the bit complexity of representing $\mathbf{P}$.

Theorem 2. *Let $(\mathcal{Z} \times \mathcal{Z}, \mathbf{P})$ and $([k] \times [k], \mathbf{Q})$ be probability spaces, and let $\mathbf{X}^n = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and $\mathbf{Y}^n = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$, where $(\mathbf{X}_i, \mathbf{Y}_i)$ are independently drawn from $\mathbf{P}$. For every $\delta > 0$, there is an algorithm running in time $O_{|\mathbf{P}|,\delta}(1)$ which distinguishes between the following two cases:*

1. *There exist $n \in \mathbb{N}$ and $f, g\colon \mathcal{Z}^n \to [k]$ such that $\mathbf{Q}$ and the distribution of $(f(\mathbf{X}^n), g(\mathbf{Y}^n))$ are δ-close in total variation distance. In this case, there is an explicit $n_0 = n_0(|\mathbf{P}|, \delta)$ such that we may choose $n \le n_0$. Further, the functions f and g can be explicitly computed.*

2. *For any $n \in \mathbb{N}$ and $f, g\colon \mathcal{Z}^n \to [k]$, $\mathbf{Q}$ and the distribution of $(f(\mathbf{X}^n), g(\mathbf{Y}^n))$ are 3δ-far in total variation distance.*

We remark here that the bound n_0, while computable, is not primitive recursive and has an Ackermann type growth, which is introduced by our application of a regularity lemma from [DS14]. It is easy to see that to prove Theorem 2, it suffices to prove the following theorem.

Theorem 3. *With the notation of Theorem 2, suppose there exist $f, g\colon \mathcal{Z}^n \to [k]$ such that $(f(\mathbf{X}^n), g(\mathbf{Y}^n)) \sim \mathbf{Q}$. Then, there exist $n_0 = n_0(|\mathbf{P}|, \delta)$ and $f_\delta, g_\delta\colon \mathcal{Z}^{n_0} \to [k]$ such that $\mathbf{Q}$ and the distribution of $(f_\delta(\mathbf{X}^{n_0}), g_\delta(\mathbf{Y}^{n_0}))$ are δ-close in total variation distance. Moreover, n_0 is computable. Further, the functions f_δ and g_δ can be explicitly computed.*

The gist of the above theorem is that if a distribution can be simulated then it can be approximately simulated with a bounded number of samples. (The crucial point in the previous sentence is that the bound is explicit, and that it depends only on $\mathbf{P}$ and the desired accuracy.) To obtain Theorem 2 from Theorem 3, we exhaustively try out all possible pairs of functions $f, g : \mathcal{Z}^{n_0} \to [k] \times [k]$ for $n_0 = n_0(|\mathbf{P}|, \delta)$. Note that this step is meaningful because $\mathcal{Z}$ is finite. If we are in case (i), then we can indeed find f, g such that $f(\mathbf{X}^{n_0}, \mathbf{Y}^{n_0})$ is 2δ-close to $\mathbf{Q}$. On the other hand, if we are in case (ii), then by definition, for every such f and g, $f(\mathbf{X}^{n_0}, \mathbf{Y}^{n_0})$ is at least 3δ far from $\mathbf{Q}$.

Remark 4. *While Theorem 2 shows the decidability of the gapped version of non-interactive simulation, the decidability of the exact version remains open (and does not seem to be amenable to our techniques). More precisely, given $\mathbf{P}$ and $\mathbf{Q}$ we do not know if the problem of checking whether $\mathbf{Q}$ can be (exactly) non-interactively simulated from $\mathbf{P}$ is decidable.*

1.1 Recent work, and the difficulty of going from two to three Ghazi, Kamath, and Sudan [GKS16b] proved Theorems 2 and 3 in the case $k = 2$. Moreover, they gave an explicit doubly exponential bound on n_0 and the running time of the algorithm. Borell's noise stability theorem (which is not available for $k > 2$) played an important role in their analysis. To explain the bottleneck in extending their result for any k, we will elaborate on the case where $\mathcal{Z} = \mathbb{R}$ and $\mathbf{P} = \mathbf{G}_{\rho,2}$. We begin by recalling Borell's inequality [Bor85] on Gaussian noise stability. A convention that we will adopt for the rest of the paper is that unless explicitly mentioned otherwise, the expectation is always w. r. t. the variable being a standard Gaussian where the ambient dimension will be clear from the context.

Theorem 5. *[Bor85] Let* $\mathbf{P} = \mathbf{G}_{\rho,2}$*. For any* $\mu_1, \mu_2 \in [0,1]$*, let* $f, g\colon \mathbb{R}^n \to \{0,1\}$ *such that* $\mathbf{E}[f] = \mu_1$ *and* $\mathbf{E}[g] = \mu_2$*. Let us choose* κ_1, κ_2 *such that for* $f_{\mathsf{LTF}}, g_{\mathsf{LTF}}\colon \mathbb{R} \to \{0,1\}$ *defined as* $f_{\mathsf{LTF}}(x) = \mathsf{sign}(x - \kappa_1)$ *and* $g_{\mathsf{LTF}}(x) = \mathsf{sign}(x - \kappa_2)$*, we have* $\mathbf{E}[f_{\mathsf{LTF}}] = \mu_1$ *and* $\mathbf{E}[g_{\mathsf{LTF}}] = \mu_2$*. Then, for* $(\mathbf{X}, \mathbf{Y}) \sim \mathbf{P}$*,*

$$\Pr[f_{\mathsf{LTF}}(\mathbf{X}) = g_{\mathsf{LTF}}(\mathbf{Y})] \geq \Pr[f(\mathbf{X}^n) = g(\mathbf{Y}^n)].$$

Likewise, if we define $h_{\mathsf{LTF}} = \mathsf{sign}(-x - \kappa_2)$*, then* $\mathbf{E}[h_{\mathsf{LTF}}] = \mu_2$ *and*

$$\Pr[f_{\mathsf{LTF}}(\mathbf{X}) = h_{\mathsf{LTF}}(\mathbf{Y})] \leq \Pr[f(\mathbf{X}^n) = g(\mathbf{Y}^n)].$$

To explain the intuitive meaning of these theorems, let us define $\mathsf{Corr}_{\max}(\rho, \mu_1, \mu_2)$ and $\mathsf{Corr}_{\min}(\rho, \mu_1, \mu_2)$ as

$$\mathsf{Corr}_{\max}(\rho, \mu_1, \mu_2) = \Pr_{(\mathbf{X},\mathbf{Y})\sim\mathbf{P}}[f_{\mathsf{LTF}}(\mathbf{X}) = g_{\mathsf{LTF}}(\mathbf{Y})],$$

$$\mathsf{Corr}_{\min}(\rho, \mu_1, \mu_2) = \Pr_{(\mathbf{X},\mathbf{Y})\sim\mathbf{P}}[f_{\mathsf{LTF}}(\mathbf{X}) = h_{\mathsf{LTF}}(\mathbf{Y})]$$

where f_{LTF}, g_{LTF} and h_{LTF} are halfspaces defined in Theorem 5. Then, Borell's result implies that for any given measures μ_1, μ_2 and functions f, g with these measures, the probability that $f(\mathbf{X})$ and $g(\mathbf{Y})$ are identical lies between $\mathsf{Corr}_{\max}(\rho, \mu_1, \mu_2)$ and $\mathsf{Corr}_{\min}(\rho, \mu_1, \mu_2)$. Further, now, it easily follows that for any η such that $\mathsf{Corr}_{\min}(\rho, \mu_1, \mu_2) \leq \eta \leq \mathsf{Corr}_{\max}(\rho, \mu_1, \mu_2)$, there is a function $g_\eta\colon \mathbb{R} \to \{0,1\}$ such that $\mathbf{E}[g_\eta] = \mu_2$ and $\eta = \Pr_{(\mathbf{X},\mathbf{Y})\sim P}[f(\mathbf{X}) = g_\eta(\mathbf{Y})]$. In fact, it is also easy to see that g_η can be assumed to be the indicator function of an interval.

Now, consider any distribution $\mathbf{Q}$ on $\{0,1\} \times \{0,1\}$, and take $(\mathbf{U}, \mathbf{V}) \sim \mathbf{Q}$. Assume that there exist $f, g\colon \mathbb{R}^n \to \{0,1\}$ such that $(f(\mathbf{X}^n), g(\mathbf{Y}^n)) \sim \mathbf{Q}$. Defining $\mu_{1,\mathbf{Q}} = \mathbf{E}[\mathbf{U}]$, $\mu_{2,\mathbf{Q}} = \mathbf{E}[\mathbf{V}]$ and $\eta_{\mathbf{Q}} = \Pr[\mathbf{U} = \mathbf{V}]$ and applying Theorem 5, we obtain that there are functions $f_{\mathbf{Q}}, g_{\mathbf{Q}}\colon \mathbb{R} \to \{0,1\}$ which satisfy

$$\mathbf{E}[f_{\mathbf{Q}}(\mathbf{X})] = \mu_{1,\mathbf{Q}}, \quad \mathbf{E}[g_{\mathbf{Q}}(\mathbf{Y})] = \mu_{2,\mathbf{Q}},$$

and

$$\Pr_{(\mathbf{X},\mathbf{Y})\sim\mathbf{P}}[f_{\mathbf{Q}}(\mathbf{X}) = g_{\mathbf{Q}}(\mathbf{Y})] = \eta_{\mathbf{Q}}.$$

Further, the functions $f_{\mathbf{Q}}$ and $g_{\mathbf{Q}}$ are in fact indicators of intervals and given $\mu_{1,\mathbf{Q}}$, $\mu_{2,\mathbf{Q}}$ and $\eta_{\mathbf{Q}}$, the functions $f_{\mathbf{Q}}$ and $g_{\mathbf{Q}}$ can be explicitly computed. Observe that any distribution $\mathbf{Q}$ over $\{0,1\} \times \{0,1\}$ is characterized by the quantities $\mu_{1,\mathbf{Q}}$, $\mu_{2,\mathbf{Q}}$ and $\eta_{\mathbf{Q}}$. Thus, it implies that $(f_{\mathbf{Q}}(\mathbf{X}), g_{\mathbf{Q}}(\mathbf{Y})) \sim \mathbf{Q}$. This completely settles the non-interactive simulation problem in the case $k = 2$, when $\mathbf{P}$ is the Gaussian measure $\mathbf{G}_{\rho,2}$ on $\mathbb{R}^2$.

In particular, we see that when $\mathbf{P}$ is Gaussian, the result of [GKS16b] is a straightforward consequence of Theorem 5. Indeed, their main contribution was to show that the general case reduces to the Gaussian case. Moreover, that part of their argument turns out to generalize to $k > 2$ (as we will discuss later). Therefore, let us continue examining the case where $\mathbf{P}$ is Gaussian, and see why $k > 2$ is more difficult. There are two problems:

1. The analogue of Borell's result for $k > 2$ is not known. In particular, the following simple question is still open: let $\boldsymbol{\mu} \in \Delta_k$ where Δ_k is the convex hull of the standard unit vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$. Let $A_{\boldsymbol{\mu}} = \{f\colon \mathbb{R}^n \to [k]\colon \mathbf{E}[f] = \boldsymbol{\mu}\}$. Among all $f \in A_{\boldsymbol{\mu}}$, what f maximizes the probability $\Pr_{(\mathbf{X},\mathbf{Y})\sim\mathbf{P}}[f(\mathbf{X}) = f(\mathbf{Y})]$? If $k = 2$, then Theorem 5 asserts that f is the indicator of some halfspace; for $k > 3$, the answer is essentially unknown. Of particular relevance to us, it is not even known whether the optimal value can be achieved in any finite dimension (whereas in the case $k = 2$, it is achieved in one dimension).

2. For $k = 2$, any distribution $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2)$ supported on $[k] \times [k]$ is completely defined by $\mathbf{E}[\mathbf{R}_1]$, $\mathbf{E}[\mathbf{R}_2]$ and $\Pr[\mathbf{R}_1 = \mathbf{R}_2]$. However, this is no longer true when $k > 2$.

In [DMN17], the authors partially circumvented the first issue. To explain the result of [DMN17], we will need to introduce two notions. The first is that of the (standard) Ornstein-Uhlenbeck noise operator. Namely, for any $t \geq 0$ and $f\colon \mathbb{R}^n \to \mathbb{R}$, we define $P_t f\colon \mathbb{R}^n \to \mathbb{R}$ as

$$P_t f(x) = \mathop{\mathbf{E}}_{y\sim\gamma_n}[f(e^{-t}x + \sqrt{1 - e^{-2t}}y)]. \tag{1.1}$$

To see the connection between P_t and our ρ-correlated Gaussian distribution $\mathbf{P} = \mathbf{G}_{\rho,2}$, choose t so that $e^{-t} = \rho$. Then

$$\mathbf{E}_{(\mathbf{X},\mathbf{Y})^n\sim\mathbf{P}^n}[f(\mathbf{X}^n)\cdot f(\mathbf{Y}^n)] = \mathbf{E}_{\mathbf{X}^n\sim\gamma_n}[f(\mathbf{X}^n)\cdot P_t f(\mathbf{X}^n)].$$

The above quantity is often referred to as the noise stability of f at noise rate $t > 0$. Note that the operator P_t is a linear operator on the space of functions mapping $\mathbb{R}^n$ to $\mathbb{R}$. In fact, the noise operator can be syntactically extended to functions $f\colon \mathbb{R}^n \to \mathbb{R}^k$ with the same definition as in (1.1). Embedding Δ_k in $\mathbb{R}^k$ and identifying $[k]$ with the vertices of Δ_k, we obtain that

$$\mathbf{E}[\langle f(\mathbf{X}^n), f(\mathbf{Y}^n)\rangle] = \mathbf{E}[\langle f(\mathbf{X}^n), P_t f(\mathbf{X}^n)\rangle].$$

Let us now recall the notion of a multivariate polynomial threshold function (PTF) from [DMN17]. Given polynomials, $p_1, \dots, p_k : \mathbb{R}^n \to \mathbb{R}$, define $f = \mathsf{PTF}(p_1, \dots, p_k)$ as

$$f(x) = \begin{cases} j & \text{if } p_j(x) > 0 \text{ and } p_i(x) \le 0 \text{ for all } j \neq i, \\ 1 & \text{otherwise.} \end{cases}$$

In [DMN17], the authors proved the following theorem.

Theorem 6. *Let $f : \mathbb{R}^n \to [k]$ such that $\mathbf{E}[f] = \boldsymbol{\mu} \in \mathbb{R}^k$. Then, given any $t > 0, \epsilon > 0$, there exists an explicitly computable $n_0 = n_0(t, k, \epsilon)$ and $d = d(t, k, \epsilon)$ such that there is a degree-d PTF $g : \mathbb{R}^{n_0} \to [k]$ with*

1. $\|\mathbf{E}[g] - \boldsymbol{\mu}\|_1 \le \epsilon$.
2. $\mathbf{E}[\langle g, P_t g\rangle] \ge \mathbf{E}[\langle f, P_t f\rangle] - \epsilon$.

