

Decidability of Non-Interactive Simulation of Joint Distributions

Badih Ghazi*

Pritish Kamath†

Madhu Sudan‡

July 14, 2016

Abstract

We present decidability results for a sub-class of “non-interactive” simulation problems, a well-studied class of problems in information theory. A *non-interactive simulation* problem is specified by two distributions $P(x, y)$ and $Q(u, v)$: The goal is to determine if two players, Alice and Bob, that observe sequences X^n and Y^n respectively where $\{(X_i, Y_i)\}_{i=1}^n$ are drawn i.i.d. from $P(x, y)$ can generate pairs U and V respectively (without communicating with each other) with a joint distribution that is arbitrarily close in total variation to $Q(u, v)$. Even when P and Q are extremely simple: e.g., P is uniform on the triples $\{(0, 0), (0, 1), (1, 0)\}$ and Q is a “doubly symmetric binary source”, i.e., U and V are uniform ± 1 variables with correlation say 0.49, it is open if P can simulate Q .

In this work, we show that whenever P is a distribution on a finite domain and Q is a 2×2 distribution, then the non-interactive simulation problem is *decidable*: specifically, given $\delta > 0$ the algorithm runs in time bounded by some function of P and δ and either gives a non-interactive simulation protocol that is δ -close to Q or asserts that no protocol gets $O(\delta)$ -close to Q . The main challenge to such a result is determining explicit (computable) convergence bounds on the number n of samples that need to be drawn from $P(x, y)$ to get δ -close to Q . We invoke contemporary results from the analysis of Boolean functions such as the invariance principle and a regularity lemma to obtain such explicit bounds.

*Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge MA 02139. Supported in part by NSF CCF-1420956, NSF CCF-1420692 and CCF-1217423. badih@mit.edu.

†Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge MA 02139. Supported in part by NSF CCF-1420956 and NSF CCF-1420692. pritish@mit.edu.

‡Harvard John A. Paulson School of Engineering and Applied Sciences. Part of this work was done while at Microsoft Research New England. Supported in part by NSF Award CCF 1565641. madhu@cs.harvard.edu.

Contents

1	Introduction	1
1.1	Proof Overview	3
1.2	Roadmap of the paper	4
2	Preliminaries	4
2.1	Notation	4
2.2	The non-interactive simulation problem	4
2.3	Reformulation of GAP-NON-INT-SIM	6
2.4	Fourier analysis and multi-linear polynomials	7
2.5	Hypercontractivity and moment bounds	9
2.6	Maximal Correlation and Witsenhausen’s rounding	10
2.7	2-dimensional Berry-Esseen theorem	11
3	Main Technical Lemma and Overview	12
3.1	Proof overview	12
3.2	Decidability of GAP-NON-INT-SIM	13
4	Smoothing of Strategies	13
5	Joint Regularity Lemma for Fourier Concentrated Functions	14
5.1	Regularity Lemma for Constant Degree Polynomials	15
5.2	Joint Regularity Lemma	17
6	Applying correlation bounds for low-influence functions	18
7	Simulating Correlated Gaussians	20
8	Putting it all together!	21
8.1	Generalizing to arbitrary binary targets	23
9	Open Questions	24
10	Acknowledgments	24

1 Introduction

Given a sequence of independent samples $(x_1, y_1), (x_2, y_2), \dots$ from a joint distribution P on $\mathcal{A} \times \mathcal{B}$ where Alice observes x_1, x_2, \dots and Bob observes y_1, y_2, \dots , what is the largest correlation that they can extract if Alice applies some function to her observations and Bob applies some function to his? The continuous version of this question – where the extracted correlation is required to be in *Gaussian* form – was solved by Witsenhausen in 1975 who gave (roughly) a $\text{poly}(|\mathcal{A}|, |\mathcal{B}|, \log(1/\delta))$ -time algorithm that estimates the best such correlation up to an additive δ [Wit75]. When the target distribution is Gaussian, the best possible correlation that is attainable is exactly the well-known “maximal correlation coefficient” which was first introduced by Hirschfeld [Hir35] and Gebelein [Geb41] and then studied by Rényi [Rén59]. However, when the target distribution is not Gaussian, the best correlation is not well-understood and this is the question explored in this paper. Specifically, we study the Boolean version of this question where the extracted correlation is required to be in the form of bits with fixed specified marginals. We give an algorithm that, given $\delta > 0$, computes the best such correlation up to an additive δ .

Questions such as the above are well-studied in the information theory literature under the label of “Non-Interactive Simulation”. The roots of this exploration go back to classical works by Gács and Körner [GK73] and Wyner [Wyn75]. In this line of work, the problem is described by a source distribution $P(X, Y)$ and a target distribution $Q(U, V)$ and the goal is to determine the maximum rate at which samples of P can be converted into samples of Q . (So the goal is to start with n samples from P and generate $R \cdot n$ samples from Q , for the largest possible R .) Gács and Körner considered the special case where Q required the output to be a pair of identical uniformly random bits, i.e., $U = V = \text{Ber}(1/2)$ and introduced what is now known as the *Gács-Körner common information* of $P(X, Y)$ to characterize the maximum rate in terms of this quantity. Wyner, on the other hand considered the “inverse” problem where $X = Y = \text{Ber}(1/2)$ and Q was arbitrary. Wyner characterized the best possible conversion rate in this setting in terms of what is now known as the *Wyner common information* of $Q(U, V)$. There is a rich history of subsequent work (see, for instance, [KA15] and the references within) exploring more general settings where neither P nor Q produces identical copies of some random variable. In such settings, even the question of when can the rate be positive is unknown and this is the question we explore in this paper.

The Non-Interactive Simulation problem is also a generalization of the Non-Interactive Correlation Distillation problem which was studied by [MO04, MOR⁺06]¹. Our setup can be thought of as a “positive-rate” version of the setup of Gács and Körner. Namely, for a known source distribution $P(X, Y)$, Alice and Bob are given an arbitrary number of i.i.d. samples and wish to generate *one sample* from the distribution $Q(U, V)$ which is given by $U = V = \text{Ber}(1/2)$. (This is possible if and only if the Gács-Körner rate is positive.)

Motivation. Our motivation for studying the best discrete correlation that can be produced is twofold. On the one hand, this question forms part of the landscape of questions arising from a quest to weaken the assumptions about randomness when it is employed in distributed computing. Computational tasks are often solved well if parties have access to a common source of randomness and there has been recent interest in cryptography [AC93, AC98, BS94, CN00, Mau93, RW05], quantum computing [Nie99, CDS08, DB14] and communication complexity [BGI14, CGMS14, GKS16] to study how the ability to solve these tasks gets affected by weakening the source of randomness. In this space of investigations, it is a very natural question to ask how well one source of randomness can be transformed to a different one, and Non-Interactive Simulation studies exactly this question.

On the other hand, from the analysis point of view, the Non-Interactive Simulation problem forms part of “tensor power” questions that have been challenging to analyze computationally. Specifically, in such questions, the quest is to understand how some quantity behaves as a function of the dimensionality of the problem as the dimension tends to infinity. Notable examples of such problems include the *Shannon capacity of a graph* [Sha56, Lov79] where the goal is to understand how the independence number of the power of a graph behaves as a function of

¹which considered the problem of maximizing agreement on a single bit, in various multi-party settings.

the exponent. Some more closely related examples arise in the problems of local state transformation of quantum entanglement [Bei12, DB13] and the problem of computing the entangled value of a game (see for eg, [KKM⁺11] and also the open problem [ope]). A more recent example is the problem of computing the amortized communication complexity of a communication problem. Braverman-Rao [BR11] showed that this equals the information complexity of the communication problem, however the task of approximating the information complexity was only recently shown to be computable [BS15]. In our case, the best non-interactive simulation to get one pair of correlated bits might require many copies of (x, y) drawn from P and the challenge is to determine how many copies get us close. Convergence results of this type are not obvious. Indeed, the task of approximating the Shannon capacity remains open to this day [AL06]. Our work is motivated in part by the quest to understand tools that can be used to analyze such questions where rate of convergence to the desired quantity is non-trivial to bound.

Estimating Binary Correlations: Previous Work and our Result. In his work generalizing the results of Gács and Körner, Witsenhausen [Wit75] gave an efficient algorithm that achieves a *quadratic* approximation to the Non-Interactive Simulation problem when $Q(U, V)$ is the distribution where U and V are marginally uniform over ± 1 and U is an ρ -correlated copy of V , i.e. $\mathbb{E}[UV] = \rho$ (henceforth, we refer to this distribution as DSBS(ρ)).² Indeed, Witsenhausen introduced the Gaussian correlation problem as an intermediate step to solving this problem and his rounding technique to convert the Gaussian random variables into Boolean ones is essentially the same as that of the Goemans-Williamson algorithm for approximating maximum cut sizes in graphs [GW95]. Already implicit from the work of Witsenhausen is that “maximum correlation” gives a way to upper bound the best achievable ρ when simulating DSBS(ρ). Recent works in the information theory community [KA12, KA15, BG15] enhance the collection of analytical tools that can be used to show stronger impossibility results. While these works produce stronger bounds, they do not necessarily converge to the optimal limit and indeed basic questions about simulation remain open. For instance, till our work, even the following question was open [Kam15]: If P is the uniform distribution on $\{(0, 0), (0, 1), (1, 0)\}$ and $Q = \text{DSBS}(.49)$ (i.e. U, V are uniformly ± 1 , with $\mathbb{E}[UV] = .49$), can P simulate Q arbitrarily well? Our work answers such questions in principle. (Specifically we do give a finite time procedure to approximate the best ρ to within arbitrary accuracy. However, we have not run this algorithm to determine the answer to this specific question.)

Below we state our main theorem informally (see Theorem 2.5 for the formal statement).

Theorem 1.1 (Informal). *There is an algorithm that takes as inputs a source distribution P , a parameter $\rho > 0$ and an error parameter $\delta > 0$, runs in time bounded by some computable function of P , ρ and δ , and either outputs a non-interactive protocol that simulates DSBS(ρ) up to additive δ in total variation distance, or asserts that there is no protocol that gets $O(\delta)$ -close to DSBS(ρ) in total variation distance.*

More generally, the proof techniques extend to deciding the non-interactive simulation problem for an arbitrary 2×2 target distribution. In particular, we also show the following (see Theorem 2.3 for the formal statement).

Theorem 1.2 (Informal). *There is an algorithm that takes as inputs a source distribution P , a 2×2 target distribution Q and an error parameter $\delta > 0$, runs in time bounded by some computable function of P , Q and δ , and either outputs a non-interactive protocol that simulates Q up to additive δ in total variation distance, or asserts that there is no protocol that gets $O(\delta)$ -close to Q in total variation distance.*

The crux of Theorems 1.1 and 1.2 is to prove *computable* bounds on the number of copies of (X, Y) that are needed in order to come δ -close to the target distribution. We now describe the challenges towards achieving such bounds, and the techniques we use.

²Henceforth, we assume that bits are in the set $\{\pm 1\}$. By a *quadratic* approximation, we mean an algorithm distinguishing between the cases (i) $\rho \geq 1 - \eta$ and (ii) $\rho < 1 - O(\sqrt{\eta})$ for any given parameter $\eta > 0$.

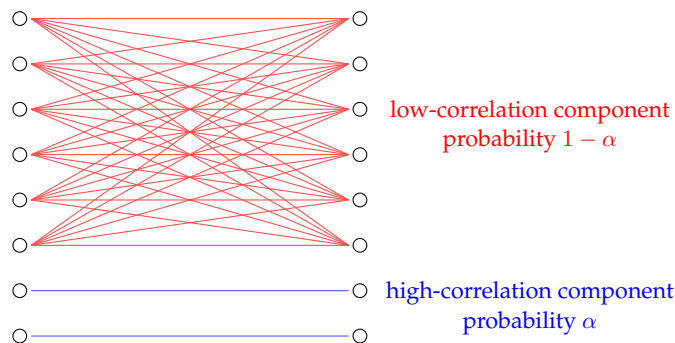


Figure 1: Example source distribution for which many copies need to be considered.