In other words, Theorem 6 shows that for any given $\boldsymbol{\mu}$ and error parameter $\epsilon > 0$, there is a low-degree, low-dimensional PTF g which approximately maximizes the noise stability and whose expectation is close to $\boldsymbol{\mu}$. We remark here that the issue of matching the expectation exactly versus approximately is insignificant since expectations can always be made to match exactly by suffering a tiny change in the correlation. The proof of Theorem 6 has two separate steps:

1. (**Smooth**) The first step is to show that given any $f : \mathbb{R}^n \to [k]$ with $\mathbf{E}[f] = \boldsymbol{\mu}$, there is a degree $d = d(t, k, \epsilon)$ PTF h on n variables such that $\|\mathbf{E}[h] - \boldsymbol{\mu}\|_1 \le \epsilon$ and $\mathbf{E}[\langle h, P_t h\rangle] \ge \mathbf{E}[\langle f, P_t f\rangle] - \epsilon$. In other words, reduce the degree but not the dimension.

   The main idea here is to modify the function f by first smoothing it and then rounding it back to the discrete set $[k]$. It is fairly easy to show that this procedure doesn't decrease the noise stability of f (as long as the amount of smoothing is chosen to match the noise parameter t). The more difficult part is to show that the result of this procedure is close to a low-degree PTF. This is done using a randomized rounding argument: we show that by rounding the smoothed function at a random threshold, the expected Gaussian surface area of the resulting partition is bounded; in particular, there exists a good way to round. A well-known link between Gaussian surface area and Hermite expansions then implies that the rounded, smoothed function is almost a low-degree PTF. This argument uses the co-area formula, gradient bounds and is inspired by ideas from [KNOW14, Nee14].

2. (**Reduce**) The second step is to show that given a multivariate PTF h, there is a multivariate PTF g on $n_0 = n_0(t, k, \epsilon)$ variables such that the noise stability of g is the same as that of the noise stability of h up to an additive error ϵ. This step uses several ideas and results from [DS14]. To give a brief overview of this part, we start with the notion of an *eigenregular* polynomial which was introduced in [DS14]. A polynomial is said to be δ-eigenregular if for the canonical tensor $\mathcal{A}_p$ associated with the polynomial, the ratio of the maximum singular value to its Frobenius norm is at most δ. Let us assume that $h = \mathsf{PTF}(p_1, \dots, p_k)$. The *regularity lemma* from [DS14], roughly speaking, shows that each of the polynomials $p_1, \dots, p_k$ can be written as a low-degree "outer" polynomial composed with a bounded number of δ-eigenregular, low-degree "inner" polynomials. Using the central limit theorem from [DS14] and several other new technical ingredients, one can replace the whole collection of inner polynomials by a new collection of inner polynomials on a bounded number of variables. Moreover, one can do this replacement while hardly affecting the distribution of the outer polynomial. In particular, this whole procedure constructs a new PTF on a bounded number of inputs, and with approximately the same noise stability as the original PTF.

How to prove Theorem 3: We will first outline the proof of Theorem 3 in the case that $\mathbf{P} = \mathbf{G}_{\rho,2}$ (the ρ-correlated Gaussian measure on $\mathbb{R}^2$). As we observed earlier, any function with codomain $[k]$ naturally maps to $\mathbb{R}^k$ by identifying $i \in [k]$ with the standard unit vector $\mathbf{e}_i \in \mathbb{R}^k$. Also, for any function $f : \mathbb{R}^n \to \mathbb{R}^k$ and $1 \le j \le k$, we let $f_j : \mathbb{R}^n \to \mathbb{R}$ denote the j^{th} coordinate of f. Then, observe that for all $1 \le i, j \le k$,

$$\Pr_{(\mathbf{X}^n, \mathbf{Y}^n) \sim \mathbf{P}^n}[f(\mathbf{X}^n) = i \wedge g(\mathbf{Y}^n) = j] = \mathbf{E}[f_i P_t g_j].$$

In particular, to prove Theorem 3 in the case $\mathbf{P} = \mathbf{G}_{\rho,2}$ it suffices to prove an improvement of Theorem 6, where the inequality $\mathbf{E}[\langle g, P_t g\rangle] \ge \mathbf{E}[\langle f, P_t f\rangle] - \epsilon$ is replaced by an almost-equality: $|\mathbf{E}[g_i P_t g_j] - \mathbf{E}[f_i P_t f_j]| \le \epsilon$ for all i, j. In fact, we will prove something slightly stronger, by starting with a tuple of functions instead of just one.

This proof will follow the same "smooth and reduce" outline as in the proof of Theorem 6. Moreover, the "reduce" step will essentially be the same as the one in [DMN17]. However, the "smooth" step here will be different and hence we outline it here. Define the set $\Delta_{k,\epsilon}$ as

$$\Delta_{k,\epsilon} = \{x \in \mathbb{R}^k : \exists y \in \Delta_k, \quad \|x - y\|_1 \le \epsilon\}.$$

Thus, if $\epsilon = 0$, then $\Delta_{k,\epsilon} = \Delta_k$. In the "smooth" step for the proof of Theorem 3, we will show that for any pair f, g of functions $\mathbf{R}^n \to [k]$ and error parameter $\epsilon > 0$, there exist functions $\tilde{f}, \tilde{g} \colon \mathbf{R}^n \to \mathbf{R}^k$ satisfying the following conditions:

(i) $\|\mathbf{E}[f] - \mathbf{E}[\tilde{f}]\|_1 \leq \epsilon$, $\|\mathbf{E}[g] - \mathbf{E}[\tilde{g}]\|_1 \leq \epsilon$;

(ii) the functions $\tilde{f}$ and $\tilde{g}$ are linear combinations of $O_{k,t,\epsilon}(1)$ low-degree PTFs (with some special structure that we will describe later);

(iii) $\Pr[\tilde{f}(\mathbf{X}^n) \in \Delta_{k,\epsilon}] \geq 1 - \epsilon$ and $\Pr[\tilde{g}(\mathbf{Y}^n) \in \Delta_{k,\epsilon}] \geq 1 - \epsilon$; and

(iv) for any $1 \leq i, j \leq k$, $|\mathbf{E}[\langle f_i P_t g_j\rangle] - \mathbf{E}[\langle \tilde{f}_i P_t \tilde{g}_j\rangle]| \leq \epsilon$.

The precise statement corresponding to this step is given in Lemma 18, which contains most of the technically new ideas in the paper. We remark that because of technical considerations, the formal statement of Lemma 18 considers tuples of functions rather than just a pair of functions which makes the notation more involved. However, the technical gist is contained in the simpler setting discussed here. To prove Lemma 18, we employ a new "boosting" based idea to obtain the functions $\tilde{f}$ and $\tilde{g}$.

The proof of Lemma 18 comes in two main steps. We start with arbitrary functions f and g. First, we show that there are projections of polynomial threshold functions f_{sm} and g_{sm} which have the same low-level Hermite spectrum as f and g. This is carried out in an iterative argument using a potential function, and is inspired by similar iterative algorithms appearing in boosting [Sch90, Fre95] from learning theory, the hardcore lemma in complexity theory [Imp95] and dense model theorems in graph theory [FK99] and additive combinatorics [Tao07, TTV09]. While these iterative algorithms have recently been used to prove structural results in complexity theory [DDFS14, LRS15, TTV09], since our algorithm is in the multidimensional setting, it is somewhat more delicate than these applications. The main argument here is carried out in Lemma 21, and we bound the degree of the resulting polynomials in Corollary 27.

The next step is to show that we can replace the projected polynomial threshold functions by polynomials that with high probability take values very close to the simplex (call them f'_{sm} and g'_{sm}). This is carried out in Lemma 31, using Bernstein approximations for Lipschitz functions. Finally, we use some probabilistic tricks to replace f'_{sm} and g'_{sm} by functions $\tilde{f}$ and $\tilde{g}$ which are linear combinations of low-degree PTFs. This finishes the proof of Lemma 18.

1.2 What happens when P is not Gaussian?

So far, the discussion has pertained to the case when $\mathbf{P} = \mathbf{G}_{\rho,2}$. What happens if $\mathbf{P}$ is a different probability distribution?

As we have remarked earlier, the main result of [GKS16b] is that the $k = 2$ case of Theorem 3 essentially reduces to the special case $\mathbf{P} = \mathbf{G}_{\rho,2}$. Their argument uses quite general tools from Boolean function analysis such as the invariance principle [MOO10, Mos10] and regularity lemmas for low-degree polynomials [DSTW10, DDS14]. A similar argument can be used to prove Theorem 3 by reducing to the Gaussian case; however, we will actually need a slightly stronger Gaussian version of Theorem 3 (note that Theorem 3 only applies to two functions):

Theorem 7. *Let $\mathbf{P} = \mathbf{G}_{\rho,2}$ and let $f^{(1)}, \ldots, f^{(\ell)} \colon \mathbb{R}^n \to [k]$ and $g^{(1)}, \ldots, g^{(\ell)} \colon \mathbb{R}^n \to [k]$ where we define $\mathbf{Q}_{i,j}$ as $\mathbf{Q}_{i,j} = (f^{(i)}(\mathbf{X}^n), g^{(j)}(\mathbf{Y}^n))$. Then, for every $\delta > 0$, there is an explicitly defined constant $n_0 = n_0(\ell, k, \delta)$ and explicitly defined functions $f^{(1)}_{\mathsf{junta}}, \ldots, f^{(\ell)}_{\mathsf{junta}} \colon \mathbb{R}^{n_0} \to [k]$ and $g^{(1)}_{\mathsf{junta}}, \ldots, g^{(\ell)}_{\mathsf{junta}} \colon \mathbb{R}^{n_0} \to [k]$ such that for every $1 \leq i, j \leq \ell$, $\mathrm{d}_{\mathrm{TV}}((f^{(i)}_{\mathsf{junta}}(\mathbf{X}^{n_0}), g^{(j)}_{\mathsf{junta}}(\mathbf{Y}^{n_0})), \mathbf{Q}_{i,j}) \leq \delta$.*

Note that the $\ell = 1$ case of Theorem 7 is exactly the $\mathbf{P} = \mathbf{G}_{\rho,2}$ case of Theorem 3, the proof of which we outlined above. Then $\ell > 1$ case has essentially the same proof, but with more notation.

In order to prove Theorem 3 from Theorem 7, Alice and Bob both execute a "decision tree." By standard arguments from Boolean function analysis (see [O'D14] for definitions of the terminology that follows), Alice and Bob can represent f and g by small decision trees, such that most of the "leaf" functions (call them $\{f^{(i)}\}_{1\leq i\leq \ell}$ and $\{g^{(i)}\}_{1\leq i\leq \ell}$) are *low-influence* functions. The invariance principle of Mossel *et al.* [MOO10, Mos10] allows us to replace $\{f^{(i)}\}_{1\leq i\leq \ell}$ and $\{g^{(i)}\}_{1\leq i\leq \ell}$ by functions of Gaussian variables; essentially, we can pretend that Alice and Bob have access to independent copies of $\mathbf{G}_{\rho,2}$ where ρ is the so-called maximal correlation coefficient of $(\mathbf{X}, \mathbf{Y})$. Finally, we apply Theorem 7 to this collection of Gaussian "leaf" functions. In the end, we have replaced Alice and Bob's initial functions by a pair of decision trees of bounded size, where every leaf function is a function of a bounded number of Gaussian variables. We give a more detailed overview of this reduction in Section A.

1.3 Subsequent work

Subsequent to the appearance of this work on `arXiv`, Ghazi, Kamath and Raghavendra [GKR17] have obtained alternate proofs of the main result here as well as the main result of

[DMN17]. Their proof techniques are different from ours with better quantitative parameters.

1.4 Acknowledgments We thank Pritish Kamath, Badih Ghazi and Madhu Sudan for pointing out that the $\ell = 1$ case of Theorem 7 is not sufficient to derive Theorem 3. (An earlier version of this paper incorrectly claimed that it was.) We also thank the anonymous reviewers who pointed out this gap and for several comments which greatly helped improve the presentation of this paper.

2 Technical preliminaries

We will start by defining some technical preliminaries which will be useful for the rest of the paper.

Definition 8. *For $k \in \mathbb{N}$ and $1 \le i \le k$, let $\mathbf{e}_i$ be the unit vector along coordinate i and let Δ_k be the convex hull formed by $\{\mathbf{e}_i\}_{1 \le i \le k}$.*

In this paper, we will be working on the space of functions $f: \mathbb{R}^n \to \mathbb{R}$ where the domain is equipped with the standard n dimensional normal measure (denoted by $\gamma_n(\cdot)$). Unless explicitly mentioned otherwise, all the functions considered in this paper will be in $L^2(\gamma_n)$. A key property of such functions is that they admit the so-called Hermite expansion. Let us define a family of polynomials $H_q: \mathbb{R} \to \mathbb{R}$ (for $q \ge 0$) as

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

Let $\mathbb{Z}^*$ denote the subset of non-negative integers and $S \in \mathbb{Z}^{*n}$. Define $H_S: \mathbb{R}^n \to \mathbb{R}$ as

$$H_S(z) = \prod_{i=1}^{n} H_{S_i}(z_i).$$

It is well known that the set $\{H_S\}_{S \in \mathbb{Z}^{*n}}$ forms an orthonormal basis for $L^2(\gamma_n)$. In other words, every $f \in L^2(\gamma_n)$ may be written as

$$f = \sum_{S \in \mathbb{Z}^{*n}} \widehat{f}(S) \cdot H_S,$$

where $\widehat{f}(S)$ are typically referred to as the *Hermite coefficients* and expansion is referred to as the *Hermite expansion.* The notion of Hermite expansion can be easily extended to $f: R^n \to \mathbb{R}^k$ as follows: Let $f = (f_1, \ldots, f_k)$ and let

$$f_i = \sum_{S \in \mathbb{Z}^{*n}} \widehat{f}_i(S) \cdot H_S.$$

Then, the Hermite expansion of f is given by $\sum_{S \in \mathbb{Z}^{*n}} \widehat{f}(S) \cdot H_S$ where $\widehat{f}(S) = (\widehat{f}_1(S), \ldots, \widehat{f}_k(S))$. In this setting, we also have Parseval's identity:

$$(2.2) \qquad \int \|f(x)\|_2^2 \, \gamma_n(x) dx = \sum_{S \in \mathbb{Z}^{*n}} \|\widehat{f}(S)\|_2^2$$

For $f: \mathbb{R}^n \to \mathbb{R}^k$ and $d \in \mathbb{N}$, define $f_{\le d}: \mathbb{R}^n \to \mathbb{R}^k$ by

$$f_{\le d}(x) = \sum_{S : |S| \le d} \widehat{f}(S) \cdot H_S(x).$$

Here $|S|$ denotes the ℓ_1 norm of the vector S. We will define $\mathsf{W}^{\le d}[f] = \|f_{\le d}\|_2^2$ and $\mathsf{W}^{>d}[f] = \sum_{|S|>d} \|\widehat{f}(S)\|_2^2$.

Ornstein-Uhlenbeck operator

Definition 9. *The Ornstein-Uhlenbeck operator P_t is defined for $t \in [0, \infty)$ such that for any $f: \mathbb{R}^n \to \mathbb{R}^k$,*

$$(P_t f)(x) = \int_{y \in \mathbb{R}^n} f(e^{-t} \cdot x + \sqrt{1 - e^{-2t}} \cdot y) d\gamma_n(y).$$

Note that if $f: \mathbb{R}^n \to \Delta_k$, then so is $P_t f$ for every $t > 0$. A basic fact about the Ornstein-Uhlenbeck operator is that the functions $\{H_S\}$ are eigenfunctions of this operator. See Proposition 11.37 in [O'D14] for a proof.

Proposition 10. *For $S \in \mathbb{Z}^{*n}$, $P_t H_S = e^{-t \cdot |S|} \cdot H_S$.*

Probabilistic inequalities The following are useful probabilistic inequalities in the analysis of Boolean functions. The first theorem is a higher degree generalization of the well-known Chernoff bound.