1.1 Proof Overview

We start by describing some illustrative special cases of the problem. In the case where $P = \text{DSBS}(\rho)$, maximal correlation based arguments imply that $\text{DSBS}(\rho)$ is the ‘best’ DSBS distribution that can be simulated [Wit75]. Thus, in this case, dictators functions achieve the optimal strategy. Consider now the case where P is a pair of ρ -correlated zero-mean unit-variance Gaussians³. Then, Borell’s isoperimetric inequality implies that the strategy where each of Alice and Bob outputs the sign of her/his Gaussian achieves the best possible DSBS [Bor85].

Given the above two examples where a *single-copy* strategy is optimal, it is tempting to try to determine the best DSBS that can be simulated using a single copy of P and hope that it would be close to the optimal DSBS (i.e., to the one that can be simulated using an arbitrary number of copies of P). But this approach cannot work as is illustrated by the following example which shows that using many copies of P is in some cases actually *needed*. Consider the source joint distribution corresponding to the bipartite graph in Figure 1 with $\alpha > 0$ being a small parameter (we interpret the distribution as the one obtained by sampling a random edge in the graph). This graph is the union of two components: a low-correlation component which has probability $1 - \alpha$ and a perfect-correlation component which has probability α . If we use a small number of copies of μ , the corresponding samples will most likely fall in the low-correlation component, and hence the best DSBS that can be produced in such a way would have a small correlation. On the other hand, as the number of used copies becomes larger than $1/\alpha$, with high probability at least one of the corresponding samples will fall in the perfect-correlation component, and hence the resulting DSBS would have correlation very close to 1. As another example, consider the distribution that is uniform on triples $\{(0, 0), (0, 1), (1, 0)\}$. It follows from [Wit75] that it is possible to simulate $\text{DSBS}(1/3)$ using many copies of this distribution. However, it can be shown that using only a single copy of this distribution (along with private randomness), Alice and Bob can at best simulate $\text{DSBS}(1/4)$.

We now describe at a high level, the main ideas that give us the computable bound on the number of samples of the joint distribution that are sufficient to obtain a δ -approximation to a given $\text{DSBS}(\rho)$. First, we observe that the problem of deciding if one can come δ -close to simulating $\text{DSBS}(\rho)$, is equivalent to checking if Alice and Bob can non-interactively come up with a distribution (X, Y) on $[-1, 1] \times [-1, 1]$ such that the marginals of X and Y have means close to 0, but $\mathbb{E}[XY]$ is large.

The results on correlation bounds for low-influence functions (obtained using the invariance principle) [MOO05, Mos10], say that if Alice and Bob are using only low-influential functions, then in fact the correlation that they get cannot be much better than that obtained by taking appropriate threshold functions on correlated gaussians. Moreover, Alice and Bob can in fact simulate correlated gaussians using only a constant number of samples from the joint distribution, by applying the maximal correlation based technique of Witsenhausen [Wit75].

In the general case, we show that we can first convert Alice and Bob’s functions to have *low degree*, after which we apply a regularity lemma (inspired from that of [DSTW10]) to conclude that after fixing a constant number of coordinates, the restricted function is in fact low-

³allowing here continuous distributions for the sake of intuition

influential. This reduces the general case to the special case of having low-influential functions and which is handled as described in the previous paragraph.

The more general case of simulating arbitrary 2×2 distribution also follows a similar outline. For a more technical overview of the proof, we refer the reader to Section 3.1.

1.2 Roadmap of the paper

In Section 2, we give some of the basic definitions, etc.. Our main theorems are also presented in this section as Theorems 2.3 and 2.5. In Section 3, we state our main technical lemma (Theorem 3.1), which is used to prove Theorem 2.5. We also give a proof overview for Theorem 3.1. In Sections 4, 5, 6 and 7, we state and prove the technical lemmas involved in proving Theorem 3.1. Finally, in Section 8, we put together everything to prove Theorem 3.1. We end with some open questions in Section 9.

2 Preliminaries

2.1 Notation

We use script letters \mathcal{A} , \mathcal{B} , etc. to denote finite sets, and μ will usually denote a probability distribution. $(\mathcal{A} \times \mathcal{B}, \mu)$ is a joint probability space. We use μ_A and μ_B to denote the marginal distributions of μ . We use letters x , y , etc to denote elements of \mathcal{A} , and bold letters \mathbf{x} , \mathbf{y} , etc. to denote elements in \mathcal{A}^n . We use x_i , y_i to denote individual coordinates of \mathbf{x} , \mathbf{y} , respectively.

For a probability space (\mathcal{A}, μ) , we will use the following definitions and notations borrowed from [AH11].

- $(\mathcal{A}^n, \mu^{\otimes n})$ denotes the product space $\mathcal{A} \times \mathcal{A} \times \cdots \times \mathcal{A}$ endowed with the product distribution.
- $\text{Supp}(\mu) \stackrel{\text{def}}{=} \{x : \mu(x) > 0\}$ is the support of μ . We would generally assume without loss of generality that $\text{Supp}(\mu) = \mathcal{A}$.
- $\alpha(\mu) \stackrel{\text{def}}{=} \min \{\mu(x) : x \in \text{Supp}(\mu)\}$ denotes the minimum non-zero probability of any atom in \mathcal{A} under the distribution μ .
- $L^2(\mathcal{A}, \mu)$ denotes the space of functions from \mathcal{A} to \mathbb{R} .
- The inner product on $L^2(\mathcal{A}, \mu)$ is denoted by $\langle f, g \rangle_\mu := \mathbb{E}_{x \sim \mu} [f(x)g(x)]$.
- The ℓ_p -norm by $\|f\|_p := \left[\mathbb{E}_{x \sim \mu} |f(x)|^p \right]^{1/p}$. Also, $\|f\|_\infty := \max_{\mu(x) > 0} |f(x)|$.
- It is easy to verify that $\|f\|_p \leq \|f\|_q$ for $1 \leq p \leq q$.
- For two distributions μ and ν , $d_{\text{TV}}(\mu, \nu)$ is the total variation distance between μ and ν .

2.2 The non-interactive simulation problem

The problem of non-interactive simulation is defined as follows,

Definition 2.1 (Non-interactive simulation [KA15]). *Let $(\mathcal{A} \times \mathcal{B}, \mu)$ and $(\mathcal{U} \times \mathcal{V}, \nu)$ be two probability spaces. We say that the distribution ν can be non-interactively simulated using distribution μ , if there exists a sequence of functions $\{f_n\}_{n \in \mathbb{N}}$ and $\{g_n\}_{n \in \mathbb{N}}$ such that,*

$$f_n : \mathcal{A}^n \rightarrow \mathcal{U} \quad g_n : \mathcal{B}^n \rightarrow \mathcal{V}$$

and the distribution $\nu_n \sim (f_n(\mathbf{x}), g_n(\mathbf{y}))_{\mu^{\otimes n}}$ over $\mathcal{U} \times \mathcal{V}$ is such that $\lim_{n \rightarrow \infty} d_{\text{TV}}(\nu_n, \nu) = 0$.

The notion of non-interactive simulation is pictorially depicted in Figure 2. We formulate a natural gap-version of the non-interactive simulation problem defined as follows,

Problem 2.2 (GAP-NON-INT-SIM($(\mathcal{A} \times \mathcal{B}, \mu)$, $(\mathcal{U} \times \mathcal{V}, \nu)$, δ)). *Given probability spaces $(\mathcal{A} \times \mathcal{B}, \mu)$ and $(\mathcal{U} \times \mathcal{V}, \nu)$, and an error parameter $\delta > 0$, distinguish between the following cases:*

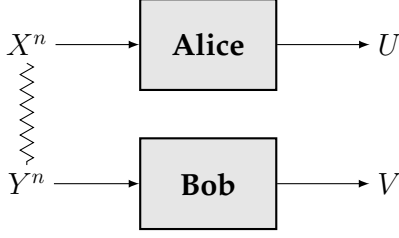


Figure 2: Non-Interactive simulation as studied in [KA12, KA15]

- (i) there exists N , and functions $f : \mathcal{A}^N \rightarrow \mathcal{U}$ and $g : \mathcal{B}^N \rightarrow \mathcal{V}$, the distribution $\nu' = (f(\mathbf{x}), g(\mathbf{y}))_{\mu^{\otimes N}}$ is such that $d_{\text{TV}}(\nu', \nu) \leq \delta$.
- (ii) for all N and all functions $f : \mathcal{A}^N \rightarrow \mathcal{U}$ and $g : \mathcal{B}^N \rightarrow \mathcal{V}$, the distribution $\nu' = (f(\mathbf{x}), g(\mathbf{y}))_{\mu^{\otimes N}}$ is such that $d_{\text{TV}}(\nu', \nu) > 8\delta$.⁴

The main result in this paper is the following theorem showing that the problem of GAP-NON-INT-SIM is decidable when $|\mathcal{U}| = |\mathcal{V}| = 2$.

Theorem 2.3 (Decidability of GAP-NON-INT-SIM for binary targets). *Given probability spaces $(\mathcal{A} \times \mathcal{B}, \mu)$ and $(\mathcal{U} \times \mathcal{V}, \nu)$ such that $|\mathcal{U}| = |\mathcal{V}| = 2$, and an error parameter δ , there exists an algorithm that runs in time $T((\mathcal{A} \times \mathcal{B}, \mu), \delta)$ (which is an explicitly computable function), and decides the problem of GAP-NON-INT-SIM $((\mathcal{A} \times \mathcal{B}, \mu), (\mathcal{U} \times \mathcal{V}, \nu), \delta)$. The run time $T((\mathcal{A} \times \mathcal{B}, \mu), \delta)$ is upper bounded by,*

$$\exp \exp \exp \left(\text{poly} \left(\frac{1}{\delta}, \frac{1}{1 - \rho_0}, \log \left(\frac{1}{\alpha} \right) \right) \right)$$

where $\rho_0 = \rho(\mathcal{A}, \mathcal{B}; \mu)$ is the maximal correlation of $(\mathcal{A} \times \mathcal{B}, \mu)$ (defined in Section 2.6) and $\alpha \stackrel{\text{def}}{=} \alpha(\mu)$ is the minimum non-zero probability in μ .

Doubly Symmetric Binary Source

In order to ease the presentation of ideas in proving the above theorem, we restrict to a special case, where the distribution $(\mathcal{U} \times \mathcal{V}; \nu)$ is a *doubly symmetric binary source* defined below.

Definition 2.4 (Doubly Symmetric Binary Source). *The distribution DSBS(ρ) is the joint distribution on ± 1 random variables (U, V) given by the following table,*

	$V = +1$	$V = -1$
$U = +1$	$(1 + \rho)/4$	$(1 - \rho)/4$
$U = -1$	$(1 - \rho)/4$	$(1 + \rho)/4$

In particular, $\mathbb{E}[U] = \mathbb{E}[V] = 0$ and $\mathbb{E}[UV] = \rho$.

We will prove a special case of Theorem 2.3, where the probability space $(\mathcal{U} \times \mathcal{V}, \nu)$ is the distribution DSBS(ρ) for some ρ (see Theorem 2.5 below). Even though we are proving only this special case, the main ideas involved here easily generalize to the proof of Theorem 2.3. We give a proof-sketch of this generalization in Section 8.1.

Theorem 2.5 (Decidability of GAP-NON-INT-SIM for DSBS targets). *Given a probability space $(\mathcal{A} \times \mathcal{B}, \mu)$, and parameters ρ and δ , there exists an algorithm that runs in time $T((\mathcal{A} \times \mathcal{B}, \mu), \delta)$ (which is an explicitly computable function), and decides the problem of GAP-NON-INT-SIM $((\mathcal{A} \times \mathcal{B}, \mu), \text{DSBS}(\rho), \delta)$.*

⁴ for sake of definition, the constant 8 could be replaced by any constant greater than 1. For a minor technical reason however our decidability results (Theorems 2.3 and 2.5) will require this constant to be strictly greater than 2. We choose to go ahead with 8 for convenience.

The run time $T((\mathcal{A} \times \mathcal{B}, \mu), \delta)$ is upper bounded by,

$$\exp \exp \exp \left(\text{poly} \left(\frac{1}{\delta}, \frac{1}{1 - \rho_0}, \log \left(\frac{1}{\alpha} \right) \right) \right)$$

where $\rho_0 = \rho(\mathcal{A}, \mathcal{B}; \mu)$ is the maximal correlation of $(\mathcal{A} \times \mathcal{B}, \mu)$ (defined in Section 2.6) and $\alpha \stackrel{\text{def}}{=} \alpha(\mu)$ is the minimum non-zero probability in μ .

We will use $\text{GAP-NON-INT-SIM}((\mathcal{A} \times \mathcal{B}, \mu), \rho, \delta)$ as a shorthand for $\text{GAP-NON-INT-SIM}((\mathcal{A} \times \mathcal{B}, \mu), \text{DSBS}(\rho), \delta)$. Theorem 2.5 will follow easily from the main technical lemma (Theorem 3.1). The proof of Theorem 2.5, assuming Theorem 3.1 is present in Section 3.2.

2.3 Reformulation of GAP-NON-INT-SIM

With the end goal of proving Theorem 2.5, we introduce a new problem of Gap-Balanced-Maximum-Inner-Product, to which we show a reduction from GAP-NON-INT-SIM. This new formulation will be better suited for applying our techniques.

Problem 2.6 ($\text{GAP-BAL-MAX-INNER-PRODUCT}((\mathcal{A} \times \mathcal{B}, \mu), \rho, \delta)$). Given a probability space $(\mathcal{A} \times \mathcal{B}, \mu)$, and parameters ρ and δ , distinguish between the following cases:

- (i) there exists N , and functions $f : \mathcal{A}^N \rightarrow [-1, 1]$ and $g : \mathcal{B}^N \rightarrow [-1, 1]$, satisfying $|\mathbb{E}[f(\mathbf{x})]| \leq \delta$ and $|\mathbb{E}[g(\mathbf{y})]| \leq \delta$, such that the following holds,

$$\mathbb{E}[f(\mathbf{x})g(\mathbf{y})] \geq \rho - \delta$$

- (ii) for all N and all functions $f : \mathcal{A}^N \rightarrow [-1, 1]$ and $g : \mathcal{B}^N \rightarrow [-1, 1]$, satisfying $|\mathbb{E}[f(\mathbf{x})]| \leq 2\delta$ and $|\mathbb{E}[g(\mathbf{y})]| \leq 2\delta$, the following holds,

$$\mathbb{E}[f(\mathbf{x})g(\mathbf{y})] < \rho - 4\delta$$

The following proposition gives a reduction from the problem of GAP-NON-INT-SIM to the problem of GAP-BAL-MAX-INNER-PRODUCT.

Proposition 2.7. For any probability space $(\mathcal{A} \times \mathcal{B}, \mu)$ and $\rho, \delta > 0$, the following reduction holds,

1. Case (i) of GAP-NON-INT-SIM $((\mathcal{A} \times \mathcal{B}, \mu), \rho, \delta)$ holds \implies
Case (i) of GAP-BAL-MAX-INNER-PRODUCT $((\mathcal{A} \times \mathcal{B}, \mu), \rho, 2\delta)$ holds
2. Case (ii) of GAP-NON-INT-SIM $((\mathcal{A} \times \mathcal{B}, \mu), \rho, \delta)$ holds \implies
Case (ii) of GAP-BAL-MAX-INNER-PRODUCT $((\mathcal{A} \times \mathcal{B}, \mu), \rho, 2\delta)$ holds

Proof. Both directions are relatively straight-forward.

1. If case (i) of GAP-NON-INT-SIM $((\mathcal{A} \times \mathcal{B}, \mu), \rho, \delta)$ holds, then there exists N and functions $f : \mathcal{A}^N \rightarrow \{1, -1\}$ and $g : \mathcal{B}^N \rightarrow \{1, -1\}$ such that the distribution $(f(\mathbf{x}), g(\mathbf{y}))_{\mu^{\otimes N}}$ is δ -close to DSBS(ρ) in total variation distance. It follows easily from the definition of total variation distance that $|\mathbb{E}[f(\mathbf{x})]| \leq 2\delta$, $|\mathbb{E}[g(\mathbf{y})]| \leq 2\delta$ and $\mathbb{E}[f(\mathbf{x})g(\mathbf{y})] \geq \rho - 2\delta$. This is exactly the conditions needed in case (i) of GAP-BAL-MAX-INNER-PRODUCT $((\mathcal{A} \times \mathcal{B}, \mu), \rho, 2\delta)$.
2. We show the contrapositive that if case (ii) of GAP-BAL-MAX-INNER-PRODUCT $((\mathcal{A} \times \mathcal{B}, \mu), \rho, 2\delta)$ does not hold, then in fact case (ii) of GAP-NON-INT-SIM $((\mathcal{A} \times \mathcal{B}, \mu), \rho, \delta)$ also does not hold. Suppose there exists N and functions $f : \mathcal{A}^N \rightarrow [-1, 1]$ and $g : \mathcal{B}^N \rightarrow [-1, 1]$ such that $|\mathbb{E}[f]| \leq 4\delta$, $|\mathbb{E}[g]| \leq 4\delta$ and $\mathbb{E}[f(\mathbf{x})g(\mathbf{y})] \geq \rho - 8\delta$. First, we observe that without loss of generality we can assume that $\mathbb{E}[f(\mathbf{x})g(\mathbf{y})] \leq \rho$. This is because, if that was not the case, we can suitably modify f and g to get $f_1 = \alpha f$ and $g_1 = \alpha g$ such that $|\mathbb{E}[f_1(\mathbf{x})]| \leq 4\delta$, $|\mathbb{E}[g_1(\mathbf{y})]| \leq 4\delta$ and $\mathbb{E}[f_1(\mathbf{x})g_1(\mathbf{y})] = \alpha^2 \cdot \mathbb{E}[f(\mathbf{x})g(\mathbf{y})]$. We can choose α suitably such that $\mathbb{E}[f_1(\mathbf{x})g_1(\mathbf{y})] \leq \rho$.

To show that case (ii) of GAP-NON-INT-SIM $((\mathcal{A} \times \mathcal{B}, \mu), \rho, \delta)$ does not hold, we obtain randomized functions $f' : \mathcal{A}^N \rightarrow \{1, -1\}$ and $g' : \mathcal{B}^N \rightarrow \{1, -1\}$ as follows, $f'(\mathbf{x})$ is equal to 1 with probability $(1 + f(\mathbf{x}))/2$ and -1 otherwise and $g'(\mathbf{y})$ is equal to 1 with probability

$(1 + g(\mathbf{y}))/2$ and -1 otherwise. [The randomness needed can be simulated using some additional copies of \mathcal{A} and \mathcal{B} .] Note that we have the following, (i) $\mathbb{E}[f'(\mathbf{x})] = \mathbb{E}[f]$ (ii) $\mathbb{E}[g'(\mathbf{y})] = \mathbb{E}[g]$ and (iii) $\rho - 8\delta \leq \mathbb{E}[f'(\mathbf{x})g'(\mathbf{y})] \leq \rho$.

Define $e_{i,j}$ for $i, j \in \{1, -1\}$ as follows,

$$\begin{aligned} e_{1,1} &= \Pr[f'(\mathbf{x}) = +1 \text{ and } g'(\mathbf{y}) = +1] - (1 + \rho)/4 \\ e_{1,-1} &= \Pr[f'(\mathbf{x}) = +1 \text{ and } g'(\mathbf{y}) = -1] - (1 - \rho)/4 \\ e_{-1,1} &= \Pr[f'(\mathbf{x}) = -1 \text{ and } g'(\mathbf{y}) = +1] - (1 - \rho)/4 \\ e_{-1,-1} &= \Pr[f'(\mathbf{x}) = -1 \text{ and } g'(\mathbf{y}) = -1] - (1 + \rho)/4 \end{aligned}$$

From (i), (ii) and (iii) above, we have the following,

$$\begin{aligned} |e_{1,1} + e_{1,-1} - e_{-1,1} - e_{-1,-1}| &\leq 4\delta \\ |e_{1,1} - e_{1,-1} + e_{-1,1} - e_{-1,-1}| &\leq 4\delta \\ |e_{1,1} - e_{1,-1} - e_{-1,1} + e_{-1,-1}| &\leq 8\delta \end{aligned}$$

In addition, we have $e_{1,1} + e_{1,-1} + e_{-1,1} + e_{-1,-1} = 0$. Combining all this, it is easy to infer that $|e_{i,j}| \leq 4\delta$ for any $i, j \in \{1, -1\}$. Hence for $\nu = (f(\mathbf{x}), g(\mathbf{y}))_{\mu^{\otimes N}}$, we get that $d_{\text{TV}}(\nu, \text{DSBS}(\rho)) \leq 8\delta$. □

2.4 Fourier analysis and multi-linear polynomials

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

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

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

It is easily seen that the functions $\{\mathcal{X}_{\sigma}\}_{\sigma \in \mathbb{Z}_q^n}$ form an orthonormal basis for the product space $L^2(\mathcal{A}^n, \mu^{\otimes n})$. Thus, every function $f \in L^2(\mathcal{A}^n, \mu^{\otimes n})$ can be written as

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

where $\widehat{f} : \mathbb{Z}_q^n \rightarrow \mathbb{R}$ can be obtained as $\widehat{f}(\sigma) = \langle f, \mathcal{X}_{\sigma} \rangle_{\mu}$. The function \widehat{f} is the Fourier transform of f with respect to the basis $\{\mathcal{X}_i\}_{i \in \mathbb{Z}_q}$. Although we will work with an arbitrary (albeit fixed) basis, many of the important properties of the Fourier transform are basis-independent. The most basic properties of \widehat{f} are summarized in the following fact which follows from the orthonormality of $\{\mathcal{X}_{\sigma}\}_{\sigma \in \mathbb{Z}_q^n}$.

Fact 2.8. *We have that:*

- *Plancherel Identity* : $\mathbb{E}[f(\mathbf{x})g(\mathbf{x})] = \sum_{\sigma} \widehat{f}(\sigma)\widehat{g}(\sigma)$.
- *As a special case, we have Parseval's identity*, $\mathbb{E}[f(\mathbf{x})^2] = \sum_{\sigma} \widehat{f}(\sigma)^2$.
- $\mathbb{E}[f] = \widehat{f}(\mathbf{0})$.
- $\text{Var}[f] = \sum_{\sigma \neq \mathbf{0}} \widehat{f}(\sigma)^2$.

In this paper, we will deal with joint probability spaces of the type $(\mathcal{A} \times \mathcal{B}, \mu)$. In such cases, we will denote the marginal probability spaces as (\mathcal{A}, μ_A) and (\mathcal{B}, μ_B) . We will abuse notations, to use \mathcal{X}_σ to denote the orthonormal basis vectors for both $L^2(\mathcal{A}^n, \mu_A^{\otimes n})$ as well as $L^2(\mathcal{B}^n, \mu_B^{\otimes n})$. The space of σ will be $\mathbb{Z}_{|\mathcal{A}|}^n$ or $\mathbb{Z}_{|\mathcal{B}|}^n$ accordingly, and will be clear from context.

For $\sigma \in \mathbb{Z}_q^n$, the *degree* of σ is denoted by $|\sigma| \stackrel{\text{def}}{=} |\{i \in [n] : \sigma_i \neq 0\}|$. We say that the degree of a function⁵ $f \in L^2(\mathcal{A}^n, \mu^{\otimes n})$, denoted by $\text{deg}(f)$, is the largest value of $|\sigma|$ such that $\widehat{f}(\sigma) \neq 0$.

Definition 2.9 (Influence). *For every coordinate $i \in [n]$, $\text{Inf}_i(f)$ is the i -th influence of f , and $\text{Inf}(f)$ is the total influence, which are defined as*

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

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

Fact 2.10. *For any function $f \in L^2(\mathcal{A}^n, \mu^{\otimes n})$, we have the following:*

- (i) $\text{Inf}_i(f) = \sum_{\sigma: \sigma_i \neq 0} \widehat{f}(\sigma)^2$ and hence, for all i , $\text{Inf}_i(f) \leq \text{Var}(f)$
- (ii) $\text{Inf}(f) = \sum_{\sigma} |\sigma| \cdot \widehat{f}(\sigma)^2$
- (iii) If $\text{deg}(f) = d$, then $\text{Inf}(f) \leq d \cdot \text{Var}[f]$.