Theorem 11. *[Jan97] Let $p: \mathbb{R}^n \to \mathbb{R}$ be a degree-d polynomial. Then, for any $t > e^d$,*

$$\Pr_x\left[|p(x) - \mathbf{E}[p(x)]| \ge t \cdot \sqrt{\mathsf{Var}[p]}\right] \le d \cdot e^{-t^{2/d}}.$$

The next theorem follows by combining Theorem 11 with the well-known result of Carbery-Wright on anti-concentration of low-degree polynomials [CW01]. See Lemma 5 in [DS14] for a short proof.

Theorem 12. *Let $a, b : \mathbb{R}^n \to \mathbb{R}$ be degree d polynomials satisfying $\mathbf{E}_x[a(x) - b(x)] = 0$ and $\mathsf{Var}[a - b] \le (\tau/d)^{3d} \cdot \mathsf{Var}[a]$. Then, $\Pr_x[\mathsf{sign}(a(x)) \ne \mathsf{sign}(b(x))] = O(\tau)$.*

Producing non-integral functions Instead of producing functions $\{f^{(j)}_{\mathsf{junta}}\}_{1 \le i \le \ell}$ and $\{g^{(j)}_{\mathsf{junta}}\}_{1 \le i \le \ell}$ (in Theorem 7) with range $[k]$, we will actually produce functions $\{\tilde{f}^{(j)}_{\mathsf{junta}}\}_{1 \le i \le \ell}$ and $\{\tilde{g}^{(j)}_{\mathsf{junta}}\}_{1 \le i \le \ell}$ whose range will be close to $\Delta_{k,\epsilon}$. The next two lemmas show that functions with range $\Delta_{k,\epsilon}$ can be converted to non-interactive simulation strategies with range $[k]$ with nearly the same guarantees. To state this more precisely, let us adopt the notation that given a point $x \in \mathbb{R}^k$, $\mathsf{Proj}(x)$ denotes the closest point to x in Δ_k in Euclidean distance.

Lemma 13. *Let $f : \mathbb{R}^n \to \mathbb{R}^k$ satisfy the following two conditions:*

1. $\Pr_x[f(x) \notin \Delta_{k,\delta}] \le \delta$.

2. *For all x, $\|f(x)\|_\infty \le k$.*

Then, there is a function $f_1 : \mathbb{R}^n \to \Delta_k$ such that $\|f - f_1\|_1 = O(k \cdot \delta)$.

Proof. Define $f_1 = \mathsf{Proj}(f)$. Note that if x is such that $f(x) \in \Delta_{k,\delta}$, then by definition, $\|f_1(x) - f(x)\|_1 \le \delta$. On the other hand, for any x, $\|f(x) - f_1(x)\|_1 \le k$. This proves the claim. □

Lemma 14. *Let $f_1, g_1 : \mathbb{R}^n \to \Delta_k$. Then, there exist (explicitly defined) $f_2, g_2 : \mathbb{R}^{n+2} \to [k]$ such that*

1. $\mathbf{E}[f_2] = \mathbf{E}[f_1]$ *and* $\mathbf{E}[g_2] = \mathbf{E}[g_1]$.

2. *For any $1 \le j, \ell \le k$,*

$$\mathbf{E}[f_{1,j} P_t g_{1,\ell}] = \mathbf{E}[f_{2,j} P_t g_{2,\ell}].$$

Further, the function f_2 (resp. g_2) is dependent only on f_1 (resp. g_1).

Proof. Let $z = (x, z_1, z_2)$ where $x \in \mathbb{R}^n$ and $z_1, z_2 \in \mathbb{R}$. For any $y \in \Delta_k$, let us divide $\mathbb{R}$ into k intervals $S_1, \ldots, S_k$ such that for $z \sim \gamma$, $\Pr[z \in S_i] = y_i$. For $y \in \Delta_k$ and $z' \in \mathbb{R}$, $\mathsf{Part}(y, z) = i$ if $z' \in S_i$. Define $f_2 : \mathbb{R}^{n+2} \to [k]$ as

$$f_2(z) = f_2(x, z_1, z_2) = \mathsf{Part}(f_1(x), z_1).$$

$$g_2(z) = g_2(x, z_1, z_2) = \mathsf{Part}(g_1(x), z_2).$$

We will now verify the claimed properties. First of all, observe that the codomain of f_2 and g_2 is indeed k. Second, by definition, it is easy to follow that $\mathbf{E}[f_1] = \mathbf{E}[f_2]$ and $\mathbf{E}[g_1] = \mathbf{E}[g_2]$. Finally, note that

$$\mathbf{E}[f_{1,j} P_t g_{1,\ell}] = \mathbf{E}_{(\mathbf{X}^n, \mathbf{Y}^n) \sim \mathbf{P}^n}[f_{1,j}(\mathbf{X}^n) g_{1,\ell}(\mathbf{Y}^n)].$$

On the other hand, suppose $z_1, z_2 \sim \gamma$. Then,

$$\Pr_{z_1, z_2 \sim \gamma}[f_2(x, z_1, z_2) = j \ \wedge \ g_2(y, z_1, z_2) = \ell]$$
$$= f_{1,j}(x) g_{1,\ell}(y).$$

Thus, we obtain that

$$\mathbf{E}[f_{2,j} P_t g_{2,\ell}] = \mathbf{E}[f_{1,j}(\mathbf{X}^n) g_{1,\ell}(\mathbf{Y}^n)] = \mathbf{E}[f_{1,j} P_t g_{1,\ell}].$$

□

2.1 Proof strategy for the main theorem

To describe the proof strategy for the main section, we first define a class of k-ary functions called *polynomial plurality functions* (PPFs) which are closely related to the multivariate PTFs defined in the introduction. For this, let us first define the function arg max as follows

Definition 15. $\arg\max : \mathbb{R}^k \to \mathbb{R}^k$ *is defined as*

$$\arg\max(x_1, \ldots, x_k) = \begin{cases} \mathbf{e}_i & \textit{if } x_i > x_j \ \ \textit{for all } j \ne i \\ 0 & \textit{otherwise} \end{cases}$$

Definition 16. *A function $f : \mathbb{R}^n \to \mathbb{R}^k$ is said to be a PPF of degree-d if there exists a polynomial $p : \mathbb{R}^n \to \mathbb{R}$ of degree d and an index $1 \le j \le x$ such that $f = \arg\max(z)$ where $z_i = \delta_{i=j} \cdot p(x)$. Given polynomial $p : \mathbb{R}^n \to \mathbb{R}$ and $1 \le j \le k$, we define the function $\mathsf{PPF}_{p,j}$ as*

$$\mathsf{PPF}_{p,j}(x) = \arg\max(\underbrace{0, \ldots, 0}_{(j-1) \textit{ times}}, p(x), \underbrace{0, \ldots, 0}_{(n-j) \textit{ times}}).$$

The following is a basic fact about PPFs.

Fact 17. *For any PPF f of degree d, if $f = \mathsf{PPF}_{p,j}$, we can assume without loss of generality that $\mathsf{Var}(p) = 1$. Further, by changing f in at most δ fraction of places, we can assume that $|\mathbf{E}[p(x)]| \le d \cdot \log^{d/2}(1/\delta)$. Such a PPF is said to be a (d, δ)-balanced PPF.*

Proof. The fact about variance follows simply by scaling. To bound $|\mathbf{E}[p(x)]|$, note that if $|\mathbf{E}[p(x)]| > d \cdot \log^{d/2}(1/\delta)$, then $\Pr_x[\mathsf{sign}(p(x)) = \mathsf{sign}(\mathbf{E}[p(x)])] \ge 1 - \delta$ (using Theorem 12). Thus, if we set $q(x) = p(x) - \mathbf{E}[p(x)] + d \cdot \log^{d/2}(1/\delta) \cdot \mathsf{sign}(\mathbf{E}[p(x)])$, then $\Pr_x[p(x) \ne q(x)] \le \delta$. The PPF defined as $\mathsf{PPF}_{q,j}$ satisfies all the desired properties. □

To prove our main theorem (Theorem 7), we will prove the following two intermediate results.

Lemma 18. *For $1 \le i \le \ell$, let $f^{(i)}, g^{(i)} : \mathbb{R}^n \to [k]$ such that $\mathbf{E}[f^{(i)}] = \boldsymbol{\mu}_f^{(i)}$ and $\mathbf{E}[g^{(i)}] = \boldsymbol{\mu}_g^{(i)}$. Then, for any $t > 0$, $\delta > 0$, $d_0 = d_0(t, k, \delta) = (2/t) \cdot \log(k^2/\delta)$ and $1 \le i \le \ell$, there are functions $f_1^{(i)}, g_1^{(i)} : \mathbb{R}^n \to \mathbb{R}^k$ which satisfy the following conditions:*

1. *For any $x \in \mathbb{R}^n$ and $1 \le i \le \ell$, $f_1^{(i)}(x), g_1^{(i)}(x)$ always lies in the positive orthant.*

2. *For any $x \in \mathbb{R}^n$ and $1 \le i \le \ell$, $\|f_1^{(i)}(x)\|_\infty, \|g_1^{(i)}(x)\|_\infty \le 1$.*

3. *For $1 \le i \le \ell$, $\Pr_x[f_1^{(i)}(x) \notin \Delta_{k,k\delta/2}] \le \delta/2$ and $\Pr_x[g_1^{(i)}(x) \notin \Delta_{k,k\delta/2}] \le \delta/2$.*

4. *For $1 \le i \le \ell$, $|\mathbf{E}[f_1^{(i)}] - \boldsymbol{\mu}_f^{(i)}|,\ |\mathbf{E}[g_1^{(i)}] - \boldsymbol{\mu}_g^{(i)}|= O(k\delta)$.*

5. *For $1 \le i,j \le \ell$ and for any $1 \le s_1, s_2 \le k$, $|\mathbf{E}[f_{1,s_1}^{(i)} P_t g_{1,s_2}^{(j)}] - \mathbf{E}[f_{s_1}^{(i)} P_t g_{s_2}^{(j)}]|= O(k \cdot \delta)$.*

6. *For $1 \le i \le \ell$, $f_1^{(i)}$ and $g_1^{(i)}$ are of the following form. There are degree-d_0 polynomials $\{p_{s,j,1}^{(i)}\}_{1\le i\le \ell, 1\le s\le k, 1\le j\le m}$ and $\{p_{s,j,2}^{(i)}\}_{1\le i\le \ell, 1\le s\le k, 1\le j\le m}$*

$$f_1^{(i)} = \sum_{s=1}^{k}\sum_{j=1}^{m} \frac{1}{m}\cdot \mathsf{PPF}_{p_{s,j,1}^{(i)},j}(x)\ ,\ g_1^{(i)} = \sum_{s=1}^{k}\sum_{j=1}^{m} \frac{1}{m}\cdot \mathsf{PPF}_{p_{s,j,2}^{(i)},j}(x),$$

such that the resulting PPFs $\mathsf{PPF}_{p_{s,j,1}^{(i)},j}(x)$ and $\mathsf{PPF}_{p_{s,j,2}^{(i)},j}(x)$ are (d_0,δ)-balanced PPFs. Here $m = O(1/\delta)$.

Further, the function $f_1^{(i)}$ (resp. $g_1^{(i)}$) is dependent only on $f^{(i)}$ (resp. $g^{(i)}$), t, k and δ.

Lemma 19. *Let $\{p_{s,j,1}^{(i)}\}_{1\le i\le \ell, 1\le s\le k, 1\le j\le m}$ and $\{p_{s,j,2}^{(i)}\}_{1\le i\le \ell, 1\le s\le k, 1\le j\le m}$ be degree-d_0 polynomials. For $1 \le i \le \ell$, let $f_1^{(i)}, g_1^{(i)} : \mathbb{R}^n \to \mathbb{R}^k$ be defined as in Lemma 18 and satisfy the following two conditions:*

1. *For $1 \le i \le \ell$, $1 \le s \le k$ and $1 \le j \le m$, all the PPFs $\mathsf{PPF}_{p_{s,j,1}^{(i)},j}$ and $\mathsf{PPF}_{p_{s,j,2}^{(i)},j}$ are (d_0,δ)-balanced PPFs.*

2. *For $1 \le i \le \ell$, $\Pr_x[f_1^{(i)}(x) \notin \Delta_{k,\delta}] \le \delta$ and $\Pr_x[g_1^{(i)}(x) \notin \Delta_{k,\delta}] \le \delta$.*

Then, there exists an explicit constant $n_0 = n_0(d_0,k,\delta,\ell)$ such that there are polynomials $\{r_{s,j,1}^{(i)}\}_{1\le i\le \ell, 1\le s\le k, 1\le j\le m}$ and $\{r_{s,j,2}^{(i)}\}_{1\le i\le \ell, 1\le s\le k, 1\le j\le m}$ satisfying the following conditions: For $1 \le i \le \ell$, let us define the functions $f_{\mathsf{junta}}^{(i)}, g_{\mathsf{junta}}^{(i)} : \mathbb{R}^{n_0} \to \mathbb{R}^k$ defined as

$$f_{\mathsf{junta}}^{(i)} = \sum_{s=1}^{k}\sum_{j=1}^{m} \frac{1}{m}\cdot \mathsf{PPF}_{r_{s,j,1}^{(i)},s}(x),$$

$$g_{\mathsf{junta}}^{(i)} = \sum_{s=1}^{k}\sum_{j=1}^{m} \frac{1}{m}\cdot \mathsf{PPF}_{r_{s,j,2}^{(i)},s}(x),$$

Then, they satisfy the following three conditions: For all $1 \le i \le \ell$,

1. *$\|\mathbf{E}[f_{\mathsf{junta}}^{(i)}] - \mathbf{E}[f_1^{(i)}]\|_1 \le \delta$ and $\|\mathbf{E}[g_{\mathsf{junta}}^{(i)}] - \mathbf{E}[g_1^{(i)}]\|_1 \le \delta$.*

2. *$\Pr_x[f_{\mathsf{junta}}^{(i)}(x) \notin \Delta_{k,\sqrt{\delta}}] \le \sqrt{\delta}$ and $\Pr_x[g_{\mathsf{junta}}^{(i)}(x) \notin \Delta_{k,\sqrt{\delta}}] \le \sqrt{\delta}$.*

3. *For any $1 \le i,j \le \ell$, $1 \le s_1, s_2 \le k$, $|\mathbf{E}[f_{1,s_1}^{(i)} P_t g_{1,s_2}^{(j)}] - \mathbf{E}[f_{\mathsf{junta},s_1}^{(i)} P_t g_{\mathsf{junta},s_2}^{(j)}]| \le \delta$.*

Proof of Theorem 7: The proof of Theorem 7 follows by applying Lemma 18 on the set $\{f^{(i)} \cup g^{(i)}\}_{1\le i\le \ell}$ and subsequently applying Lemma 19. While the range of functions produced by $\{f_{\mathsf{junta}}^{(i)} \cup g_{\mathsf{junta}}^{(i)}\}_{1\le i\le \ell}$ is not Δ_k, by applying Lemma 13 and Lemma 14, we can rectify this issue. We note here that the functions obtained in this process, namely $\{f_{\mathsf{junta}}^{(i)} \cup g_{\mathsf{junta}}^{(i)}\}_{1\le i\le \ell}$ are explicit. Namely, the functions obtained before applying Lemma 13 and Lemma 14 are low-degree PPFs. Lemma 13 applies a projection on to the standard simplex Δ_k. Likewise, Lemma 14 also produces an explicit function as its output. We now explain why $\{f_{\mathsf{junta}}^{(i)} \cup g_{\mathsf{junta}}^{(i)}\}_{1\le i\le \ell}$ satisfy the stated guarantees.

In particular, overloading notation, let us denote the functions obtained by application of Lemma 13 and Lemma 14 as $f_{\mathsf{junta}}^{(i)}$ and $g_{\mathsf{junta}}^{(i)}$. Then, we see that

$$\|\mathbf{E}[f_{\mathsf{junta}}^{(i)}] - \mathbf{E}[f_1^{(i)}]\|_1,\ \|\mathbf{E}[g_{\mathsf{junta}}^{(i)}] - \mathbf{E}[g_1^{(i)}]\|_1 \le O(k \cdot \sqrt{\delta}),$$

$$\text{For any } 1 \le i,j \le \ell,\ 1 \le s_1, s_2 \le k,$$
$$|\mathbf{E}[f_{1,s_1}^{(i)} P_t g_{1,s_2}^{(j)}] - \mathbf{E}[f_{\mathsf{junta},s_1}^{(i)} P_t g_{\mathsf{junta},s_2}^{(j)}]| \le \delta$$

Note that the functions $\{f^{(i)} \cup g^{(i)}\}_{1\le i\le \ell}$ have arity n_0. Further, observe that for $1 \le s_1, s_2 \le k$ and $1 \le i,j \le \ell$,

$$\begin{aligned}&\Pr[f_{\mathsf{junta}}^{(i)}(\mathbf{X}^{n_0}) = s_1\ \wedge\ g_{\mathsf{junta}}^{(j)}(\mathbf{Y}^{n_0}) = s_2]\\ &= \mathbf{E}[f_{\mathsf{junta},s_1}^{(i)} P_t g_{\mathsf{junta},s_2}^{(j)}] \text{ and}\\ &\Pr[f^{(i)}(\mathbf{X}^{n}) = s_1\ \wedge\ g^{(j)}(\mathbf{Y}^{n}) = s_2] = \mathbf{E}[f_{s_1}^{(i)} P_t g_{s_2}^{(j)}].\end{aligned}$$

Thus, for $1 \le s_1, s_2 \le k$,

$$\begin{aligned}&|\Pr[f_{\mathsf{junta}}^{(i)}(\mathbf{X}^{n_0}) = s_1\ \wedge\ g_{\mathsf{junta}}^{(j)}(\mathbf{Y}^{n_0}) = s_2]\\ &\quad - \Pr[f^{(i)}(\mathbf{X}^{n}) = s_1\ \wedge\ g^{(j)}(\mathbf{Y}^{n}) = s_2]| \le \delta.\end{aligned}$$

This immediately implies that

$$\mathrm{d}_{\mathrm{TV}}((f_{\mathsf{junta}}^{(i)}(\mathbf{X}^{n_0}), g_{\mathsf{junta}}^{(j)}(\mathbf{Y}^{n_0})), (f^{(i)}(\mathbf{X}), g^{(j)}(\mathbf{Y}))) = O(k^2\delta),$$

which finishes the proof.

3 Proof of Lemma 18

The proof of Lemma 18 shall proceed in several steps. Note that Lemma 18 claims existence of $\{f_1^{(i)}\}$ and $\{g_1^{(i)}\}$ which satisfies six different properties. The functions $\{f^{(i)}\}$ and $\{g^{(i)}\}$ themselves satisfy the first five properties and thus, the only non-trivial task that remains is to achieve the sixth property. The sixth property will be achieved by gradual modification of $\{f^{(i)}\}$ and $\{g^{(i)}\}$ in a sequence of steps which are explained below.

1. Corollary 22 allows us to replace $f^{(i)}$ (resp. $g^{(i)}$) with $f_{\mathsf{sm}}^{(i)}$ (resp. $g_{\mathsf{sm}}^{(i)}$), which is the projection onto Δ_k of a polynomial, and which shares the same low-degree Hermite expansion as $f^{(i)}$ (resp. $g^{(i)}$). Coupled with Claim 20, this shows that if $f^{(i)}$ is replaced by $f_{\mathsf{sm}}^{(i)}$ and $g^{(i)}$ is replaced by $g_{\mathsf{sm}}^{(i)}$, then the first five properties in Lemma 18 hold. On the other hand, note that while $f_{\mathsf{sm}}^{(i)}$ and $g_{\mathsf{sm}}^{(i)}$ do not have the full structure claim in Property 6, they do have some resemblance to PPFs. Corollary 22 is the technically most innovative part of the proof and in turn relies on Lemma 21. A crucial point for the application to non-interactive simulation is that the construction of $f_{\mathsf{sm}}^{(i)}$ (resp. $g_{\mathsf{sm}}^{(i)}$) is dependent only on $f^{(i)}$ (resp. $g^{(i)}$) and the error parameters.

2. Applying Bernstein-type approximations for Lipschitz functions in terms of low-degree polynomials, Lemma 31 shows that $f_{\mathsf{sm}}^{(i)}$ and $g_{\mathsf{sm}}^{(i)}$ can be replaced by $f_{\mathsf{sm}}^{'(i)}$ and $g_{\mathsf{sm}}^{'(i)}$ where each coordinate of $f_{\mathsf{sm}}^{'(i)}$ and $g_{\mathsf{sm}}^{'(i)}$ is a low-degree multivariate polynomial. Again, crucially for the application to non-interactive simulation, the function $f_{\mathsf{sm}}^{'(i)}$ (resp. $g_{\mathsf{sm}}^{'(i)}$) is dependent only on $f_{\mathsf{sm}}^{(i)}$ (resp. $g_{\mathsf{sm}}^{(i)}$) and the error parameters.

3. Finally, the functions $f_{\mathsf{sm}}^{'(i)}$ and $g_{\mathsf{sm}}^{'(i)}$ are changed to $f_1^{(i)}$ and $g_1^{(i)}$ which are linear combinations of PPFs (as promised in Lemma 18) using some simple probabilistic observations. Again, the conversion of $f_{\mathsf{sm}}^{'(i)}$ to $f_1^{(i)}$ is only dependent on $f_{\mathsf{sm}}^{'(i)}$ and desired error parameters.

3.1 Projections of polynomials We begin with the first step described above. The first lemma relates the (by now, well-known) connection between the low-degree Hermite expansion of a function and its noise stability. In particular, it shows that if a pair of functions $(f^{(1)}, g^{(1)})$ (whose range is Δ_k) is replaced by another pair $(\underline{f}^{(1)}, \underline{g}^{(1)})$ such that low-degree Hermite spectrum of $f^{(1)}$ (resp. $g^{(1)}$) is close to that of $\underline{f}^{(1)}$ (resp. $\underline{g}^{(1)}$) are close to each other, then for any $1 \le s_1, s_2 \le k$, $\mathbf{E}[f_{s_1}^{(1)} P_t g_{s_2}^{(1)}] \approx \mathbf{E}[\underline{f}_{s_1}^{(1)} P_t \underline{g}_{s_2}^{(1)}]$.

Claim 20. *Let $f^{(1)}, g^{(1)}, \underline{f}^{(1)}, \underline{g}^{(1)} : \mathbb{R}^n \to \Delta_k$ such that for $d_1 = d_1(\delta, t) = \frac{1}{t}\log(k^2/\delta)$ we have*

$$\mathsf{W}^{\le d_1}[(f^{(1)} - \underline{f}^{(1)})],\ \mathsf{W}^{\le d_1}[g^{(1)} - \underline{g}^{(1)}] \le \delta^2/k^4.$$

Then, $\sum_{1 \le s_1, s_2 \le k} |\mathbf{E}[f_{s_1}^{(1)} P_t g_{s_2}^{(1)}] - \mathbf{E}[\underline{f}_{s_1}^{(1)} P_t \underline{g}_{s_2}^{(1)}]| \le \delta$.

Proof. For any $1 \le s_1, s_2 \le k$,

$$\begin{aligned}&|\mathbf{E}[f_{s_1}^{(1)} P_t g_{s_2}^{(1)}] - \mathbf{E}[\underline{f}_{s_1}^{(1)} P_t \underline{g}_{s_2}^{(1)}]|\\ &\le |\mathbf{E}[(f_{s_1}^{(1)} - \underline{f}_{s_1}^{(1)}) P_t g_{s_2}^{(1)}]| + |\mathbf{E}[\underline{f}_{s_1}^{(1)} P_t (g_{s_2}^{(1)} - \underline{g}_{s_2}^{(1)})]|\end{aligned}$$

By using the self-adjointness of the noise operator and applying the Jensen's inequality, the first term can be bounded as

$$\begin{aligned}|\mathbf{E}[(f_{s_1}^{(1)} - \underline{f}_{s_1}^{(1)}) P_t g_{s_2}^{(1)}]| &\le \sqrt{\mathbf{E}[P_t(f^{(1)} - \underline{f}^{(1)})_{s_1}^2]}\sqrt{\mathbf{E}[(g^{(1)})_{s_2}^2]}\\ &\le \sqrt{\mathbf{E}[P_t(f^{(1)} - \underline{f}^{(1)})_{s_1}^2]}.\end{aligned}$$

Similarly bounding $|\mathbf{E}[\underline{f}_{s_1}^{(1)} P_t(g_{s_2}^{(1)} - \underline{g}_{s_2}^{(1)})]|$, we obtain

$$\begin{aligned}&|\mathbf{E}[(f_{s_1}^{(1)} - \underline{f}_{s_1}^{(1)}) P_t g_{s_2}^{(1)}]| + |\mathbf{E}[\underline{f}_{s_1}^{(1)} P_t (g_{s_2}^{(1)} - \tilde{g}_{s_2}^{(1)})]|\\ &\le \sqrt{\mathbf{E}[P_t(f_{s_1}^{(1)} - \underline{f}_{s_1}^{(1)})^2]} + \sqrt{\mathbf{E}[P_t(g_{s_2}^{(1)} - \underline{g}_{s_2}^{(1)})^2]}.\end{aligned}$$

Now, applying the condition that $\mathsf{W}^{\le d_1}[(f^{(1)} - f^{(2)})] \le \delta^2/k^4$, we get that

$$\mathbf{E}[\|P_t(f^{(1)} - \underline{f}^{(1)})\|_2^2] \le \frac{\delta^2}{k^4} + e^{-2td_1} \cdot \mathbf{E}[\|(f^{(1)} - \underline{f}^{(1)})\|_2^2] \le \frac{2\delta^2}{k^4}.$$

The last inequality uses the fact that for all x, $\|f^{(1)}(x) - \underline{f}^{(1)}(x)\|_1 \le 1$. Likewise, we also get $\mathbf{E}[\|P_t(g^{(1)} - \underline{g}^{(1)})\|_2^2 \le 2\delta^2/k^4$. Combining this, we obtain that for all $1 \le s_1, s_2 \le k$,

$$|\mathbf{E}[f_{s_1}^{(1)} P_t g_{s_2}^{(1)}] - \mathbf{E}[\underline{f}_{s_1}^{(1)} P_t \underline{g}_{s_2}^{(1)}]| \le \frac{2\delta}{k^2}.$$

Summing over all $1 \le s_1, s_2 \le k$, we get the stated bound. □

Next, we state the main technical lemma of this section. To state the lemma, we define the function $\mathsf{Proj} : \mathbb{R}^k \to \Delta_k$ such that $\mathsf{Proj}(x) = y$ if y is the closest point (in Euclidean distance) to x in Δ_k. While the authors are aware that technically, we require Proj to be quantified by the parameter k, the relevant k shall always be clear from the context. Though we do use other ℓ_p norms in this paper, it is crucial that the projection is defined in terms of ℓ_2 norm.

Lemma 21. *Let $F : \mathbb{R}^n \to \Delta_k$ and let $g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}^k$ be an orthonormal sequence of functions under the standard n-dimensional Gaussian measure γ_n. Here the function $g_1 : x \mapsto (1, \ldots, 1)$. Then, for any $\delta > 0$, there exists a function $F_{\mathsf{proj}} : \mathbb{R}^n \to \Delta_k$ of the form $F_{\mathsf{proj}} = \mathsf{Proj}(\sum_{i=1}^m \kappa_i g_i)$ satisfying*

$$\sum_{i=1}^{m} (\mathbf{E}[g_i F] - \mathbf{E}[g_i F_{\mathsf{proj}}])^2 \leq \delta.$$

Further, $\sum_{i=1}^m \|\kappa_i\|_2^2 \leq \delta^{-2}$.

Before proving Lemma 21, we first see why this lemma is useful. In particular, we have the following corollary. Essentially, the corollary says that given $f, g : \mathbb{R}^n \to \Delta_k$, there are functions f_{sm} and g_{sm} such that (i) the low-level Hermite spectrum of f (resp. g) is close to f_{sm} (resp. g_{sm}) (ii) Both f_{sm} and g_{sm} are obtained by applying the function Proj on a low-degree polynomial. In essence, we are obtaining *simple* functions f_{sm} and g_{sm} which simultaneously (i) have the same low-level Hermite spectrum as f and g (ii) and have range Δ_k.

Corollary 22. *Given function $f : \mathbb{R}^n \to [k]$, $d \in \mathbb{N}$ and error parameter $\delta > 0$, there is a function $f_{\mathsf{sm}} : \mathbb{R}^n \to \Delta_k$ which has the following properties:*

1. *The function f_{sm} has the following form:*

   $$f_{\mathsf{sm}}(x) = \mathsf{Proj}\bigg(\sum_{|S| \leq d} \alpha_{f,s} H_S(x) \bigg),$$

   where $H_S(x)$ is the Hermite polynomial corresponding to the multiset S.

2. *$\sum_{|S| \leq d} \|\alpha_{f,S}\|_2^2 \leq \delta^{-2}$.*

3. *Define $\beta_{f,S} = \mathbf{E}[f_{\mathsf{sm}}(x) \cdot H_S(x)]$. Then, $\sum_{|S| \leq d} \|\beta_{f,S} - \alpha_{f,S}\|_2^2 \leq \delta$.*

We note that for a scalar-valued function H_S and a vector-valued function f_{sm}, we compute $\mathbf{E}[f_{\mathsf{sm}} \cdot H_S]$ pointwise for each coordinate of the vector valued function f_{sm}.

The proof of this corollary follows straightaway by instantiating Lemma 21 with $\{g_1, \ldots, g_m\} = \{H_S\}_{|S| \leq d}$ with $F = f$ and $F = g$.

Proof of Lemma 21: We will prove this lemma via an iterative argument. We will define a sequence of functions $\{F_t\}_{t \geq 0}$ iteratively such that for all $t \geq 0$, $F_t : \mathbb{R}^n \to \Delta_k$. Define the vector $\beta \in \mathbb{R}^m$ by $\beta_j = \langle F, g_j \rangle$. Also, for every $t \geq 0$, we will define $\beta_t \in \mathbb{R}^m$ by $\beta_{t,j} = \langle F_t, g_j \rangle$. The iterative process has the following property: If for any t, $\|\beta_t - \beta\|_2^2 \leq \delta$, then we terminate the process. Else, we modify F_t to obtain the function F_{t+1}. We now define the initial function F_0 as well as the modification to obtain F_{t+1} from F_t (when $t \geq 0$).