Restrictions of polynomials

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

We now state a lemma that will be needed,

Lemma 2.11 (cf. Lemma 3.3 in [DSTW10]). *For any function $P \in L^2(\mathcal{A}^n, \mu^{\otimes n})$, consider a random assignment $\xi \sim \mu_A^H$ to the variables \mathbf{x}_H . Let $T = [n] \setminus H$. Then, for all $i \in T$, it holds that $\mathbb{E}_\xi[\text{Inf}_i(P_\xi)] = \text{Inf}_i(P)$. Also, $\mathbb{E}_\xi[\text{Var}(P_\xi)] \leq \text{Var}(P)$.*

To prove the lemma, we first recall the following fact about the expected value of Fourier coefficients under random restrictions.

Fact 2.12. *Let $P \in L^2(\mathcal{A}^n, \mu^{\otimes n})$. For any subset $H \subseteq [n]$, consider an assignment ξ to the variables \mathbf{x}_H . Let $T = [n] \setminus H$. Then, we have*

$$\widehat{P}_\xi(\sigma_T) = \sum_{\sigma_H} \widehat{P}(\sigma_H \circ \sigma_T) \cdot \mathcal{X}_{\sigma_H}(\xi)$$

and therefore

$$\mathbb{E}_\xi \left[\widehat{P}_\xi(\sigma_T)^2 \right] = \sum_{\sigma_H} \widehat{P}(\sigma_H \circ \sigma_T)^2$$

Proof. The first part follows from simply substituting $P_\xi(\mathbf{x}_T) = P(\mathbf{x}_H \leftarrow \xi, \mathbf{x}_T)$, and taking inner product with $\mathcal{X}_{\sigma_T}(\mathbf{x}_T)$.

$$\widehat{P}_\xi(\sigma_T) = \left\langle \sum_{\sigma_H \circ \sigma'_T} \widehat{P}(\sigma_H \circ \sigma'_T) \cdot \mathcal{X}_{\sigma_H}(\xi) \mathcal{X}_{\sigma'_T}(\mathbf{x}_T), \mathcal{X}_{\sigma_T}(\mathbf{x}_T) \right\rangle_\mu = \sum_{\sigma_H} \widehat{P}(\sigma_H \circ \sigma_T) \cdot \mathcal{X}_{\sigma_H}(\xi)$$

⁵we will interchangeably use the word *polynomial* to talk about any function in $L^2(\mathcal{A}^n, \mu^{\otimes n})$.

The second part simply follows from the orthonormality of the characters \mathcal{X}_{σ_H} and $\mathcal{X}_{\sigma'_H}$ for $\sigma_H \neq \sigma'_H$. In particular, we have the following,

$$\begin{aligned}
\mathbb{E}_\xi \left[\widehat{P}_\xi(\sigma_T)^2 \right] &= \mathbb{E}_\xi \left[\left(\sum_{\sigma_H} \widehat{P}(\sigma_H \circ \sigma_T) \cdot \mathcal{X}_{\sigma_H}(\xi) \right)^2 \right] \\
&= \mathbb{E}_\xi \left[\sum_{\sigma_H \sigma'_H} \widehat{P}(\sigma_H \circ \sigma_T) \cdot \widehat{P}(\sigma'_H \circ \sigma_T) \cdot \mathcal{X}_{\sigma_H}(\xi) \cdot \mathcal{X}_{\sigma'_H}(\xi) \right] \\
&= \sum_{\sigma_H \sigma'_H} \widehat{P}(\sigma_H \circ \sigma_T) \cdot \widehat{P}(\sigma'_H \circ \sigma_T) \cdot \mathbb{E}_\xi [\mathcal{X}_{\sigma_H}(\xi) \cdot \mathcal{X}_{\sigma'_H}(\xi)] \\
&= \sum_{\sigma_H} \widehat{P}(\sigma_H \circ \sigma_T)^2
\end{aligned}$$

□

Intuitively, the above fact says that all the Fourier weight on degree sequences $\{\sigma_H \circ \sigma_T\}_{\sigma_H}$ collapses down onto σ_T in expectation. Consequently, the influence of an unrestricted variable does not change, and the variance does not increase in expectation under random restrictions, as both these quantities are sums of Fourier weight on certain σ_T 's.

Proof of Lemma 2.11. We simply use Facts 2.8 and 2.10 in addition to Fact 2.12 to prove the lemma.

Basically, from Facts 2.8 and 2.12 we get,

$$\mathbb{E}_\xi [\text{Var}(P_\xi)] = \mathbb{E}_\xi \left[\sum_{\sigma_T \neq 0} \widehat{P}_\xi(\sigma_T)^2 \right] = \sum_{\sigma_T \neq 0} \mathbb{E}_\xi \left[\widehat{P}_\xi(\sigma_T)^2 \right] = \sum_{\sigma_T \neq 0} \sum_{\sigma_H} \widehat{P}(\sigma_H \circ \sigma_T)^2 \leq \text{Var}(P)$$

Similarly, from Facts 2.10 and 2.12 we get that for all $i \in T$,

$$\mathbb{E}_\xi [\text{Inf}_i(P_\xi)] = \mathbb{E}_\xi \left[\sum_{\substack{\sigma_T: \\ (\sigma_T)_i \neq 0}} \widehat{P}_\xi(\sigma_T)^2 \right] = \sum_{\substack{\sigma_T: \\ (\sigma_T)_i \neq 0}} \mathbb{E}_\xi \left[\widehat{P}_\xi(\sigma_T)^2 \right] = \sum_{\substack{\sigma_T: \\ (\sigma_T)_i \neq 0}} \sum_{\sigma_H} \widehat{P}(\sigma_H \circ \sigma_T)^2 = \text{Inf}_i(P)$$

□

2.5 Hypercontractivity and moment bounds

The following moment bound for low-degree polynomials appears as Theorem 2.7 in [AH11], which in turn follows from hypercontractivity.

Theorem 2.13 ([Wol07]). *Let (\mathcal{A}, μ) be a finite probability space in which the minimum non-zero probability is $\alpha(\mu) \leq 1/2$. Then, for $p \geq 2$, every degree- d polynomial $f \in L^2(\mathcal{A}^n, \mu^{\otimes n})$ satisfies*

$$\|f\|_p \leq C_p(\alpha)^{d/2} \|f\|_2$$

Here, C_p is defined by

$$C_p(\alpha) = \frac{A^{1/p'} - A^{-1/p'}}{A^{1/p} - A^{-1/p}}$$

where $A = (1 - \alpha)/\alpha$ and $1/p + 1/p' = 1$. The value at $\alpha = 1/2$ is taken to be the limit of the above expression as $\alpha \rightarrow 1/2$, i.e., $C_p(1/2) = p - 1$.

We will use the following known concentration bound for low-degree polynomials.

Theorem 2.14 ([AH11]; Theorem 2.12). *Let $f \in L^2(\mathcal{A}^n, \mu^{\otimes n})$ be a degree- d polynomial. Then, for any $t > e^{d/2}$,*

$$\Pr[|f| > t \cdot \|f\|_2] \leq \exp(-ct^{2/d})$$

where $c := \frac{\alpha(\mu)d}{e}$.

Definition 2.15 (Bonami-Beckner operator). For any $\rho \in [0, 1]$, the Bonami-Beckner operator T_ρ on a probability space (\mathcal{A}, μ) is given by its action on any $f : \mathcal{A} \rightarrow \mathbb{R}$, as follows,

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

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

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

2.6 Maximal Correlation and Witsenhausen's rounding

The ‘‘maximal correlation coefficient’’ was first introduced by Hirschfeld [Hir35] and Gebelein [Geb41] and then studied by Rényi [Rén59].

Definition 2.16 (Maximal correlation). Given a joint probability space $(\mathcal{A} \times \mathcal{B}, \mu)$, we define the maximal correlation of the joint distribution $\rho(\mathcal{A}, \mathcal{B}; \mu)$ as follows,

$$\rho(\mathcal{A}, \mathcal{B}; \mu) = \sup \left\{ \mathbb{E}_{(x,y) \sim \mu} [f(x)g(y)] \mid \begin{array}{l} f : \mathcal{A} \rightarrow \mathbb{R}, \quad \mathbb{E}[f] = \mathbb{E}[g] = 0 \\ g : \mathcal{B} \rightarrow \mathbb{R}, \quad \text{Var}(f) = \text{Var}(g) = 1 \end{array} \right\}$$

Maximal correlation has the following properties which imply necessary conditions for when non-interactive simulation could be possible.

Fact 2.17 (Properties of maximal correlation (cf. [BDK05])).

1. (Tensorization) : For all joint probability spaces $(\mathcal{A} \times \mathcal{B}, \mu)$, it is the case that $\rho(\mathcal{A}^n, \mathcal{B}^n; \mu^{\otimes n}) = \rho(\mathcal{A}, \mathcal{B}; \mu)$.
2. (Data processing) : For all joint probability spaces $(\mathcal{A} \times \mathcal{B}, \mu)$, and any functions $f : \mathcal{A} \rightarrow \mathcal{U}$ and $g : \mathcal{B} \rightarrow \mathcal{V}$, it is the case that $\rho(\mathcal{A}, \mathcal{B}; \mu) \geq \rho(\mathcal{U}, \mathcal{V}; \nu)$, where ν is the distribution $(f(x), g(y))_{(x,y) \sim \mu}$.
3. (Lower Semi-Continuous) : If distributions $(\mathcal{U} \times \mathcal{V}; \nu_n)$ are such that $\lim_{n \rightarrow \infty} \nu_n = \nu$, then $\lim_{n \rightarrow \infty} \rho(\mathcal{U}, \mathcal{V}; \nu_n) \geq \rho(\mathcal{U}, \mathcal{V}; \nu)$.

Corollary 2.18 (Necessary condition for non-interactive simulation). Let $(\mathcal{A} \times \mathcal{B}, \mu)$ and $(\mathcal{U} \times \mathcal{V}, \nu)$ be two probability spaces. If the distribution ν can be non-interactively simulated using distribution μ , then it must be the case that $\rho(\mathcal{A}, \mathcal{B}; \mu) \geq \rho(\mathcal{U}, \mathcal{V}; \nu)$.

A simple fact that can be easily verified is that the maximal correlation of the distribution DSBS(ρ) is ρ . And hence if $(\mathcal{A} \times \mathcal{B}, \mu)$ can non-interactively simulate DSBS(ρ^*), then $\rho^* \leq \rho(\mathcal{A}, \mathcal{B}; \mu)$. In addition, Witsenhausen [Wit75] showed that any joint probability space $(\mathcal{A} \times \mathcal{B}, \mu)$ can simulate DSBS(ρ^*) for $\rho^* = 1 - \frac{2 \arccos(\rho(\mathcal{A}, \mathcal{B}; \mu))}{\pi}$. All together, we have the following theorem,

Theorem 2.19 (Witsenhausen [Wit75]). For any joint probability space $(\mathcal{A} \times \mathcal{B}, \mu)$, with $\rho = \rho(\mathcal{A}, \mathcal{B}; \mu)$, then the largest ρ^* for which $(\mathcal{A} \times \mathcal{B}, \mu)$ can non-interactively simulate DSBS(ρ^*) is bounded as follows,

$$1 - \frac{2 \arccos(\rho)}{\pi} \leq \rho^* \leq \rho$$

Note that, maximal correlation is an easily computable quantity, namely, it is the second largest singular value of the Markov operator⁶ corresponding to $(\mathcal{A} \times \mathcal{B}, \mu)$.

Remark 2.20. The astute reader might have noticed a strong resemblance between Theorem 2.19 and the random hyperplane rounding of Goemans-Williamson [GW95] used in the approximation algorithm for MAX-CUT. This is not a coincidence and indeed the bounds in Theorem 2.19 come from morally the same technique as in [GW95].

In this context we will use the following shorthand for ρ -correlated 2-dimensional gaussian.

Definition 2.21 (2-dimensional Gaussian). Let $\mathcal{G}(\rho)$ denote a 2-dimensional gaussian distribution with mean $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and covariance matrix $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.

⁶The Markov operator corresponding to $(\mathcal{A} \times \mathcal{B}, \mu)$ is a $|\mathcal{A}| \times |\mathcal{B}|$ matrix T which is given by $T(x, y) = \mu(y|X = x)$.

2.7 2-dimensional Berry-Esseen theorem

We will need the following 2-dimensional Berry-Esseen theorem. The proof is very similar to Theorem 68 of [MORS10]. The main difference is that in our case the random variables are not necessarily binary-valued, but they do have finite support. We include the proof for completeness.