The function $F_0 : \mathbb{R}^n \to \Delta_k$ is defined as $F_0 : x \to (1/k, \ldots, 1/k)$. Next, given F_t, we define F_{t+1}. To do this, we will also need to define an auxiliary sequence of functions $\{G_t\}_{t \geq 0}$ where $G_0 = F_0$. The iterative process is defined in Figure 1.

Description of iterative process

1. Define $\rho_t = \|\beta_t - \beta\|_2$.
2. If $\rho_t^2 \leq \delta$, then stop the process. Else, we define $J_t = \sum_{j=1}^m (\beta - \beta_t)_j \cdot g_j$.
3. Define $G_{t+1} = G_t + J_t/2$. Define $F_{t+1} = \mathsf{Proj}(G_{t+1})$ and $t \leftarrow t + 1$. Go to Step 1.

Figure 1: Iterative process describing the sequence $\{F_t\}$

It is clear that if this process terminates at step $t = t_0$, then the function $F_{\mathsf{proj}} = F_{t_0}$ satisfies the required properties. Thus, we now need to bound the convergence rate of the process. To do this, we introduce a potential function $\Psi(t)$ defined as follows:

$$\Psi(t) = \mathbf{E}[\langle F - F_t, F - 2G_t + F_t \rangle].$$

The basic observation here is that $\Psi(0) = O(1)$. We will prove two main lemmas. The first will prove that in every iteration of the process in Figure 1, $\Psi(t)$ decreases by a fixed amount. The second is that $\Psi(t)$ is always non-negative. These two facts, in conjunction, automatically imply an upper bound on the maximum number of steps in the algorithm.

Claim 23.

$$\mathbf{E}[\langle F - F_t, J_t \rangle] = \rho_t^2.$$

Proof. By orthogonality of the functions $\{g_j\}_{j=1}^m$,

$$\begin{aligned} \mathbf{E}[\langle F - F_t, J_t \rangle] &= \sum_{j=1}^{m} (\beta - \beta_t)_j \mathbf{E}[\langle g_j, F - F_t \rangle] \\ &= \sum_{j=1}^{m} (\beta - \beta_t)_j \cdot (\beta - \beta_t)_j = \|\beta - \beta_t\|_2^2. \end{aligned}$$

□

We now recall a basic fact about projective maps (see, e.g. [CG59, Theorem 3]).

Fact 24. *Let C be a closed, convex set and let $\mathsf{Proj}_C : \mathbb{R}^n \to C$ be defined as $x \mapsto \arg\min_{y \in C} \|x -$*

$y\|_2$. *Then the map* Proj_C *is uniquely defined, and always contractive i.e. for any* $z, z' \in \mathbb{R}^n$, $\|\mathsf{Proj}_C(z) - \mathsf{Proj}_C(z')\|_2 \leq \|z - z'\|_2$. *Moreover, for any* $x \in C$ *and any* $z \in \mathbf{R}^n$, $\langle z - \mathsf{Proj}_C(z), x - \mathsf{Proj}_C(z)\rangle \leq 0$.

Claim 25. *For all* t, $\Psi(t) \geq 0$.

Proof.

$$\begin{aligned}\Psi(t) &= \mathbf{E}[\langle F - F_t, F - 2G_t + F_t\rangle] \\ &= \mathbf{E}[\langle F - F_t, F - F_t\rangle] + 2 \cdot \mathbf{E}[\langle F - F_t, F_t - G_t\rangle].\end{aligned}$$

The first term is clearly non-negative. The second is non-negative by Fact 24, taking $z = G_t$ and $x = F$. □

The next lemma shows that the potential function always decreases by a fixed quantity.

Lemma 26.

$$\Psi(t+1) - \Psi(t) \leq -\frac{\rho_t^2}{4}.$$

Proof.

$$\begin{aligned} & \Psi(t+1) - \Psi(t) \\ = \; & \mathbf{E}[\langle F - F_{t+1}, F - 2G_{t+1} + F_{t+1}\rangle] \\ - \; & \mathbf{E}[\langle F - F_t, F - 2G_t + F_t\rangle] \\ = \; & \mathbf{E}[\langle F - F_t, 2(G_t - G_{t+1})\rangle] \\ + \; & \mathbf{E}[\langle F_{t+1} - F_t, 2G_{t+1} - F_t - F_{t+1}\rangle] \\ = \; & \mathbf{E}[\langle F - F_t, -J_t\rangle] \\ + \; & \mathbf{E}[\langle F_{t+1} - F_t, 2G_{t+1} - F_t - F_{t+1}\rangle] \\ = \; & -\rho_t^2 + \mathbf{E}[\langle F_{t+1} - F_t, 2G_{t+1} - F_t - F_{t+1}\rangle] \\ & \text{(applying Claim 23)} \\ = \; & -\rho_t^2 + 2 \cdot \mathbf{E}[\langle F_{t+1} - F_t, G_{t+1} - F_{t+1}\rangle] \\ + \; & \mathbf{E}[\langle F_{t+1} - F_t, F_{t+1} - F_t\rangle] \\ = \; & -\rho_t^2 + \mathbf{E}[\|F_{t+1} - F_t\|_2^2] \\ + \; & 2 \cdot \mathbf{E}[\langle F_{t+1} - F_t, G_{t+1} - F_{t+1}\rangle] \\ \leq \; & -\rho_t^2 + \mathbf{E}[\|G_{t+1} - G_t\|_2^2] \\ + \; & 2 \cdot \mathbf{E}[\langle F_{t+1} - F_t, G_{t+1} - F_{t+1}\rangle] \text{ (applying Fact 24)} \\ = \; & -\frac{3\rho_t^2}{4} + 2 \cdot \mathbf{E}[\langle F_{t+1} - F_t, G_{t+1} - F_{t+1}\rangle] \end{aligned}$$

It remains to show that $\mathbf{E}[\langle F_{t+1} - F_t, G_{t+1} - F_{t+1}\rangle] \leq \frac{\rho_t^2}{4}$. Indeed, the Cauchy-Schwarz inequality yields

$$\begin{aligned} & \|F_{t+1} - F_t\|_2 \|G_{t+1} - G_t\|_2 \geq \langle G_{t+1} - G_t, F_{t+1} - F_t\rangle \\ & = \langle G_{t+1} - F_{t+1}, F_{t+1} - F_t\rangle + \langle F_{t+1} - F_t, F_{t+1} - F_t\rangle \\ & + \langle F_t - G_t, F_{t+1} - F_t\rangle \end{aligned}$$

In the last line above, the second term is obviously non-negative. Moreover, the third term is non-negative by Fact 24 (take $z = G_t$ and $x = F_{t+1}$). Hence,

$$\begin{aligned}\langle G_{t+1} - F_{t+1}, F_{t+1} - F_t\rangle &\leq \|F_{t+1} - F_t\|_2 \|G_{t+1} - G_t\|_2 \\ &\leq \|G_{t+1} - G_t\|_2^2 = \frac{\rho_t^2}{4},\end{aligned}$$

where the second inequality follows from Fact 24. □

Combining Claim 25 and Lemma 26, we obtain that the iterative process described in Figure 1 stops in at most $4/\delta$ steps. If the above iteration stops after $t = t_0$ steps, we let $F_{\mathsf{proj}} = F_{t_0}$. Note that $F_{\mathsf{proj}} = \mathsf{Proj}(\sum_{0 \leq t < t_0} J_t/2)$. Thus, it is clear that $F_{\mathsf{proj}} = \mathsf{Proj}(\sum_{i=1}^m \kappa_i g_i)$. To bound $\sum_{i=1}^m \|\kappa_i\|_2^2$, note that

$$\begin{aligned}\sum_{i=1}^m \|\kappa_i\|_2^2 = \|\sum_{0 \leq t < t_0} J_t/2\|_2^2 &\leq t_0 \cdot \sum_{0 \leq t < t_0} \|J_t/2\|_2^2 \\ &\leq t_0^2 \cdot \max_t \|J_t/2\|_2^2 \leq t_0^2.\end{aligned}$$

The very last inequality uses the fact that $\|J_t\|_2 \leq \|(F_t - F)\|_2 \leq 1$. Plugging the upper bound of $O(1/\delta^2)$ on t_0^2, we obtain that $\sum_{i=1}^m \|\kappa_i\|_2^2 \leq O(1/\delta^2)$. This concludes the proof. □

Corollary 27. *For* $t > 0$, *error parameter* $\delta > 0$ *and any function* $f : \mathbb{R}^n \to [k]$, *there is a function* $f_{\mathsf{sm}} : \mathbb{R}^n \to \Delta_k$ *such that for* $d = (2/t) \cdot \log(k^2/\delta)$, *we have the following:*

1. $\|\mathbf{E}[f_{\mathsf{sm}}] - \mathbf{E}[f]\|_1 \leq \delta$.
2. *The function* $f_{\mathsf{sm}} = \mathsf{Proj}(p_{f,1}(x), \ldots, p_{f,k}(x))$ *where for all* $1 \leq s \leq k$, $p_{f,s} : \mathbb{R}^n \to \mathbb{R}$ *are polynomials of degree* d *and* $\mathsf{Var}(p_{f,s}) \leq k^8/\delta^4$.
3. *For any* $g : \mathbb{R}^n \to [k]$ *and the corresponding function* $g_{\mathsf{sm}} : \mathbb{R}^n \to \Delta_k$, *we have* $\sum_{1 \leq s_1, s_2 \leq k} |\mathbf{E}[f_{\mathsf{sm},s_1} P_t g_{\mathsf{sm},s_2}] - \mathbf{E}[f_{s_1} P_t g_{s_2}]| \leq \delta$.

Proof. Given the function $f : \mathbb{R}^n \to [k]$, we apply Corollary 22 to obtainthe function $f_{\mathsf{sm}} : \mathbb{R}^n \to \Delta_k$ where

$$f_{\mathsf{sm}} = \mathsf{Proj}(p_{f,1}(x), \ldots, p_{f,k}(x)),$$

where for all $1 \leq s \leq k$, $p_{f,s} : \mathbb{R}^n \to \mathbb{R}$ are polynomials of degree $d = (1/t) \cdot \log(k^2/\delta)$ such that $\mathsf{W}^{\leq d}[(f_{\mathsf{sm}} - f)] \leq \delta^2/k^4$. Further, for each $1 \leq s \leq k$, $\mathsf{Var}(p_{f,s}) \leq (k^8/\delta^4)$. This immediately implies both items 1 and 2. To prove Item 3, note that we also have $\mathsf{W}^{\leq d}[(g_{\mathsf{sm}} - g)] \leq \delta^2/k^4$. Applying Claim 20, we obtain that $\sum_{1 \leq s_1, s_2 \leq k} |\mathbf{E}[f_{\mathsf{sm},s_1} P_t g_{\mathsf{sm},s_2}] - \mathbf{E}[f_{s_1} P_t g_{s_2}]| \leq \delta$. This proves Item 3. □

This completes the first step in the outline of Lemma 5: we have replaced arbitrary functions by projections of polynomials.

3.2 Bernstein approximation

The next step in the proof of Lemma 5 is the removal of the projection. The basic idea is just to approximate the projection map by a polynomial. Then, the projection of a polynomial becomes the composition of two polynomials, which is still a polynomial.

Definition 28. *For $0 \leq k \leq d$, efine $p_{k,d}(x) = \binom{d}{k} x^k (1-x)^{d-k}$. For a function $f : [0,1]^\ell \to \mathbb{R}$, define the polynomial $\mathsf{BP}_{f,d_1,\ldots,d_\ell}$ by*

$$\mathsf{BP}_{f,d_1,\ldots,d_\ell}(x) = \sum_{k_1,\ldots,k_\ell} f\left(\frac{k_1}{d_1}, \ldots, \frac{k_\ell}{d_\ell}\right) p_{k_1,d_1}(x_1) \cdots p_{k_\ell,d_\ell}(x_\ell).$$

We call $\mathsf{BP}_{f,d_1,\ldots,d_\ell}$ the multivariate Bernstein approximation for f with degrees $(d_1, \ldots, d_\ell)$.

Theorem 29. ***Multivariate Bernstein approximations*** *Let $f : [0,1]^\ell \to \mathbb{R}$ be a L-Lipschitz function in $[0,1]^\ell$. In other words, $\|f(x) - f(y)\|_2 \leq L \cdot \|x - y\|_2$. Then $\mathsf{BP}_{f,d_1,\ldots,d_\ell}$ satisfies the inequality*

$$\sup_{z \in [0,1]^\ell} |f(z) - \mathsf{BP}_{f,d_1,\ldots,d_\ell}(z)| \leq \frac{L}{2} \cdot \left(\sum_{j=1}^{\ell} \frac{1}{d_j}\right)^{1/2}$$

The proof of Theorem 29 is folklore; we provide one for completeness.

Proof. Fix $z \in [0,1]^\ell$. Note that each $p_{k_i,d_i}(z_i)$ is non-negative, and that $\sum_{k_i=0}^{d_i} p_{k_i,d_i}(z_i) = 1$. Hence,

$$\begin{aligned}
&f(z) - \mathsf{BP}_{f,d_1,\ldots,d_\ell}(z) \\
&= \sum_{k_1,\ldots,k_\ell} \left[f(z) - f\left(\frac{k_1}{d_1}, \ldots, \frac{k_\ell}{d_\ell}\right)\right] \prod_{j=1}^{\ell} p_{k_j,d_j}(z_j) \\
&\leq L \sum_{k_1,\ldots,k_\ell} \left\|z - \left(\frac{k_1}{d_1}, \ldots, \frac{k_\ell}{d_\ell}\right)\right\|_2 \prod_{j=1}^{\ell} p_{k_j,d_j}(z_j) \\
&\leq L \left[\sum_{k_1,\ldots,k_\ell} \left\|z - \left(\frac{k_1}{d_1}, \ldots, \frac{k_\ell}{d_\ell}\right)\right\|_2^2 \prod_{j=1}^{\ell} p_{k_j,d_j}(z_j)\right]^{1/2} \\
&= L \left[\sum_{i=1}^{\ell} \sum_{k_i=0}^{d_i} \left(z_i - \frac{k_i}{d_i}\right)^2 p_{k_i,d_i}(z_i)\right]^{1/2}.
\end{aligned}$$

Finally, note that $\sum_{k=0}^{d} (x - k/d)^2 p_{k,d}(x)$ is just the variance of a binomial random variable with d trials and success probability x. This is bounded by $\frac{1}{4d}$. Plugging in this bound for each i separately completes the proof. □

Rescaling the function, we have the following corollary. To state this corollary, we let $B(x,r) = \{z : \|z - x\|_2 \leq r\}$ i.e. the ℓ_2 of radius r at x.

Corollary 30. *Let $f : B(x,r) \to \mathbb{R}$ be a 1-Lipschitz function (where $B(x,r) \subseteq \mathbb{R}^\ell$). Then, given any error parameter $\eta > 0$, there is a polynomial $p_{f,r,\eta}$ whose degree in every variable is at most $d_B(\eta, r, \ell) = \ell \cdot 4r^2 \cdot (1/\eta^2)$ such that*

$$\sup_{z \in B(x,r)} |p_{f,r,\eta}(z) - f(z)| \leq \eta.$$

Proof. To prove this, we will rely on Theorem 29. First, define $B_\infty(x,r) = \{z : \|z - x\|_\infty \leq r\}$. We extend f to $B_\infty(x,r)$ as follows: $f(z) = f(\mathsf{Proj}_{B(x,r)}(z))$. Note that the extension is 1-Lipschitz (using Fact 24). Define the function $g : [0,1]^\ell \to \mathbb{R}$ as

$$g(z) = f\left(x + \left(z - \mathbf{\frac{1}{2}}\right) \cdot 2r\right).$$

Here $\mathbf{\frac{1}{2}}$ is the point in $\mathbb{R}^\ell$ which is $1/2$ in every coordinate. It is easy to see that the function g is $2r$-Lipschitz. Thus, if we choose the function $\mathsf{BP}_{g,d_1,\ldots,d_\ell}$, then we have

$$\sup_{z \in [0,1]^\ell} |\mathsf{BP}_{g,d_1,\ldots,d_\ell} - g(z)| \leq 2r \cdot \left(\sum_{j=1}^{\ell} \frac{1}{d_j}\right)^{1/2}.$$

In particular, we set all the degrees $d_1 = \ldots = d_\ell = \ell \cdot 4r^2 \cdot (1/\eta^2)$, then $\sup_{z \in [0,1]^\ell} |\mathsf{BP}_{g,d_1,\ldots,d_\ell} - g(z)| \leq \eta$. Thus, if we set $p_{f,r,\eta}(z)$ as

$$p_{f,r,\eta}(z) = \mathsf{BP}_{g,d_1,\ldots,d_\ell}\left(\frac{z - x}{2r} + \mathbf{\frac{1}{2}}\right).$$

It is clear that the polynomial $p_{f,r,z}$ satisfies $\sup_{z \in B(x,r)} |p_{f,r,\eta}(z) - f(z)| \leq \eta$. □

We next modify the function $f_{\mathsf{sm}} : \mathbb{R}^n \to \Delta_k$ obtained in Corollary 27 to obtain the function $f'_{\mathsf{sm}} : \mathbb{R}^n \to \mathbb{R}^k$ which is a (i) low-degree polynomial and (ii) f_{sm} is close to f'_{sm} with high probability on the Gaussian measure γ_n.

Lemma 31. *Given the function $f_{\mathsf{sm}} : \mathbb{R}^n \to \Delta_k$ from Corollary 27, there is a function $f'_{\mathsf{sm}} : \mathbb{R}^n \to \mathbb{R}^k$ such that $f'_{\mathsf{sm}} = (p'_{f,1}(x), \ldots, p'_{f,k}(x))$ where for all $1 \leq s \leq k$, $p'_{f,s} : \mathbb{R}^n \to \mathbb{R}$ are polynomials satisfying the following conditions:*

1. *For $1 \leq s \leq k$, the polynomials $\{p'_{f,s}\}$ have degree $d' = \log^d(dk/\delta) \cdot \mathsf{poly}(k/\delta) \cdot d$ where d is the degree appearing in Corollary 27.*

2. $\Pr_{x \sim \gamma_n}[\|f_{\mathsf{sm}}(x) - f'_{\mathsf{sm}}(x)\|_\infty \leq \delta/4] \leq \delta/2.$

Proof. Let the function $f_{\mathsf{sm}}(x) = \mathsf{Proj}(p_{f,1}(x), \dots, p_{f,k}(x))$. Since all the polynomials are degree d and have variance at most $\sigma_{\mathsf{sm}}^2 = k^8/\delta^4$, using Theorem 11, we obtain the following:

(3.3)
$$\Pr_{x \sim \gamma_n} \sup_{1 \le s \le k} [|p_{f,s} - \mathbf{E}[p_{f,s}]| \le \log^{d/2}(2dk/\delta) \cdot \sigma_{\mathsf{sm}}] \le \frac{\delta}{2}.$$

Define the point $\boldsymbol{\mu}_{sm,f} = (\mathbf{E}[p_{f,1}], \dots, \mathbf{E}[p_{f,s}])$. Also, let $r_{sm} = \log^{d/2}(2dk/\delta) \cdot \sigma_{\mathsf{sm}}$. Since the projection from $\mathbf{R}^k$ to Δ_k is Lipschitz, Corollary 30 implies that there exist polynomials $p_{\mathsf{sm},s} : \mathbb{R}^k \to \mathbb{R}$ (for $1 \le s \le k$) whose degree in every variable is at most $k \cdot 4r_{sm}^2 \cdot 16/\delta^2 = \log^d(dk/\delta) \cdot \mathsf{poly}(k/\delta)$, and which satisfy for all $z \in B(\boldsymbol{\mu}_{sm,f}, r_{sm})$, we have

$$|p_{\mathsf{sm},s}(z) - \mathsf{Proj}_s(z)| \le \frac{\delta}{4} \tag{3.4}$$

Let $p_{\mathsf{sm}} : \mathbb{R}^k \to \mathbb{R}^k$ be defined as the map $p_{\mathsf{sm}}(x) = (p_{\mathsf{sm},1}(x), \dots, p_{\mathsf{sm},k}(x))$. Recall that $f_{\mathsf{sm}} = \mathsf{Proj}(p_{f,1}(x), \dots, p_{f,k}(x))$. We define $p'_f = p_{\mathsf{sm}} \circ (p_{f,1}, \dots, p_{f,k})$. We now define $f'_{\mathsf{sm}} = (p'_{f,1}(x), \dots, p'_{f,k}(x))$. It is clear that for $1 \le s \le k$, $p'_{f,s}$ is a polynomials of degree $\log^d(dk/\delta) \cdot \mathsf{poly}(k/\delta) \cdot d$. Likewise, combining (3.4) and (3.3), we obtain that $\Pr_{x \sim \gamma_n}[\| f_{\mathsf{sm}}(x) - f'_{\mathsf{sm}}(x)\|_\infty \le \delta/2] \le \delta/2$. □

3.3 Converting to PPFs

Before we finish the proof of Lemma 18, we will need to make a couple of elementary observations. First of all, observe that if α is uniformly random in $[0,1]$, then for any $x \in [0,1]$, $\mathbf{E}[\mathbf{1}_{x-\alpha \ge 0}] = x$. Here $\mathbf{1}_{x-\alpha\ge 0}$ denotes the function which is 1 if $x - \alpha \ge 0$ and 0 otherwise. Now, for any parameter $\eta > 0$, define the distribution Int_η to be uniformly random over the set $\{i \cdot \eta\}_{i \ge 0} \cap [0,1]$. Then, we have the following simple claim.

Claim 32. *Let $\zeta > 0$ and $y \in \Delta_{k,\zeta}$. Then,*

$$\left\| \mathbf{E}\left[\sum_{s=1}^{k} \arg\max(\underbrace{0,\dots,0}_{s-1\ times}, y_s - \alpha_s, \underbrace{0,\dots,0}_{k-s\ times})\right] - y\right\|_1 \le 2(\zeta + k \cdot \eta).$$

Here, the expectation is with respect to $(\alpha_1, \dots, \alpha_k) \sim \mathsf{Int}_\eta^k$

Proof. Let the point closest to y in Δ_k be x. Then, we have $\|x - y\|_1 = \zeta$. We have the following:

$$\left\| \mathbf{E}\left[\sum_{s=1}^{k} \arg\max(\underbrace{0,\dots,0}_{s-1 \text{ times}}, x_s - \alpha_s, \underbrace{0,\dots,0}_{k-s \text{ times}})\right] - x\right\|_1 \le k \cdot \eta.$$

Combining this with $\|x - y\|_1 \le \zeta$, we obtain

(3.5)
$$\left\| \mathbf{E}\left[\sum_{s=1}^{k} \arg\max(\underbrace{0,\dots,0}_{s-1 \text{ times}}, x_s - \alpha_s, \underbrace{0,\dots,0}_{k-s \text{ times}})\right] - y\right\|_1 \le k \cdot \eta + \zeta.$$

Next, for any $1 \le s \le k$,

$$\|\mathbf{E} \arg\max(\underbrace{0,\dots,0}_{s-1 \text{ times}}, x_s - \alpha_s, \underbrace{0,\dots,0}_{k-s \text{ times}}) - \arg\max(\underbrace{0,\dots,0}_{s-1 \text{ times}}, y_s - \alpha_s, \underbrace{0,\dots,0}_{k-s \text{ times}})\|_1 \le |x_s - y_s| + \eta.$$

Summing over all $1 \le s \le k$ and combining with (3.5), we obtain the claim. □

Proof of Lemma 18: For $1 \le i \le \ell$, let $\{f_{\mathsf{sm}}^{'(i)}\}$ and $\{g_{\mathsf{sm}}^{'(i)}\}$ be the functions obtained by applying Corollary 27 and Lemma 31 to the family of functions $\{f^{(i)}\}$ and $\{g^{(i)}\}$. In particular, let $f_{\mathsf{sm}}^{'(i)} = (p_{f,1}^{'(i)}, \dots, p_{f,k}^{'(i)})$ and $g_{\mathsf{sm}}^{'(i)} = (p_{g,1}^{'(i)}, \dots, p_{g,k}^{'(i)})$. For $\eta > 0$ (to be fixed later), let us define $f_1^{(i)}$ and $g_1^{(i)}$ as follows:

$$f_1^{(i)} = \sum_{s=1}^{k} \mathbf{E} \arg\max\,(\underbrace{0,\dots,0}_{s-1 \text{ times}}, p_{f,s}^{'(i)} - \alpha_s, \underbrace{0,\dots,0}_{k-s \text{ times}})$$

$$g_1^{(i)} = \sum_{s=1}^{k} \mathbf{E} \arg\max\,(\underbrace{0,\dots,0}_{s-1 \text{ times}}, p_{g,s}^{'(i)} - \alpha_s, \underbrace{0,\dots,0}_{k-s \text{ times}})$$

We will now verify the properties of the construction.

Proof of Items 1 and 2: Both these items are straight forward from the construction.

Proof of Item 3: By the second item of Lemma 31, we have $\Pr_{x\sim\gamma_n}[f_{\mathsf{sm}}^{'(i)}(x) \in \Delta_{k,k\delta/4}] \ge 1 - \delta/2$. By applying Claim 32, we obtain that whenever $f_{\mathsf{sm}}^{'(i)}(x) \in \Delta_{k,k\delta/4}$, $f_1^{(i)}(x) \in \Delta_{k,O(k\delta+k\eta)}$. Thus, as long as $\eta \le \delta/k$, this proves Item 3 for $f_1^{(i)}$. The proof for $g_1^{(i)}$ is similar.

Proof of Items 4 and 5: We first observe that $\Pr_{x\sim\gamma_n}[\|f_{\mathsf{sm}}^{'(i)}(x) - f_{\mathsf{sm}}^{(i)}(x)\|_1 \le k \cdot \delta/4] \ge 1 - \delta/2$. By applying Claim 32, we obtain that $\Pr_{x\sim\gamma_n}[\|f_1^{(i)}(x) - f_{\mathsf{sm}}^{(i)}(x)\|_1 \le O(k\delta + k\eta)] \ge 1 - \delta/2$. However, note that by definition, $\|f_1^{(i)}(x) - f_{\mathsf{sm}}^{(i)}(x)\|_\infty \le k$. This implies that $\mathbf{E}[\|f_{\mathsf{sm}}^{'(i)}(x) - f_1^{(i)}(x)\|_1] = O(k\delta + k\eta)$. As long as $\eta \le \delta/k$, we have $\mathbf{E}[\|f_{\mathsf{sm}}^{(i)}(x) - f_1^{(i)}(x)\|_1] = O(k\delta)$. Combining with the guarantees of Corollary 22 yields Items 4 and 5.

Proof of Item 6: To prove Item 6, note that for any $1 \leq s \leq k$ and $\alpha_s \in [0,1]$,

$$\arg\max\,(\underbrace{0,\ldots,0}_{s-1 \text{ times}}, p^{'(i)}_{f,s} - \alpha_s, \underbrace{0,\ldots,0}_{k-s \text{ times}}) = \mathsf{PPF}_{p^{'(i)}_{f,s}-\alpha_s, s}.$$

Thus, if we define $p^{(i)}_{s,j,1} = p^{'(i)}_{f,s} - \eta \cdot j$ and $p^{(i)}_{s,j,2} = p^{'(i)}_{g,s} - \eta \cdot j$, then

$$f^{(i)}_1 = \sum_{s=1}^{k}\sum_{j=0}^{m} \frac{1}{m} \mathsf{PPF}_{p^{(i)}_{s,j,1},s} \text{ and}$$

$$g^{(i)}_1 = \sum_{s=1}^{k}\sum_{j=0}^{m} \frac{1}{m} \mathsf{PPF}_{p^{(i)}_{s,j,2},s},$$

where $m = \lceil 1/\eta \rceil$. As $\eta \leq \delta/k$, $m = O(k/\delta)$. By Lemma 31, $\mathsf{deg}(p^{'(i)}_{f,s})$ and $\mathsf{deg}(p^{'(i)}_{g,s})$ is at most $d' = d \cdot \mathsf{poly}(k/\delta) \cdot \log^d(dk/\delta)$ where $d = 2/t \cdot \log(dk/\delta)$ (coming from Corollary 22). If we set $d_0(t,k,\delta) = d'$, then $\mathsf{deg}(p^{'(i)}_{f,s})$ and $\mathsf{deg}(p^{'(i)}_{g,s})$ is at most $d_0(t,k,\delta)$. As $\mathsf{deg}(p^{(i)}_{s,j,1}) = \mathsf{deg}(p^{'(i)}_{f,s})$ and $\mathsf{deg}(p^{(i)}_{s,j,2}) = \mathsf{deg}(p^{'(i)}_{g,s})$, this proves Item 6. (We can make the PPFs balanced by applying Fact 17). □

4 Construction of junta polynomials

This section is dedicated to the proof of Lemma 19. To prove this lemma, we will first recall the following important result from [DMN17] (Theorem 41 in that paper).

Theorem 33. *Let $p_1, \ldots, p_\ell : \mathbb{R}^n \to \mathbb{R}$ be degree-d polynomials and for $\delta > 0$, the following two conditions: (i) For all $1 \leq s \leq \ell$, $\mathsf{Var}(p_s) = 1$ and (ii) For all $1 \leq s \leq \ell$, $|\mathbf{E}[p_s]| \leq \log^{d/2}(k \cdot d/\delta)$. For $1 \leq s \leq \ell$ and $t > 0$, define $u_s : \mathbb{R}^{2n} \to \mathbb{R}$ as follows: $u_s(x,y) = p_s(e^{-t}x + \sqrt{1-e^{-2t}}y)$. Then, there is an explicitly computable $n_0 = n_0(\ell, d, \xi)$ and polynomials $r_1, \ldots, r_\ell : \mathbb{R}^{n_0} \to \mathbb{R}$ with the following properties: For $1 \leq s \leq \ell$, define $v_s : \mathbb{R}^{2n_0} \to \mathbb{R}$ as $v_s(x,y) = r_s(e^{-t}x + \sqrt{1-e^{-2t}}y)$. Then, for $1 \leq s, s' \leq \ell$,*

1. $|\Pr_{x\sim\gamma_n}[p_s \geq 0] - \Pr_{x\sim\gamma_n}[r_s \geq 0]| \leq \xi$.
2. $|\Pr_{x,y\sim\gamma_n}[u_s \geq 0] - \Pr_{x,y\sim\gamma_{n_0}}[v_s \geq 0]| \leq \xi$.
3. $|\Pr_{x\sim\gamma_n}[p_s \cdot p_{s'} \geq 0] - \Pr_{x\sim\gamma_{n_0}}[r_s \cdot r_{s'} \geq 0]| \leq \xi$.
4. $|\Pr_{x,y\sim\gamma_n}[u_s \cdot u_{s'} \geq 0] - \Pr_{x,y\sim\gamma_{n_0}}[v_s \cdot v_{s'} \geq 0]| \leq \xi$.
5. $|\Pr_{x,y\sim\gamma_n}[p_s \cdot u_{s'} \geq 0] - \Pr_{x,y\sim\gamma_{n_0}}[v_s \cdot v_{s'} \geq 0]| \leq \xi$.