Lemma 2.22 (2-dimensional Berry-Esseen). *Let (X, Y) be any pair of correlated real-valued random variables with finite support such that, $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ and $\text{Var}(X) = \text{Var}(Y) = 1$ and $\mathbb{E}[XY] = \rho$. For every $\zeta > 0$, there exists $w \stackrel{\text{def}}{=} w((X, Y), \zeta) \in \mathbb{N}$, such that for every $a, b \in \mathbb{R}$, it is the case that,*

$$|\Pr[\bar{X} \leq a, \bar{Y} \leq b] - \Pr[\mathcal{G}_1 \leq a, \mathcal{G}_2 \leq b]| \leq \zeta$$

where $\bar{X} = \frac{\sum_{i=1}^w X_i}{\sqrt{w}}$, $\bar{Y} = \frac{\sum_{i=1}^w Y_i}{\sqrt{w}}$ (with (X_i, Y_i) draw i.i.d. from (X, Y)) and $(\mathcal{G}_1, \mathcal{G}_2) \sim \mathcal{G}(\rho)$.

In particular, one may take $w = O\left(\frac{1+\rho}{\alpha \cdot (1-\rho)^3 \cdot \zeta^2}\right)$, where α is the minimum non-zero probability in the distribution (X, Y) .

In order to prove Lemma 2.22, we need the following statement that appears as Theorem 16 in [KKMO07] and as Corollary 16.3 in [BR86].

Theorem 2.23. *Let Z_1, \dots, Z_w be independent random variables taking values in \mathbb{R}^k and satisfying:*

- $\mathbb{E}[Z_j]$ is the all-zero vector for every $j \in \{1, \dots, w\}$.
- $\sum_{j=1}^w \text{Cov}[Z_j]/w = V$ where Cov denotes the covariance matrix.
- λ is the smallest eigenvalue of V and Λ is the largest eigenvalue of V .
- $\sum_{j=1}^w \mathbb{E}[\|Z_j\|^3]/w = \rho_3 < \infty$.

Let Q_w denote the distribution of $(Z_1 + \dots + Z_w)/\sqrt{w}$, let $\Phi_{0,V}$ denote the distribution of the k -dimensional Gaussian with mean 0 and covariance matrix V , and let $\eta = C\lambda^{-3/2}\rho_3 w^{-1/2}$, where C is a certain universal constant. Then, for any Borel set A ,

$$|Q_w(A) - \Phi_{0,V}(A)| \leq \eta + B(A)$$

where $B(A)$ is the following measure of the boundary of A : $B(A) = 2 \sup_{y \in \mathbb{R}^k} \Phi_{0,V}((\partial A)^{\eta'} + y)$, $\eta' = \Lambda^{1/2}\eta$ and $(\partial A)^{\eta'}$ denotes the set of points within distance η' of the topological boundary of A .

Proof of Lemma 2.22. We apply Theorem 2.23 with $k = 2$. Let $Z = (X, Y)$, and hence we have that,

$$\mathbb{E}[Z] = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \text{Cov}[Z] = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

Let $Z_i = (X_i, Y_i)$. Since all Z_i are i.i.d. distributed according to Z , we have $V = \sum_{j=1}^w \text{Cov}[(X_j, Y_j)]/w$ is also $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. It follows that the smallest and largest eigenvalues of V are $\lambda = 1 - \rho$ and $\Lambda = 1 + \rho$ respectively. Moreover, since the underlying distribution has finite support, we have that,

$$\sum_{j=1}^w \frac{\mathbb{E}[\|Z_j\|^3]}{w} = \mathbb{E}[\|Z\|^3] < \max \|Z\| \cdot \mathbb{E}[\|Z\|^2] \leq 1/\sqrt{\alpha}$$

(where α is the smallest atom in the distribution (X, Y)). Thus, we get $\rho_3 \leq 1/\sqrt{\alpha}$. Hence, $\eta = O((1 - \rho)^{-3/2} \alpha^{-1/2} w^{-1/2})$. As in [KKMO07], one can check that the topological boundary of any set of the form $(-\infty, a] \times (-\infty, b]$ is $O(\eta')$, where $\eta' = (1 + \rho)^{1/2} \eta$. Thus, from Lemma 2.22 it follows by choosing w sufficiently large so that $O((1 + (1 + \rho)^{1/2})(1 - \rho)^{-3/2} \alpha^{-1/2} w^{-1/2}) \leq \zeta$.

In particular it suffices to choose $w = O\left(\frac{(1+(1+\rho)^{1/2})^2}{\alpha \cdot (1-\rho)^3 \cdot \zeta^2}\right) = O\left(\frac{1+\rho}{\alpha \cdot (1-\rho)^3 \cdot \zeta^2}\right)$. \square

3 Main Technical Lemma and Overview

In this section we state the main technical lemma which will be used to solve GAP-BAL-MAX-INNER-PRODUCT. We also give a high level overview of the proof techniques.

Theorem 3.1. *Given any joint probability space $(\mathcal{A} \times \mathcal{B}, \mu)$ and any $\delta > 0$, there exists $n_0 = n_0((\mathcal{A} \times \mathcal{B}, \mu), \delta)$ such that for any n and any functions $f : \mathcal{A}^n \rightarrow [-1, 1]$ and $g : \mathcal{B}^n \rightarrow [-1, 1]$, there exist functions $\tilde{f} : \mathcal{A}^{n_0} \rightarrow [-1, 1]$ and $\tilde{g} : \mathcal{B}^{n_0} \rightarrow [-1, 1]$ such that $|\mathbb{E}[\tilde{f}] - \mathbb{E}[f]| \leq \delta/3$, $|\mathbb{E}[\tilde{g}] - \mathbb{E}[g]| \leq \delta/3$ and*

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n_0}} [\tilde{f}(\mathbf{x}) \cdot \tilde{g}(\mathbf{y})] \geq \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n}} [f(\mathbf{x}) \cdot g(\mathbf{y})] - \delta$$

Most importantly, n_0 is a computable function in the parameters of the problem. In particular, one may take,

$$n_0 = \exp \left(\text{poly} \left(\frac{1}{\delta}, \frac{1}{1-\rho}, \log \left(\frac{1}{\alpha} \right) \right) \right)$$

where $\rho \stackrel{\text{def}}{=} \rho(\mathcal{A}, \mathcal{B}; \mu)$ is the maximal correlation of $(\mathcal{A} \times \mathcal{B}, \mu)$ and $\alpha \stackrel{\text{def}}{=} \alpha(\mu)$ is the minimum non-zero probability in μ .

3.1 Proof overview

The proof of Theorem 3.1 goes through a series of intermediate steps, which we describe at a high level here. At each step we lose only a small amount in the correlation. The first three steps preserve the marginals $\mathbb{E}[f]$ and $\mathbb{E}[g]$ exactly, while the fourth step incurs a small additive error in the same. The full proof is presented in Section 8.

Step 1: Smoothing of strategies. We transform f and g into functions f_1, g_1 such that f_1 and g_1 have ‘most’ of their Fourier mass concentrated on terms of degree at most d , where d is a constant that depends on the distribution $(\mathcal{A} \times \mathcal{B}, \mu)$ and a tolerance parameter, but is independent of n . This transformation is described in Section 4.

Step 2: Regularity lemma for low degree functions. We first prove a *regularity lemma* (similar to the one in [DSTW10]) which roughly shows that for any degree- d polynomial, there exists a h -sized subset of variables, such that under a random restriction of the variables in this subset, the resulting function on the remaining variables has low individual influences (i.e. $\leq \tau$). Note that h will be a constant depending on the degree d and τ , but will be independent of n .

We apply this regularity lemma on the degree- d truncated versions of both f_1 and g_1 obtained from Step 1. We take the union of the subsets obtained for f_1 and g_1 . We show that with high probability over random restrictions of the variables in this subset, the resulting restriction of f_1 and g_1 on the remaining variables has low individual influences. This step is described in Section 5.

Note that this step does not change the functions f_1 and g_1 at all, but we gain some structural knowledge about the same.

Step 3: Correlation bounds for low influence functions. We use results about correlation bounds for low influential functions [MOO05, Mos10]. Intuitively, these results suggest that if the functions f_1 and g_1 were low influential functions to begin with, then the correlation $\mathbb{E}[f_1(\mathbf{x})g_1(\mathbf{y})]$ will not be ‘much’ better than the correlation between certain threshold functions applied on correlated gaussians.

We apply the above correlation bounds for the low influential functions obtained by restrictions of the small subset of variables in f_1 and g_1 , to obtain functions $f_2 : \mathcal{A}^h \times \mathbb{R} \rightarrow [-1, 1]$ and $g_2 : \mathcal{B}^h \times \mathbb{R} \rightarrow [-1, 1]$, where Alice and Bob together have access to h samples from $(\mathcal{A} \times \mathcal{B}, \mu)$ and a single copy of ρ -correlated gaussians, that is, $\mathcal{G}(\rho)$ (see Defn. 2.21). Here the correlation ρ is same as the maximal correlation $\rho(\mathcal{A}, \mathcal{B}; \mu)$. This step is described in Section 6.

Step 4: Simulating correlated gaussians. Finally, Alice and Bob can non-interactively simulate the distribution $\mathcal{G}(\rho)$ using constantly many samples from $(\mathcal{A} \times \mathcal{B}, \mu)$. This is done using the technique of Witsenhausen [Wit75], which primarily uses a 2-dimensional central limit theorem. This step is described in Section 7.

3.2 Decidability of GAP-NON-INT-SIM

Assuming Theorem 3.1, we now give the algorithm as described in Theorem 2.5.

Proof of Theorem 2.5. We have from Proposition 2.7 that, in order to decide $\text{GAP-NON-INT-SIM}((\mathcal{A} \times \mathcal{B}, \mu), \rho, \delta)$, it suffices to decide $\text{GAP-BAL-MAX-INNER-PRODUCT}((\mathcal{A} \times \mathcal{B}, \mu), \rho, 2\delta)$.

If we were in the YES case of $\text{GAP-BAL-MAX-INNER-PRODUCT}((\mathcal{A} \times \mathcal{B}, \mu), \rho, 2\delta)$, then we have that there exists an n and functions $f : \mathcal{A}^n \rightarrow [-1, 1]$ and $g : \mathcal{B}^n \rightarrow [-1, 1]$, such that $|\mathbb{E}[f(\mathbf{x})]| \leq 2\delta$, $|\mathbb{E}[g(\mathbf{y})]| \leq 2\delta$ and $\mathbb{E}[f(\mathbf{x}) \cdot g(\mathbf{y})] \geq \rho - 2\delta$. Using Theorem 3.1, with parameter δ , we get that there exists functions $\tilde{f} : \mathcal{A}^{n_0} \rightarrow [-1, 1]$ and $\tilde{g} : \mathcal{B}^{n_0} \rightarrow [-1, 1]$ such that $|\mathbb{E}[\tilde{f}(\mathbf{x})]| \leq 8\delta/3$, $|\mathbb{E}[\tilde{g}(\mathbf{y})]| \leq 8\delta/3$ and $\mathbb{E}[\tilde{f}(\mathbf{x}) \cdot \tilde{g}(\mathbf{y})] \geq \rho - 3\delta$.

In the NO case of $\text{GAP-BAL-MAX-INNER-PRODUCT}((\mathcal{A} \times \mathcal{B}, \mu), \rho, 2\delta)$, we have that for all n , in particular for $n = n_0$, there do not exist functions $f : \mathcal{A}^n \rightarrow [-1, 1]$ and $g : \mathcal{B}^n \rightarrow [-1, 1]$ such that $|\mathbb{E}[f(\mathbf{x})]| \leq 4\delta$, $|\mathbb{E}[g(\mathbf{y})]| \leq 4\delta$ and $\mathbb{E}[f(\mathbf{x}) \cdot g(\mathbf{y})] \geq \rho - 8\delta$.

This naturally gives us a brute force algorithm: Analyze all possible functions $\tilde{f} : \mathcal{A}^{n_0} \rightarrow [-1, 1]$ and $\tilde{g} : \mathcal{B}^{n_0} \rightarrow [-1, 1]$ to check if there exist functions satisfying $|\mathbb{E}[\tilde{f}(\mathbf{x})]| \leq 8\delta/3$, $|\mathbb{E}[\tilde{g}(\mathbf{y})]| \leq 8\delta/3$ and $\mathbb{E}[\tilde{f}(\mathbf{x}) \cdot \tilde{g}(\mathbf{y})] \geq \rho - 3\delta$. For purposes of our algorithm we can treat the range $[-1, 1]$ as a discrete set $R \stackrel{\text{def}}{=} \{k\delta^2/10 : k \in \mathbb{Z}, |k| < 10/\delta^2\}$. This ensures that if indeed such a desired \tilde{f} and \tilde{g} exist, then we will find functions $\tilde{f}' : \mathcal{A}^{n_0} \rightarrow R$ and $\tilde{g}' : \mathcal{B}^{n_0} \rightarrow R$ such that $|\mathbb{E}[\tilde{f}'(\mathbf{x})]|, |\mathbb{E}[\tilde{g}'(\mathbf{y})]| \leq 8\delta/3 + O(\delta^2)$ and $\mathbb{E}[\tilde{f}'(\mathbf{x}) \cdot \tilde{g}'(\mathbf{y})] \geq \rho - 3\delta - O(\delta^2)$. In the YES case, we will find such functions, whereas in the NO case, \tilde{f}' and \tilde{g}' as above simply don't exist.

It is easy to see that this brute force can be done in $(\frac{1}{\delta^2})^{O((|\mathcal{A}|+|\mathcal{B}|)^{n_0})}$ time, which is upper bounded by the running time claimed in Theorem 2.5. \square

4 Smoothing of Strategies

The first step in our approach is to obtain smoothed versions of the functions $f : \mathcal{A}^n \rightarrow [-1, 1]$ and $g : \mathcal{B}^n \rightarrow [-1, 1]$, which have *small Fourier tails*, without hurting the correlation by much. In particular, we show the following lemma.

Lemma 4.1 (Smoothing of strategies). *Given any joint probability space $(\mathcal{A} \times \mathcal{B}, \mu)$ and parameters $\lambda, \eta > 0$, there exists $d = d((\mathcal{A} \times \mathcal{B}, \mu), \lambda, \eta)$ such that for any n and any functions $f : \mathcal{A}^n \rightarrow [-1, 1]$ and $g : \mathcal{B}^n \rightarrow [-1, 1]$, there exist functions $f_1 : \mathcal{A}^n \rightarrow [-1, 1]$ and $g_1 : \mathcal{B}^n \rightarrow [-1, 1]$ such that $\mathbb{E}[f_1] = \mathbb{E}[f]$ and $\mathbb{E}[g_1] = \mathbb{E}[g]$, and*

$$|\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n}} [f_1(\mathbf{x}) \cdot g_1(\mathbf{y})] - \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n}} [f(\mathbf{x}) \cdot g(\mathbf{y})]| \leq \lambda$$

such that f_1 and g_1 have low energy Fourier tails, namely,

$$\sum_{|\boldsymbol{\sigma}| > d} \widehat{f}_1(\boldsymbol{\sigma})^2 \leq \eta \quad \text{and} \quad \sum_{|\boldsymbol{\sigma}| > d} \widehat{g}_1(\boldsymbol{\sigma})^2 \leq \eta$$

In particular, one may take $d = \frac{\log \eta}{2 \log \gamma}$, where $\gamma = 1 - C \frac{(1-\rho)\lambda}{\log(1/\lambda)}$, and $\rho = \rho(\mathcal{A}, \mathcal{B}; \mu)$.

To prove Lemma 4.1, we use Lemma 6.1 of Mossel [Mos10]. We state a specialized version of Mossel's lemma, which suffices for our application.

Lemma 4.2 ([Mos10]). *Let $(\mathcal{A} \times \mathcal{B}, \mu)$ be finite joint probability space, such that $\rho(\mathcal{A} \times \mathcal{B}, \mu) \leq \rho$.*

Let $P \in L^2(\mathcal{A}^n, \mu_{\mathcal{A}}^{\otimes n})$ and $Q \in L^2(\mathcal{B}^n, \mu_{\mathcal{B}}^{\otimes n})$ be multi-linear polynomials. Let $\varepsilon > 0$ and γ be chosen sufficiently close to 1 so that,

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

Then:

$$|\mathbb{E}[P(\mathbf{x})Q(\mathbf{y})] - \mathbb{E}[T_\gamma P(\mathbf{x})T_\gamma Q(\mathbf{y})]| \leq 2\varepsilon \text{Var}[P] \text{Var}[Q]$$

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

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

Proof of Lemma 4.1. Given parameters λ and η , we first choose ε and γ in Lemma 4.2, such that $\varepsilon = \lambda/2$ and $\gamma = 1 - C((1 - \rho)\varepsilon) / (\log(1/\varepsilon))$ as required. We choose d to be large enough such that $\gamma^{2d} \leq \eta$, that is, $d = (\log \eta) / (2 \log \gamma)$. Now, given functions $f : \mathcal{A}^n \rightarrow [-1, 1]$ and $g : \mathcal{B}^n \rightarrow [-1, 1]$, we obtain functions f_1 and g_1 as follows: $f_1(\mathbf{x}) = T_\gamma f(\mathbf{x})$ and $g_1(\mathbf{y}) = T_\gamma g(\mathbf{y})$. It is easy to see that, $\mathbb{E}[f_1(\mathbf{x})] = \mathbb{E}[f(\mathbf{x})]$ and $\mathbb{E}[g_1(\mathbf{y})] = \mathbb{E}[g(\mathbf{y})]$. From Lemma 4.2, and the fact that $\text{Var}[f], \text{Var}[g] \leq 1$, we get $|\mathbb{E}[f_1(\mathbf{x})g_1(\mathbf{y})] - \mathbb{E}[f(\mathbf{x})g(\mathbf{y})]| \leq 2\varepsilon = \lambda$ as desired. Also, note that $\hat{f}_1(\boldsymbol{\sigma}) = \hat{f}(\boldsymbol{\sigma}) \cdot \gamma^{|\boldsymbol{\sigma}|}$ (similarly for $\hat{g}_1(\boldsymbol{\sigma})$). Thus, we get that,

$$\begin{aligned} \sum_{|\boldsymbol{\sigma}| > d} \hat{f}_1(\boldsymbol{\sigma})^2 &\leq \gamma^{2d} \cdot \sum_{|\boldsymbol{\sigma}| > d} \hat{f}(\boldsymbol{\sigma})^2 \leq \gamma^{2d} \leq \eta \\ \sum_{|\boldsymbol{\sigma}| > d} \hat{g}_1(\boldsymbol{\sigma})^2 &\leq \gamma^{2d} \cdot \sum_{|\boldsymbol{\sigma}| > d} \hat{g}(\boldsymbol{\sigma})^2 \leq \gamma^{2d} \leq \eta \end{aligned}$$

□

5 Joint Regularity Lemma for Fourier Concentrated Functions

The second step in our approach is to apply a *regularity lemma* on the functions $f_1 : \mathcal{A}^n \rightarrow [-1, 1]$ and $g_1 : \mathcal{B}^n \rightarrow [-1, 1]$ obtained from the previous step of smoothing. *Regularity lemma* is a loosely referred term which shows that for various types of combinatorial objects, an arbitrary object can be approximately decomposed into a constant number of “pseudorandom” sub-objects.

Our version of the regularity lemma draws inspiration from that of [DSTW10]; in fact our proofs also closely follow theirs. Formally, we show the following lemma.

Lemma 5.1 (Joint regularity lemma for Fourier-concentrated functions). *Let $(\mathcal{A} \times \mathcal{B}, \mu)$ be a joint probability space. Let $d \in \mathbb{N}$ and $\tau > 0$ be any given constant parameters. There exists an $\eta \stackrel{\text{def}}{=} \eta(\tau) > 0$ and $h \stackrel{\text{def}}{=} h((\mathcal{A} \times \mathcal{B}, \mu), d, \tau)$ such that the following holds:*

For all $P \in L^2(\mathcal{A}^n, \mu_{\mathcal{A}}^{\otimes n})$ and $Q \in L^2(\mathcal{B}^n, \mu_{\mathcal{B}}^{\otimes n})$ satisfying $\sum_{|\boldsymbol{\sigma}| > d} \hat{P}(\boldsymbol{\sigma})^2 \leq \eta$, $\sum_{|\boldsymbol{\sigma}| > d} \hat{Q}(\boldsymbol{\sigma})^2 \leq \eta$, and $\text{Var}[P] \leq 1$ and $\text{Var}[Q] \leq 1$: there exists a subset of indices $H \subseteq [n]$ with $|H| \leq h$, such that the restrictions of the functions P and Q obtained by evaluating the coordinates in H according to distribution μ , satisfy the following (where we denote $T = [n] \setminus H$),

- *With probability at least $1 - \tau$ over $\xi \sim \mu_{\mathcal{A}}^{\otimes h}$, the restriction $P_\xi(\mathbf{x}_T)$ is such that for all $i \in T$, it is the case that $\text{Inf}_i(P_\xi(\mathbf{x}_T)) \leq \tau$*
- *With probability at least $1 - \tau$ over $\xi \sim \mu_{\mathcal{B}}^{\otimes h}$, the restriction $Q_\xi(\mathbf{x}_T)$ is such that for all $i \in T$, it is the case that $\text{Inf}_i(Q_\xi(\mathbf{x}_T)) \leq \tau$*

In particular, one may take $\eta = \tau^2/16$ and $h = \frac{d}{\tau^2} \cdot \left(\frac{C_4(\alpha)}{\alpha} \log \frac{C_4(\alpha)}{\alpha \cdot d \cdot \tau} \right)^{O(d)}$ which is a constant that depends on d, τ and $\alpha \stackrel{\text{def}}{=} \alpha(\mu)$, which is the minimum non-zero probability in μ .

5.1 Regularity Lemma for Constant Degree Polynomials

We first prove a version of the above regularity lemma for degree- d functions, as opposed to Fourier-concentrated functions.

Lemma 5.2 (Regularity Lemma for degree- d functions). *Let (\mathcal{A}, μ_A) be a probability space. Let $d \in \mathbb{N}$ and $\tau > 0$ be any given constant parameters. There exists $h \stackrel{\text{def}}{=} h((\mathcal{A}, \mu_A), d, \tau)$ such that the following holds:*

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

$$\Pr_{\xi \sim \mu_A^{\otimes |H|}} [\forall i \in T : \text{Inf}_i(P_\xi(\mathbf{x}_T)) \leq \tau] \geq 1 - \tau$$

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

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

The intuitive explanation of the regularity lemma is as follows: If P is a degree d polynomial with $\text{Var}(P) \leq 1$, then the total influence of P is at most d . Hence for all $\beta > 0$, there can only be at most $h \stackrel{\text{def}}{=} d/\beta$ variables with influence greater than β . Indeed, our subset H_0 will essentially be the set of all the variables with influence at least β (we will choose β to be suitably smaller than τ , but with no dependence on n). Clearly, $|H_0| \leq h$. For any superset $H \supseteq H_0$, and for a random restriction of \mathbf{x}_H to ξ , it will follow from well known hypercontractivity bounds (Theorem 2.14) and a careful union bound, that the influence of all the remaining variables will be less than τ with high probability.

Our regularity lemma draws inspiration from the one in [DSTW10]. In fact, our proof of the above regularity lemma also closely follows the proof steps in [DSTW10]. However their regularity lemma was much more involved as they were dealing with low-degree polynomial threshold functions, whereas we are directly dealing with low-degree polynomials. In particular, a major difference in our regularity lemmas is that [DSTW10] obtain a (potentially) *adaptive* decision tree, whereas we obtain just a single subset H . Also, our notion of ‘regularity’ is much simpler in that we only need all influences to be small. Another aspect of our regularity lemma is that it is robust enough to also work for Fourier concentrated functions, as opposed to only low-degree functions (potentially, [DSTW10] could also be modified to have this feature, although it was not required for their application). Another minor difference is that our Fourier analysis is for functions in $L^2(\mathcal{A}^n, \mu_A^{\otimes n})$, as opposed to functions on the boolean hypercube. But this is not really a significant difference and the proof steps go through as it is, albeit with slightly different parameters which depend on the hypercontractivity parameters of the distribution (\mathcal{A}, μ_A) .

Before we give a proof of Lemma 5.2, we would need the following claim.