We now derive an additional property of the polynomials $\{p_s\}_{1\leq s\leq\ell}$ and $\{r_s\}_{1\leq s\leq\ell}$ defined in Theorem 33 which will be useful later.

Corollary 34. *Let $p_1, \ldots, p_\ell : \mathbb{R}^n \to \mathbb{R}$ and $u_1, \ldots, u_\ell : \mathbb{R}^n \to \mathbb{R}$ be as defined in Theorem 33. Then, for any $1 \leq s, s' \leq k$,*

$$\begin{aligned}|&\Pr_{x\sim\gamma_n}[(p_s(x) \geq 0) \wedge (p_{s'}(x) \geq 0)] \\ &- \Pr_{x\sim\gamma_{n_0}}[(r_s(x) \geq 0) \wedge (r_{s'}(x) \geq 0)]| \leq 2\xi.\end{aligned}$$

Proof. The main observation here is that if $A, B \neq 0$, then

$$\mathbf{1}[A \geq 0] \cdot \mathbf{1}[B \geq 0] = \frac{1}{2}(\mathbf{1}[A \cdot B \geq 0] + \mathbf{1}[A \geq 0] + \mathbf{1}[B \geq 0] - 1).$$

Now, note that because p_s, $p_{s'}$, r_s and $r_{s'}$ are degree-d polynomials, any of these functions vanish over the Gaussian measure with probability 0. Thus,

$$\begin{aligned}&\Pr_{x\sim\gamma_n}[(p_s(x) \geq 0) \wedge (p_{s'}(x) \geq 0)] \\ &= \frac{1}{2}(\Pr_{x\sim\gamma_n}[p_s(x) \geq 0] + \Pr_{x\sim\gamma_n}[p_{s'}(x) \geq 0] \\ &+ \Pr_{x\sim\gamma_n}[p_s \cdot p_{s'}(x) \geq 0] - 1)\end{aligned}$$

$$\begin{aligned}&\Pr_{x\sim\gamma_{n_0}}[(r_s(x) \geq 0) \wedge (r_{s'}(x) \geq 0)] \\ &= \frac{1}{2}(\Pr_{x\sim\gamma_{n_0}}[r_s(x) \geq 0] + \Pr_{x\sim\gamma_{n_0}}[r_{s'}(x) \geq 0] \\ &+ \Pr_{x\sim\gamma_{n_0}}[r_s \cdot r_{s'}(x) \geq 0] - 1)\end{aligned}$$

Combining the above equations with items 1 and 3 in Theorem 33 yields the corollary. □

We now describe the proof of Lemma 19.

Proof of Lemma 19: Let us consider the collection of degree-d_0 polynomials $\{p^{(i)}_{s,j,1}\}_{1\leq i\leq\ell, 1\leq s\leq k, 1\leq j\leq m} \cup \{p^{(i)}_{s,j,2}\}_{1\leq i\leq\ell, 1\leq s\leq k, 1\leq j\leq m}$. We now apply Theorem 33 to obtain polynomials $\{r^{(i)}_{s,j,1}\}_{1\leq i\leq\ell, 1\leq s\leq k, 1\leq j\leq m} \cup \{r^{(i)}_{s,j,2}\}_{1\leq i\leq\ell, 1\leq s\leq k, 1\leq j\leq m}$ with $\xi = \delta/(40k^2)$. We now define

$$f^{(i)}_{\mathsf{junta}} = \sum_{s=1}^{k}\sum_{j=1}^{m} \frac{1}{m} \cdot \mathsf{PPF}_{r^{(i)}_{s,j,1},s}(x)\ ,\ g^{(i)}_{\mathsf{junta}} = \sum_{s=1}^{k}\sum_{j=1}^{m} \frac{1}{m} \cdot \mathsf{PPF}_{r^{(i)}_{s,j,2},s}(x)$$

We now verify the properties of the construction.

Proof of Item 1: Observe that for $1 \leq s \leq k$, we have the following

$$\mathbf{E}[(f_1^{(i)}(x))_s] = \sum_{j=1}^{m} \frac{1}{m} \cdot \mathbf{E}_x[\mathsf{PPF}_{p_{s,j,1}^{(i)},s}(x)] = \sum_{j=1}^{m} \frac{1}{m} \cdot \Pr_x[p_{s,j,1}^{(i)}(x) \geq 0]$$

$$\mathbf{E}[(f_{\mathsf{junta}}^{(i)}(x))_s] = \sum_{j=1}^{m} \frac{1}{m} \cdot \mathbf{E}_x[\mathsf{PPF}_{r_{s,j,1}^{(i)},s}(x)] = \sum_{j=1}^{m} \frac{1}{m} \cdot \Pr_x[r_{s,j,1}^{(i)}(x) \geq 0]$$

Thus, we obtain

$$|\mathbf{E}[(f_1^{(i)}(x))_s] - \mathbf{E}[(f_{\mathsf{junta}}^{(i)}(x))_s]| \leq \sup_{1\leq j\leq m} |\Pr_x[p_{s,j,1}^{(i)}(x) \geq 0] - \Pr_x[r_{s,j,1}^{(i)}(x) \geq 0]| \leq \xi.$$

The penultimate inequality follows by applying Theorem 33 to $p_{s,j,1}^{(i)}$ and $r_{s,j,1}^{(i)}$. This immediately implies that $\|\mathbf{E}[f_1^{(i)}(x)] - \mathbf{E}[f_{\mathsf{junta}}^{(i)}(x)]\|_1 \leq k \cdot \xi \leq \delta$. The proof for $|\mathbf{E}[(g_1^{(i)}(x))_s] - \mathbf{E}[(g_{\mathsf{junta}}^{(i)}(x))_s]| \leq \delta$. is exactly identical.

Proof of Item 2: Like Item 1, we will only prove that $\Pr_x[f_{\mathsf{junta}}^{(i)}(x) \in \Delta_{k,\sqrt{\delta}}] \leq \sqrt{\delta}$. The proof for $\Pr_x[g_{\mathsf{junta}}^{(i)}(x) \in \Delta_{k,\sqrt{\delta}}] \leq \sqrt{\delta}$. To prove this, we first observe that for all x both $f_1^{(i)}(x)$ and $f_{\mathsf{junta}}^{(i)}(x)$ always lie in the positive orthant and secondly, $\|f_1^{(i)}(x)\|_\infty, \|f_{\mathsf{junta}}^{(i)}(x)\|_\infty \leq 1$. Next,

$$\mathbf{E}[(\|f_1^{(i)}(x)\|_1 - 1)^2] \leq \Pr_x[f_1^{(i)}(x) \in \Delta_{k,\delta}] \cdot \delta^2 + \Pr_x[f_1^{(i)}(x) \notin \Delta_{k,\delta}] \cdot k^2 \leq \delta^2 + k^2 \cdot \delta. \tag{4.6}$$

The first inequality uses $\sup_x \|f_1^{(i)}(x)\|_1 \leq k$ and the second inequality uses $\Pr_x[f_1^{(i)}(x) \notin \Delta_{k,\delta}] \leq \delta$. Next, observe that

$$\|f_1^{(i)}(x)\|_1 = \sum_{s=1}^{k} \sum_{j=1}^{m} \frac{1}{m} \cdot \mathbf{1}[p_{s,j,1}^{(i)}(x) \geq 0],$$

$$\|f_{\mathsf{junta}}^{(i)}(x)\|_1 = \sum_{s=1}^{k} \sum_{j=1}^{m} \frac{1}{m} \cdot \mathbf{1}[r_{s,j,1}^{(1)}(x) \geq 0]$$

This implies

$$\begin{aligned} &(\|f_1^{(i)}(x)\|_1 - 1)^2 \\ = \quad & \sum_{s=1}^{k} \sum_{s'=1}^{k} \sum_{j=1}^{m} \sum_{j'=1}^{m} \frac{1}{m^2} \mathbf{1}[p_{s,j,1}^{(i)}(x) \geq 0] \cdot \mathbf{1}[p_{s',j',1}^{(i)}(x) \geq 0] + 1 \\ - \quad & \frac{2}{m} \sum_{s=1}^{k} \sum_{j=1}^{m} \mathbf{1}[p_{s,j,1}^{(i)}(x) \geq 0]. \end{aligned} \tag{4.7}$$

$$\begin{aligned} &(\|f_{\mathsf{junta}}^{(i)}(x)\|_1 - 1)^2 \\ = \quad & \sum_{s,s'=1}^{k} \sum_{j,j'=1}^{m} \frac{1}{m^2} \mathbf{1}[r_{s,j,1}^{(i)}(x) \geq 0] \cdot \mathbf{1}[r_{s',j',1}^{(i)}(x) \geq 0] + 1 \\ - \quad & \frac{2}{m} \sum_{s=1}^{k} \sum_{j=1}^{m} \mathbf{1}[r_{s,j,1}^{(i)}(x) \geq 0]. \end{aligned} \tag{4.8}$$

Recall that by construction, we have

$$\sup_{1\leq s\leq k,\ 1\leq j\leq m} |\Pr_x[p_{s,j,1}^{(i)}(x) \geq 0] - \Pr_x[r_{s,j,1}^{(i)}(x) \geq 0]| \leq \xi \tag{4.9}$$

Applying Corollary 34, we also obtain

$$\begin{aligned} &\sup_{1\leq s,s'\leq k,\ 1\leq j,j'\leq m} |\Pr_x[(p_{s,j,1}^{(i)}(x) \geq 0) \wedge (p_{s',j',1}^{(i)}(x) \geq 0)] \\ &- \Pr_x[(r_{s,j,1}^{(i)}(x) \geq 0) \wedge (r_{s',j',1}^{(i)}(x) \geq 0)]| \leq 2\xi. \end{aligned} \tag{4.10}$$

Applying (4.9) and (4.10) to (4.7) and (4.8), we obtain

$$|\mathbf{E}[(\|f_{\mathsf{junta}}^{(i)}(x)\|_1 - 1)^2] - \mathbf{E}[(\|f_1^{(i)}(x)\|_1 - 1)^2]| \leq 2k^2 \cdot \xi + 2k \cdot \xi \leq \delta.$$

Combining this with (4.6), we obtain $\mathbf{E}[(\|f_{\mathsf{junta}}^{(i)}(x)\|_1 - 1)^2] \leq 2k^2 \cdot \delta$. Applying Markov's inequality, we obtain that $\Pr[|\ \|f_{\mathsf{junta}}^{(i)}(x)\|_1 - 1| > k\sqrt{\delta}] \leq 2k\sqrt{\delta}$. Since $f_{\mathsf{junta}}^{(i)}(x)$ lies in the positive orthant for any x, this proves Item 2.

Proof of Item 3: To prove Item 3, we observe that for any $1 \le s_1, s_2 \le k$,

$$
\begin{aligned}
(4.11)\quad & \mathbf{E}[f_{1,s_1} P_t g_{1,s_2}] \\
= \quad & \frac{1}{m^2} \sum_{j=1}^{m} \sum_{j'=1}^{m} \mathbf{E}[\mathsf{PPF}_{p^{(1)}_{s_1,j_1}}(x) P_t \ \mathsf{PPF}_{p^{(2)}_{s_2,j_2}}(x)] \\
= \quad & \frac{1}{m^2} \sum_{j=1}^{m} \sum_{j'=1}^{m} \mathbf{E}_{x,y}[\mathsf{PPF}_{p^{(1)}_{s_1,j_1}}(x) \mathsf{PPF}_{p^{(2)}_{s_2,j_2}}(z)] \\
= \quad & \frac{1}{m^2} \sum_{j=1}^{m} \sum_{j'=1}^{m} \Pr_{x,y}[(p^{(1)}_{s_1,j_1}(x) \ge 0) \wedge (p^{(2)}_{s_2,j_2}(z) \ge 0)] \\
= \quad & \frac{1}{m^2} \sum_{j=1}^{m} \sum_{j'=1}^{m} \Pr_{x,y}[(p^{(1)}_{s_1,j_1}(x) \ge 0) \wedge (u^{(2)}_{s_2,j_2}(z) \ge 0)].
\end{aligned}
$$

In the above, $z = e^{-t}x + \sqrt{1 - e^{-2t}}y$. Likewise, we can obtain

$$
\begin{aligned}
(4.12)\quad & \\
& \mathbf{E}[f_{\mathsf{junta},s_1} P_t g_{\mathsf{junta},s_2}] \\
& = \frac{1}{m^2} \sum_{j=1}^{m} \sum_{j'=1}^{m} \Pr_{x,y}[(r^{(1)}_{s_1,j_1}(x) \ge 0) \wedge (v^{(2)}_{s_2,j_2}(z) \ge 0)].
\end{aligned}
$$

Combining (4.11) and (4.12) with Item 5 in Theorem 33 yields

$$
|\mathbf{E}[f_{1,s_1} P_t g_{1,s_2}] - \mathbf{E}[f_{\mathsf{junta},s_1} P_t g_{\mathsf{junta},s_2}]| \le \xi.
$$

This finishes the proof. □

References

[AC93] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography - I: Secret sharing. *IEEE Trans. Information Theory*, 39(4):1121–1132, 1993.

[AC98] R. Ahlswede and I. Csiszár. Common Randomness in Information Theory and Cryptography - Part II: CR Capacity. *IEEE Trans. Information Theory*, 44(1):225–240, 1998.

[AL06] N. Alon and E. Lubetzky. The shannon capacity of a graph and the independence numbers of its powers. *IEEE Transactions on Information Theory*, 52(5):2172–2176, 2006.

[BGI14] M. Bavarian, D. Gavinsky, and T. Ito. On the Role of Shared Randomness in Simultaneous Communication. In *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014*, pages 150–162, 2014.

[Bor85] C. Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Probability Theory and Related fields*, 70:1–13, 1985.

[CG59] Ward Cheney and Allen A Goldstein. Proximity maps for convex sets. *Proceedings of the American Mathematical Society*, 10(3):448–450, 1959.

[CGMS15] C. Canonne, V. Guruswami, R. Meka, and M. Sudan. Communication with Imperfectly Shared Randomness. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pages 257–262, 2015.

[CN00] I. Csiszár and P. Narayan. Common randomness and secret key generation with a helper. *IEEE Trans. Information Theory*, 46(2):344–366, 2000.

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

[DDFS14] A. De, I. Diakonikolas, V. Feldman, and R. Servedio. Near-optimal solutions for the Chow Parameters Problem and low-weight approximation of halfspaces. *Journal of the ACM*, 61(2), 2014.

[DDS14] Anindya De, Ilias Diakonikolas, and Rocco A. Servedio. Deterministic approximate counting for juntas of degree-2 polynomial threshold functions. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 229–240, 2014.

[DLTW08] A. Doherty, Y. Liang, B. Toner, and S. Wehner. The quantum moment problem and bounds on entangled multi-prover games. In *Computational Complexity, 2008. CCC'08. 23rd Annual IEEE Conference on*, pages 199–210. IEEE, 2008.

[DMN17] A. De, E. Mossel, and J. Neeman. Noise stability is computable and low-dimensional. In *Conference on Computational Complexity*, pages 28:1–28:10, 2017. Full version available at https://arxiv.org/abs/1701.01483.