Claim 5.3 (cf. Claim 3.12 in [DSTW10]). *Let $P \in L^2(\mathcal{A}^n, \mu_A^{\otimes n})$ be a degree- d polynomial. Let $H \subseteq [n]$ and $T = [n] \setminus H$. Let ξ be a random restriction fixing H . For all $r \geq e^d$ and all $i \in T$, we have the following,*

$$\Pr_{\xi} [\text{Inf}_i(P_\xi) > r \cdot C_4(\alpha)^d \cdot \text{Inf}_i(P)] \leq \exp(-c \cdot r^{1/d})$$

where, $c = \alpha(\mu_A)d/e$ (see Theorem 2.14) and $C_4(\alpha)$ is obtained as in Theorem 2.13.

Proof. The identity $\text{Inf}_i(P_\xi) = \sum_{\sigma_T: (\sigma_T)_i \neq 0} \widehat{P}_\xi(\sigma_T)^2$ and Fact 2.12 imply that $\text{Inf}_i(P_\xi)$ is a degree- $2d$ polynomial in ξ . Hence, the claim would follow from the concentration bound for low-degree polynomials, i.e., Theorem 2.14, if we can appropriately upper bound the ℓ_2 -norm of the polynomial $\text{Inf}_i(P_\xi)$. So, to prove Claim 5.3, it suffices to show that

$$\|\text{Inf}_i(P_\xi)\|_2 \leq C_4(\alpha)^d \cdot \text{Inf}_i(P) \tag{1}$$

By the triangle inequality for norms we have that,

$$\|\text{Inf}_i(P_\xi)\|_2 = \left\| \sum_{\sigma_T: (\sigma_T)_i \neq 0} \widehat{P}_\xi(\sigma_T)^2 \right\|_2 \leq \sum_{\sigma_T: (\sigma_T)_i \neq 0} \left\| \widehat{P}_\xi(\sigma_T)^2 \right\|_2$$

Since $\widehat{P}_\xi(\sigma_T)$ is a degree- d polynomial, the moment bound for low-degree polynomials, i.e., Theorem 2.13, yields that

$$\left\| \widehat{P}_\xi(\sigma_T)^2 \right\|_2 = \left\| \widehat{P}_\xi(\sigma_T) \right\|_4^2 \leq C_4(\alpha)^d \left\| \widehat{P}_\xi(\sigma_T) \right\|_2^2$$

and hence

$$\begin{aligned} \|\text{Inf}_i(P_\xi)\|_2 &\leq C_4(\alpha)^d \sum_{\sigma_T: (\sigma_T)_i \neq 0} \left\| \widehat{P}_\xi(\sigma_T) \right\|_2^2 \\ &= C_4(\alpha)^d \sum_{\sigma_T: (\sigma_T)_i \neq 0} \mathbb{E}_\xi \left[\widehat{P}_\xi(\sigma_T)^2 \right] \\ &= C_4(\alpha)^d \cdot \mathbb{E}_\xi [\text{Inf}_i(P_\xi)] \\ &= C_4(\alpha)^d \cdot \text{Inf}_i(P) \end{aligned}$$

where the last equality follows from Lemma 2.11. Thus, Equation (1) and the claim follows from Theorem 2.14. \square

Proof of Lemma 5.2. Let $P \in L^2(\mathcal{A}^n, \mu_A^{\otimes n})$ be the given degree- d multilinear polynomial with $\text{Var}[P] \leq 1$. From part (iii) of Fact 2.10, we have that $\text{Inf}(P) \leq d$. Let $H_0 \subset [n]$ be the set of indices $i \in [n]$ such that $\text{Inf}_i(P) \geq \beta$. Since, $d \geq \text{Inf}(P) \geq \sum_i \text{Inf}_i(P)$, we have that $|H_0| \leq d/\beta$. We will choose β as a suitable constant less than τ , but with no dependence on n .

Fix $H \supseteq H_0$ and let $T = [n] \setminus H$. From Claim 5.3, we have for any $i \in T$, that $\Pr_\xi [\text{Inf}_i(P_\xi) > r \cdot C_4(\alpha)^d \cdot \text{Inf}_i(P)] \leq \exp(-\Omega(c \cdot r^{1/d}))$. However, to prove that $\text{Inf}_i(P_\xi) \leq \tau$ for all $i \in T$, with high probability, we cannot simply use a naïve union bound over all $i \in T$, as that will introduce a dependence of n in β and thereby in h . Instead, we use a *bucketing* argument, as done in [DSTW10], as follows:

We partition the indices $i \in T$ into buckets $\{B_j\}_{j \in \mathbb{N}}$ as $B_j = \left\{ i \in T : \text{Inf}_i(P) \in \left(\frac{\beta}{2^{j+1}}, \frac{\beta}{2^j} \right] \right\}$. Since $\text{Inf}(P) \leq d$, we have that $|B_j| \leq 2^{j+1}d/\beta$. For all $i \in B_j$, we use the concentration $\Pr_\xi [\text{Inf}_i(P_\xi) \leq r \cdot C_4(\alpha)^d \cdot \text{Inf}_i(P)] \geq 1 - \exp(-c \cdot r^{1/d})$ by choosing $r = \frac{\tau \cdot 2^j}{\beta \cdot C_4(\alpha)^d}$. We then do a union bound over all the buckets. Thus, we get that,

$$\begin{aligned} \Pr_\xi [\forall i \in T : \text{Inf}_i(P_\xi(\mathbf{x}_T)) \leq \tau] &\geq 1 - \sum_{j=0}^{\infty} \Pr_\xi [\exists i \in B_j : \text{Inf}_i(P_\xi(\mathbf{x}_T)) > \tau] \\ &\geq 1 - \sum_{j=0}^{\infty} \exp \left(-c \left(\frac{\tau \cdot 2^j}{\beta \cdot C_4(\alpha)^d} \right)^{1/d} \right) \cdot \frac{2^{j+1}d}{\beta} \end{aligned}$$

It can be verified that for $\frac{1}{\beta} = \frac{(2 \cdot C_4(\alpha))^d}{c^d \cdot \tau} \cdot \log \left(\frac{(2 \cdot C_4(\alpha))^d}{c^d \cdot \tau} \right)^d$ it holds that,

$$\sum_{j=0}^{\infty} \exp \left(-c \left(\frac{\tau \cdot 2^j}{\beta \cdot C_4(\alpha)^d} \right)^{1/d} \right) \cdot \frac{2^{j+1}d}{\beta} \leq \tau$$

Thus, we have the regularity lemma as desired with $|H_0| \leq h = \frac{d}{\beta} = \frac{d}{\tau} \cdot \left(\frac{C_4(\alpha)}{\alpha} \log \frac{C_4(\alpha)}{\alpha \cdot d \cdot \tau} \right)^{O(d)}$ which is a constant that depends on d, τ and $\alpha \stackrel{\text{def}}{=} \alpha(\mu_A)$. \square

5.2 Joint Regularity Lemma

In this section, we use Lemma 5.2 to prove the joint regularity lemma, namely Lemma 5.1.

Proof of Lemma 5.1. We have $P \in L^2(\mathcal{A}^n, \mu_A^{\otimes n})$ and $Q \in L^2(\mathcal{B}^n, \mu_B^{\otimes n})$ satisfying $\sum_{|\sigma|>d} \widehat{P}(\sigma)^2 \leq \eta$, $\sum_{|\sigma|>d} \widehat{Q}(\sigma)^2 \leq \eta$, and $\text{Var}[P] \leq 1$ and $\text{Var}[Q] \leq 1$. First, we split P and Q into low and high degree components. That is, $P(\mathbf{x}) = P^\ell(\mathbf{x}) + P^h(\mathbf{x})$ and $Q(\mathbf{y}) = Q^\ell(\mathbf{y}) + Q^h(\mathbf{y})$, where $P^\ell(\mathbf{x})$ and $Q^\ell(\mathbf{y})$ contain all the monomials of degree at most d in $P(\mathbf{x})$ and $Q(\mathbf{y})$ respectively. Note that $\text{Var}[P^\ell] \leq \text{Var}[P] \leq 1$. Similarly, $\text{Var}[Q^\ell] \leq 1$.

We apply the regularity lemma for degree- d functions (Lemma 5.2), with parameter τ equal to $\tau/4$, on functions P^ℓ and Q^ℓ separately, to obtain subsets $H_A, H_B \subseteq [n]$ respectively. The subset H is then obtained as $H_A \cup H_B$. Note that, $|H| \leq h((\mathcal{A}, \mu_A), d, \tau/4) + h((\mathcal{B}, \mu_B), d, \tau/4)$, which is a computable in terms of the parameters of the problem, but more importantly has no dependence on n .

From Lemma 5.2, we know that for $T = [n] \setminus H$ (note that $H \supseteq H_A$ and $H \supseteq H_B$),

$$\Pr_{\xi \sim \mu_A^{\otimes |H|}} [\forall i \in T : \text{Inf}_i(P_\xi^\ell(\mathbf{x}_T)) \leq \tau/4] \geq 1 - \tau/4 \quad (2)$$

$$\Pr_{\xi \sim \mu_B^{\otimes |H|}} [\forall i \in T : \text{Inf}_i(Q_\xi^\ell(\mathbf{y}_T)) \leq \tau/4] \geq 1 - \tau/4 \quad (3)$$

Now, we show that after adding P^h to P^ℓ , the influences $\text{Inf}_i(P_\xi(\mathbf{x}_T))$ are still upper bounded by τ , with high probability over ξ .

$$\begin{aligned} \text{Inf}_i(P_\xi(\mathbf{x}_T)) &= \sum_{\sigma_T: (\sigma_T)_i \neq 0} \left(\sum_{\sigma_H} \widehat{P}(\sigma_H \circ \sigma_T) \cdot \chi_{\sigma_H}(\xi) \right)^2 \\ &= \sum_{\sigma_T: (\sigma_T)_i \neq 0} \left(\sum_{\sigma_H} \widehat{P}^\ell(\sigma_H \circ \sigma_T) \cdot \chi_{\sigma_H}(\xi) + \sum_{\sigma_H} \widehat{P}^h(\sigma_H \circ \sigma_T) \cdot \chi_{\sigma_H}(\xi) \right)^2 \\ &\leq 2 \cdot \sum_{\sigma_T: (\sigma_T)_i \neq 0} \left(\sum_{\sigma_H} \widehat{P}^\ell(\sigma_H \circ \sigma_T) \cdot \chi_{\sigma_H}(\xi) \right)^2 + \left(\sum_{\sigma_H} \widehat{P}^h(\sigma_H \circ \sigma_T) \cdot \chi_{\sigma_H}(\xi) \right)^2 \\ &= 2 \cdot (\text{Inf}_i(P_\xi^\ell(\mathbf{x}_T)) + \text{Inf}_i(P_\xi^h(\mathbf{x}_T))) \end{aligned} \quad (4)$$

Since $\mathbb{E}_\xi [\text{Var}(P_\xi^h(\mathbf{x}_T))] \leq \text{Var}(P^h(\mathbf{x}_T)) \leq \eta$ (see Lemma 2.11), we have by Markov's inequality that,

$$\Pr_{\xi \sim \mu_A^{\otimes |H|}} [\text{Var}(P_\xi^h(\mathbf{x}_T)) \leq 4\eta/\tau] \geq 1 - \tau/4$$

Since for all $i \in T$, we have $\text{Inf}_i(P_\xi^h(\mathbf{x}_T)) \leq \text{Var}(P_\xi^h(\mathbf{x}_T))$ (see Fact 2.10), we get that

$$\Pr_{\xi \sim \mu_A^{\otimes |H|}} [\forall i \in T : \text{Inf}_i(P_\xi^h(\mathbf{x}_T)) \leq 4\eta/\tau] \geq 1 - \tau/4 \quad (5)$$

We will choose $\eta = (\tau/4)^2$, and thus, by union bound (using Equations 4, 3 and 5), we have that,

$$\Pr_{\xi \sim \mu_A^{\otimes |H|}} [\forall i \in T : \text{Inf}_i(P_\xi(\mathbf{x}_T)) \leq \tau] \geq 1 - \tau/2 > 1 - \tau$$

By exactly same flow of calculations for $Q(\mathbf{y})$, we can have,

$$\Pr_{\xi \sim \mu_B^{\otimes |H|}} [\forall i \in T : \text{Inf}_i(Q_\xi(\mathbf{y}_T)) \leq \tau] \geq 1 - \tau/2 > 1 - \tau$$

This completes the proof of Lemma 5.1. \square

6 Applying correlation bounds for low-influence functions

The third step in our approach is to use *correlation bounds for low-influence functions* obtained from the invariance principle [MOO05, Mos10], to convert the functions $f_1 : \mathcal{A}^n \rightarrow [-1, 1]$ and $g_1 : \mathcal{B}^n \rightarrow [-1, 1]$ into functions $f_2 : \mathcal{A}^h \times \mathbb{R} \rightarrow [-1, 1]$ and $g_2 : \mathcal{B}^h \times \mathbb{R} \rightarrow [-1, 1]$ using the following lemma.