[DS14] A. De and R. Servedio. Efficient deterministic approximate counting for low-degree polynomial threshold functions. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA*, pages 832–841, 2014. Full version at http://arxiv.org/abs/1311.7178.

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

[FK99] A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[Fre95] Y. Freund. Boosting a weak learning algorithm by majority. *Information and Computation*, 121(2):256–285, 1995.

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

[GKR17] B. Ghazi, P. Kamath, and P. Raghavendra. Dimension Reduction for Polynomials over Gaussian Space and Applications. *arXiv preprint arXiv:1708.03808*, 2017.

[GKS16a] B. Ghazi, P. Kamath, and M. Sudan. Communication Complexity of Permutation-Invariant Functions. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms,* *SODA 2016*, pages 1902–1921, 2016.

[GKS16b] B. Ghazi, P. Kamath, and M. Sudan. Decidability of Non-Interactive Simulation of Joint Distributions. In *56th Annual IEEE Symposium on Foundations of Computer Science*, pages 545–554, 2016.

[Hae79] W. Haemers. On some problems of Lovász concerning the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25(2):231–232, 1979.

[Imp95] R. Impagliazzo. Hard-Core Distributions for Somewhat Hard Problems. In *36th Annual Symposium on Foundations of Computer Science*, pages 538–545, 1995.

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

[KA16] S. Kamath and V. Anantharam. On non-interactive simulation of joint distributions. *IEEE Transactions on Information Theory*, 62(6):3419–3435, 2016.

[KKM$^+$11] J. Kempe, H. Kobayashi, K. Matsumoto, B. Toner, and T. Vidick. Entangled games are hard to approximate. *SIAM Journal on Computing*, 40(3):848–877, 2011.

[KNOW14] P. Kothari, A. Nayyeri, R. O'Donnell, and C. Wu. Testing surface area. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1204–1214, 2014.

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

[LRS15] J. Lee, P. Raghavendra, and D. Steurer. Lower Bounds on the Size of Semidefinite Programming Relaxations. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015*, pages 567–576, 2015.

[MO05] E. Mossel and R. O'Donnell. Coin flipping from a cosmic source: On error correction of truly random bits. *Random Structures Algorithms*, 26(4):418–436, 2005.

[MOO10] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Ann. Math.*, 171(1):295–341, 2010.

[MOR$^+$06] E. Mossel, R. O'Donnell, O. Regev, J. Steif, and B. Sudakov. Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality. *Israel J. Math.*, 154:299–336, 2006.

[Mos10] E. Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010.

[Nee14] J. Neeman. Testing surface area with arbitrary accuracy. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 393–397, 2014.

[O'D14] R. O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, Cambridge, 2014.

[Reg16] O. Regev. . Personal communication, 2016.

[Sch90] R. Schapire. The strength of weak learnability. *Machine Learning*, 5(2):197–227, 1990.

[Tao07] T. Tao. Structure and randomness in combinatorics. In *Proc. 48th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2007.

[TTV09] Luca Trevisan, Madhur Tulsiani, and Salil P. Vadhan. Regularity, Boosting, and Efficiently Simulating Every High-Entropy Distribution. In *IEEE Conference on Computational Complexity*, pages 126–136, 2009.

[Wit75] H. S. Witsenhausen. On Sequences of Pairs of Dependent Random Variables. *SIAM Journal on Applied Mathematics*, 28(1):100–113, 1975.

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

A Reduction from arbitrary P to the Gaussian case

We first restate Theorem 3 below.

Theorem. *3 Suppose there exist $f, g : \mathcal{Z}^n \to [k]$ such that $(f(\mathbf{X}^n), g(\mathbf{Y}^n)) \sim \mathbf{Q}$. Then, there exist $n_0 = n_0(|\mathbf{P}|, \delta)$ and $f_{\mathsf{junta}}, g_{\mathsf{junta}} : \mathcal{Z}^{n_0} \to [k]$ such that $\mathbf{Q}$ and the distribution of $(f_{\mathsf{junta}}(\mathbf{X}^{n_0}), g_{\mathsf{junta}}(\mathbf{Y}^{n_0}))$ are δ-close in total variation distance. Moreover, n_0 is computable. Further, the functions f_{junta} and g_{junta} can be explicitly computed.*

Next, we restate Theorem 7.

Theorem. *7 Let $\mathbf{P} = (\mathbf{X}, \mathbf{Y}) = \mathbf{G}_{\rho,2}$ and let $f^{(1)}, \ldots, f^{(\ell)} : \mathbb{R}^n \to [k]$ and $g^{(1)}, \ldots, g^{(\ell)} : \mathbb{R}^n \to [k]$ where we define $\mathbf{Q}_{i,j}$ as $\mathbf{Q}_{i,j} = (f^{(i)}(\mathbf{X}^n), g^{(j)}(\mathbf{Y}^n))$. Then, for every $\delta > 0$, there is an explicitly defined constant $n_0 = n_0(\ell, k, \delta)$ and explicitly defined functions $f^{(1)}_{\mathsf{junta}}, \ldots, f^{(\ell)}_{\mathsf{junta}} : \mathbb{R}^{n_0} \to [k]$ and $g^{(1)}_{\mathsf{junta}}, \ldots, g^{(\ell)}_{\mathsf{junta}} : \mathbb{R}^{n_0} \to [k]$ such that for every $1 \leq i, j \leq \ell$, $\mathrm{d}_{\mathrm{TV}}((f^{(i)}_{\mathsf{junta}}(\mathbf{X}^{n_0}), g^{(j)}_{\mathsf{junta}}(\mathbf{Y}^{n_0})), \mathbf{Q}_{i,j}) \leq \delta$.*

The main purpose of this section is to show how proving Theorem 3 reduces to proving Theorem 7. While the reduction essentially follows just going over the steps in [GKS16b] *mutatis mutandis* (which in turn relies on standard tools from Boolean function analysis), for the purpose of clarity, we give a brief overview of the reduction here.

First, let us fix some notation.

1. We recall the notion of maximal correlation coefficient: Namely, given a probability space $(\mathbf{X}, \mathbf{Y})$, we let $\rho(\mathbf{X}, \mathbf{Y})$ be defined as

$$\rho(\mathbf{X}, \mathbf{Y}) = \sup \mathbf{E}[\Psi_1(\mathbf{X}) \cdot \Psi_2(\mathbf{Y})],$$

where the supremum is taken over all functions which satisfy $\mathbf{E}[\Psi_1(\mathbf{X})] = \mathbf{E}[\Psi_2(\mathbf{Y})] = 0$ and $\mathsf{Var}[\Psi_1(\mathbf{X})] = \mathsf{Var}[\Psi_2(\mathbf{Y})] = 1$.

2. For a given set $H \subseteq [n]$, $x_H \in \mathbf{X}^{|H|}$ and function $f : \mathbf{X}^n \to \mathbb{R}^k$, we let $f(x_H, .) : \mathbf{X}^{[n]\setminus H} \to \mathbb{R}^k$ denote the function obtained by fixing the coordinates of f in H to x_H.

As we have stated before, for the case $k = 2$, Ghazi, Kamath and Sudan [GKS16b] reduce Theorem 3 for the general $\mathbf{P}$ case to the case when $\mathbf{P} = \mathbf{G}_{\rho,2}$. In other words, for $k = 2$, [GKS16b] reduces Theorem 3 for the general $\mathbf{P}$ case to Theorem 7 with $\ell = 1$. We now give a sketch of why Theorem 3 reduces to Theorem 7 for $k > 1$.

Overview of the reduction: Using the regularity lemma for low-degree polynomials [DSTW10, DDS14] and other ideas from Boolean function analysis (along the lines of [GKS16b]), one can easily show the following: Let $\tau > 0$ be any error parameter. Then, there exists a set $H \subseteq [n]$ such that $|H| = O_{\tau,|\mathbf{P}|,k}(1)$ and for $(x_H, y_H) \sim (\mathbf{X}, \mathbf{Y})^H$, with probability $1 - \tau$, the following holds: The functions $f(x_H, \cdot)$ and $g(y_H, \cdot)$ are *low-influence* functions namely,

$$\max_{i \in [n]\setminus H} \mathsf{Inf}_i(f(x_H, \cdot)) \le \tau, \quad \max_{i \in [n]\setminus H} \mathsf{Inf}_i(g(y_H, \cdot)) \le \tau.$$

In the above definition, for $f : \mathbb{R}^n \to \mathbb{R}^k$, we let $\mathsf{Inf}_i(f)$ denotes the quantity

$$\mathsf{Inf}_i(f) = \sum_{i \in S : S \in \mathbb{Z}^{*n}} \|\widehat{f}(S)\|_2^2,$$

where $\widehat{f}(S)$ denotes the Hermite coefficient of f corresponding to S. Note that this is the standard definition of "influence" from Boolean function analysis (see [O'D14, Mos10]). In fact, one can also additionally assume that every coordinate of f and g is essentially a low-degree polynomial.

To understand why the low-influence condition is useful, let $\mathbf{P}_G = \mathbf{G}_{\rho,2}$ where $\rho = \rho(\mathbf{X}, \mathbf{Y})$. Further, let $(\mathbf{X}_G, \mathbf{Y}_G) = \mathbf{P}_G$. Likewise, let $\tilde{f}(x_H, \cdot)$ (resp. $\tilde{g}(y_H, \cdot)$) be the multilinear extension of $f(x_H, \cdot)$ (resp. $g(y_H, \cdot)$) to the Gaussian space. Then, the invariance principle of Mossel *et al.* [MOO10, Mos10] shows that as long as τ is chosen to be sufficiently small in δ, for any pair (x_H, y_H) where $f(x_H, \cdot)$ and $g(y_H, \cdot)$ are low-influence functions, the following holds:

$$\begin{aligned}\mathrm{d_{TV}}((&\tilde{f}(x_H, \mathbf{X}_G^{[n]\setminus H}), \tilde{g}(y_H, \mathbf{Y}_G^{[n]\setminus H})),\\ &(f(x_H, \mathbf{X}^{[n]\setminus H}), f(y_H, \mathbf{Y}^{[n]\setminus H})) \le \delta/4.\end{aligned}$$

Note that the total number of (x_H, y_H) pairs is bounded by $|\mathsf{supp}(\mathbf{P})|^{2|H|}$. Let us denote this number by $\mathbf{N}_{sup}$. By applying Theorem 7,we obtain that for any $\delta > 0$, there is $n_0 = n_0(\mathbf{N}_{sup}, k, \delta)$ such that corresponding to every function $\tilde{f}(x_H, \cdot)$ (resp. $\tilde{g}(y_H, \cdot)$), there is a function $\underline{f}_{x_H} : \mathbb{R}^{n_0} \to [k]$ (resp. $\underline{g}_{y_H} : \mathbb{R}^{n_0} \to [k]$) such that

$$\begin{aligned}\mathrm{d_{TV}}((&\tilde{f}(x_H, \mathbf{X}_G^{[n]\setminus H}), \tilde{g}(y_H, \mathbf{Y}_G^{[n]\setminus H}),\\ &(\underline{f}_{x_H}(\mathbf{X}_G^{n_0}), \underline{g}_{y_H}(\mathbf{Y}_G^{n_0}))) \le \delta/4.\end{aligned}$$

Note that here we are crucially using the fact that Theorem 7 is valid for an arbitrary $\ell \ge 1$ and not just $\ell = 1$. Let us define $m_0 = n_0 \cdot (1/\kappa^2)$. We next define $\underline{f}_{\mathsf{low},\ x_H} : \mathbb{R}^{m_0} \to [k]$ as

$$\underline{f}_{\mathsf{low},\ x_H}(x_{1,1}, \ldots, x_{n_0,\kappa^{-2}}) = \underline{f}_{x_H}(\kappa \cdot z_1, \ldots, \kappa \cdot z_{n_0}).$$

Here $z_j = (x_{j,1} + \ldots + x_{j,\kappa^{-2}})$ for $1 \le j \le n_0$. Likewise, letting $w_j = (y_{j,1} + \ldots + x_{y,\kappa^{-2}})$, we define,

$$\underline{g}_{\mathsf{low},\ y_H}(y_{1,1}, \ldots, y_{n_0,\kappa^{-2}}) = \underline{g}_{y_H}(\kappa \cdot w_1, \ldots, \kappa \cdot w_{n_0}).$$

From the definition of $\underline{f}_{x_H}$ and $\underline{g}_{y_H}$, it easily follows that,

$$(\underline{f}_{x_H}(\mathbf{X}_G^{n_0}), \underline{g}_{y_H}(\mathbf{Y}_G^{n_0})) = (\underline{f}_{\mathsf{low},x_H}(\mathbf{X}_G^{m_0}), \underline{g}_{\mathsf{low},y_H}(\mathbf{Y}_G^{m_0}))$$

Let f_{low,x_H} and g_{low,y_H} denote the multilinear extensions of $\underline{f}_{\mathsf{low},x_H}$ and $\underline{g}_{\mathsf{low},y_H}$ to the space $(\mathbf{X}^{m_0}, \mathbf{Y}^{m_0})$. Observe that the functions $\underline{f}_{\mathsf{low},x_H}$ and $\underline{g}_{\mathsf{low},y_H}$ have influence bounded by κ. Thus, as long as κ is chosen to be a sufficiently small function of δ, the invariance principle [Mos10] implies that

$$\begin{aligned}\mathrm{d_{TV}}((&f_{\mathsf{low},x_H}(\mathbf{X}^{m_0}), g_{\mathsf{low},y_H}(\mathbf{Y}^{m_0})),\\ &(\underline{f}_{\mathsf{low},x_H}(\mathbf{X}_G^{m_0}), \underline{g}_{\mathsf{low},y_H}(\mathbf{Y}_G^{m_0}))) \le \delta/4.\end{aligned}$$

Combining the above three equations, we get that

$$\begin{aligned}\mathrm{d_{TV}}((&f_{\mathsf{low},x_H}(\mathbf{X}^{m_0}), g_{\mathsf{low},y_H}(\mathbf{Y}^{m_0})),\\ &(f(x_H, \mathbf{X}^{[n]\setminus H}), g(y_H, \mathbf{Y}^{[n]\setminus H}))) \le \frac{3\delta}{4}.\end{aligned}$$

With this, we define functions $f_{\mathsf{junta}} : \mathbb{R}^{m_0+|H|} \to [k]$ and $g_{\mathsf{junta}} : \mathbb{R}^{m_0+|H|} \to [k]$ as follows. Split $x \in \mathbb{R}^{m_0+H}$ as (x_H, x_{m_0}) and $y \in \mathbb{R}^{m_0+H}$ as (y_H, y_{m_0}).

$$\begin{aligned}f_{\mathsf{junta}}(x_H, x_{m_0}) &= f_{\mathsf{low},x_H}(x_{m_0});\\ g_{\mathsf{junta}}(y_H, y_{m_0}) &= g_{\mathsf{low},y_H}(y_{m_0}).\end{aligned}$$

This immediately implies

$$\begin{aligned}\mathrm{d_{TV}}((&f(\mathbf{X}^n), g(\mathbf{Y}^n)),\\ &(f_{\mathsf{junta}}(\mathbf{X}^{m_0+|H|}), g_{\mathsf{junta}}(\mathbf{Y}^{m_0+|H|})) \le \frac{3\delta}{4} + \tau.\end{aligned}$$

Once we choose $\tau \le \delta/4$, the reduction is complete.