Lemma 6.1 (Applying correlation bounds for low-influence functions). *Let $(\mathcal{A} \times \mathcal{B}, \mu)$ be a joint probability space. Let $\gamma > 0$ be any given constant parameter. There exists a $\tau \stackrel{\text{def}}{=} \tau((\mathcal{A} \times \mathcal{B}, \mu), \gamma) > 0$ such that the following holds:*

For all functions $f_1 : \mathcal{A}^n \rightarrow [-1, 1]$ and $g_1 : \mathcal{B}^n \rightarrow [-1, 1]$, and a subset $H \subseteq [n]$ with $|H| = h$, such that the restrictions of the functions f_1 and g_1 obtained by evaluating the coordinates in H according to distribution μ , satisfy the following (where we denote $T = [n] \setminus H$),

- *With probability at least $1 - \tau$ over $\xi \sim \mu_{\mathcal{A}}^{\otimes h}$, the restriction $(f_1)_{\xi}(\mathbf{x}_T)$ is such that for all $i \in T$, it is the case that $\text{Inf}_i((f_1)_{\xi}(\mathbf{x}_T)) \leq \tau$*
- *With probability at least $1 - \tau$ over $\xi \sim \mu_{\mathcal{B}}^{\otimes h}$, the restriction $(g_1)_{\xi}(\mathbf{x}_T)$ is such that for all $i \in T$, it is the case that $\text{Inf}_i((g_1)_{\xi}(\mathbf{x}_T)) \leq \tau$*

There exist functions $f_2 : \mathcal{A}^h \times \mathbb{R} \rightarrow [-1, 1]$ and $g_2 : \mathcal{B}^h \times \mathbb{R} \rightarrow [-1, 1]$, such that,

$$\mathbb{E}_{\mathbf{x} \sim \mu_{\mathcal{A}}^{\otimes n}} f_1(\mathbf{x}) = \mathbb{E}_{\substack{\mathbf{x} \sim \mu_{\mathcal{A}}^{\otimes h} \\ r_A \sim \mathcal{N}(0,1)}} f_2(\mathbf{x}, r_A) \quad \text{and} \quad \mathbb{E}_{\mathbf{y} \sim \mu_{\mathcal{B}}^{\otimes n}} g_1(\mathbf{y}) = \mathbb{E}_{\substack{\mathbf{y} \sim \mu_{\mathcal{B}}^{\otimes h} \\ r_B \sim \mathcal{N}(0,1)}} g_2(\mathbf{y}, r_B)$$

and,

$$\mathbb{E}_{\substack{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes h} \\ (r_A, r_B) \sim \mathcal{G}(\rho)}} [f_2(\mathbf{x}, r_A) \cdot g_2(\mathbf{y}, r_B)] \geq \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n}} [f_1(\mathbf{x}) \cdot g_1(\mathbf{y})] - \gamma$$

Additionally, f_2 and g_2 will have the following special form: there exist functions $f'_2 : \mathcal{A}^h \rightarrow \mathbb{R}$ and $g'_2 : \mathcal{B}^h \rightarrow \mathbb{R}$ such that,

$$f_2(\mathbf{x}, r) = \begin{cases} 1 & r \geq f'_2(\mathbf{x}) \\ -1 & r < f'_2(\mathbf{x}) \end{cases} \quad \text{and} \quad g_2(\mathbf{y}, r) = \begin{cases} 1 & r \geq g'_2(\mathbf{y}) \\ -1 & r < g'_2(\mathbf{y}) \end{cases}$$

Also, one may take $\tau = \gamma^{O(\frac{\log(1/\gamma)\log(1/\alpha)}{(1-\rho)\gamma})}$, where $\rho = \rho(\mathcal{A}, \mathcal{B}; \mu)$ and $\alpha \stackrel{\text{def}}{=} \alpha(\mu)$ is the minimum non-zero probability in μ .

As mentioned before, the main technical tool in proving Lemma 6.1 is a result about correlation bounds for low influence functions (which are generalizations of the ‘Majority is Stablest’ theorem). Before we state that theorem, we need the following definition, which is a slightly modified version of Definition 1.12 in [Mos10].

Definition 6.2 (Gaussian stability). *Let Φ be the cumulative distribution function (CDF) of a standard $\mathcal{N}(0, 1)$ Gaussian. Given $\rho \in [-1, 1]$ and $\mu, \nu \in [-1, 1]$, we define,*

$$\begin{aligned} \bar{\Gamma}_{\rho}(\mu, \nu) &= \mathbb{E}[\bar{P}_{\mu}(X) \cdot \bar{Q}_{\nu}(Y)] \\ \underline{\Gamma}_{\rho}(\mu, \nu) &= -\mathbb{E}[\bar{P}_{\mu}(X) \cdot \bar{Q}_{-\nu}(Y)] \end{aligned}$$

where (X, Y) is distributed according to $\mathcal{G}(\rho)$ and

$$\bar{P}_{\mu}(X) = \begin{cases} 1 & X \leq \Phi^{-1}(\frac{1+\mu}{2}) \\ -1 & \text{otherwise} \end{cases} \quad \text{and} \quad \bar{Q}_{\nu}(Y) = \begin{cases} 1 & Y \leq \Phi^{-1}(\frac{1+\nu}{2}) \\ -1 & \text{otherwise} \end{cases}$$

Note that for $(X, Y) \sim \mathcal{G}(\rho)$, we have that,

$$\mathbb{E}_X[\bar{P}_{\mu}(X)] = \mu \quad \text{and} \quad \mathbb{E}_Y[\bar{Q}_{\nu}(Y)] = \nu = \mathbb{E}_Y[-\bar{Q}_{-\nu}(Y)]$$

With this definition in hand, we can state the correlation bounds for low influential functions that are obtained from invariance principle.

Theorem 6.3 (Correlation bounds from invariance principle; [MCO05, Mos10]). *Let $(\mathcal{A} \times \mathcal{B}, \mu)$ be a joint probability space. As before, let $\alpha = \alpha(\mu)$ be the minimum probability of any atom in $\mathcal{A} \times \mathcal{B}$. Let $\rho = \rho(\mathcal{A}, \mathcal{B}; \mu)$ be the maximal correlation of the joint probability space (see Definition 2.16).*

Then, for all $\varepsilon > 0$, there exists $\tau \stackrel{\text{def}}{=} \tau((\mathcal{A} \times \mathcal{B}, \mu), \varepsilon) > 0$ such that if

$$P : \mathcal{A}^n \rightarrow [-1, 1] \quad \text{and} \quad Q : \mathcal{B}^n \rightarrow [-1, 1]$$

satisfy $\text{Inf}_i(P) \leq \tau$ and $\text{Inf}_i(Q) \leq \tau$ for all $i \in [n]$, then

$$\underline{\Gamma}_\rho \left(\mathbb{E}_{\mathbf{x}}[P(\mathbf{x})], \mathbb{E}_{\mathbf{y}}[Q(\mathbf{y})] \right) - \varepsilon \leq \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n}} [P(\mathbf{x})Q(\mathbf{y})] \leq \overline{\Gamma}_\rho \left(\mathbb{E}_{\mathbf{x}}[P(\mathbf{x})], \mathbb{E}_{\mathbf{y}}[Q(\mathbf{y})] \right) + \varepsilon$$

Furthermore, one may take

$$\tau = \varepsilon^{O\left(\frac{\log(1/\varepsilon)\log(1/\alpha)}{(1-\rho)\varepsilon}\right)}$$

Intuitively, this theorem says that if P and Q are low-influential, then their correlation is not much more than that of appropriate threshold functions applied on ρ -correlated gaussians. With this tool in hand, we are now ready to prove Lemma 6.1.

Proof of Lemma 6.1. Suppose we have $f_1 : \mathcal{A}^n \rightarrow [-1, 1]$ and $g_1 : \mathcal{B}^n \rightarrow [-1, 1]$, and a subset $H \subseteq [n]$ with $|H| = h$, such that the restrictions of the functions f_1 and g_1 obtained by evaluating the coordinates in H according to distribution μ , satisfy the properties as stated in the lemma. We construct function $f_2 : \mathcal{A}^h \times \mathbb{R} \rightarrow [-1, 1]$ and $g_2 : \mathcal{B}^h \times \mathbb{R} \rightarrow [-1, 1]$ by replacing the functions obtained after restricting the variables in H by appropriate threshold functions acting on ρ -correlated gaussians, namely,

$$\begin{aligned} \forall (\mathbf{x}, r) \in \mathcal{A}^h \times \mathbb{R} & : f_2(\mathbf{x}, r) = \overline{P}_{\nu_1}(r) \quad \text{where} \quad \nu_1 \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{x}_T \sim \mu_A^{\otimes n-h}} [f_1(\mathbf{x}_H \leftarrow \mathbf{x}, \mathbf{x}_T)] \\ \forall (\mathbf{y}, r) \in \mathcal{B}^h \times \mathbb{R} & : g_2(\mathbf{y}, r) = \overline{Q}_{\nu_2}(r) \quad \text{where} \quad \nu_2 \stackrel{\text{def}}{=} \mathbb{E}_{\mathbf{y}_T \sim \mu_B^{\otimes n-h}} [g_1(\mathbf{y}_H \leftarrow \mathbf{y}, \mathbf{y}_T)] \end{aligned}$$

where \overline{P}_ν and \overline{Q}_ν are as defined in Definition 6.2.⁷

It follows from definition, that $\mathbb{E}[f_2(\mathbf{x}, r)] = \mathbb{E}[f_1(\mathbf{x})]$ and $\mathbb{E}[g_2(\mathbf{y}, r)] = \mathbb{E}[g_1(\mathbf{y})]$. That is, this process has not changed the individual means of f_1 and g_1 . We now need to prove that the correlation is not hurt by much. From Lemma 5.1 and a simple union bound, we know that with probability $1 - 2\tau$, a random restriction $(\mathbf{x}_H, \mathbf{y}_H)$ for the coordinates in H is such that,

$$\forall i \in T \quad : \text{Inf}_i((f_1)_{\mathbf{x}_H}(\mathbf{x}_T)) \leq \tau \quad \text{and} \quad \text{Inf}_i((g_1)_{\mathbf{y}_H}(\mathbf{y}_T)) \leq \tau$$

Let's call all the tuples $(\mathbf{x}_H, \mathbf{y}_H)$ for which the above happens as 'good'.

⁷For simplicity, we will abuse notations in the following sense: when we say $f_1(\mathbf{x})$, we mean $\mathbf{x} \in \mathcal{A}^n$, but when we say $f_2(\mathbf{x}, r)$, we mean $\mathbf{x} \in \mathcal{A}^h$ and $r \in \mathbb{R}$.

$$\begin{aligned}
& \mathbb{E}_{\mathbf{x}, \mathbf{y}} f_1(\mathbf{x})g_1(\mathbf{y}) \\
&= \mathbb{E}_{\mathbf{x}_H, \mathbf{y}_H} \left[\mathbb{E}_{\mathbf{x}_T, \mathbf{y}_T} f_1(\mathbf{x}_H, \mathbf{x}_T) \cdot g_1(\mathbf{y}_H, \mathbf{y}_T) \right] \\
&= \Pr[(\mathbf{x}_H, \mathbf{y}_H) \text{ is not 'good'}] \cdot \mathbb{E}_{\mathbf{x}_H, \mathbf{y}_H} \left[\mathbb{E}_{\mathbf{x}_T, \mathbf{y}_T} f_1(\mathbf{x}_H, \mathbf{x}_T) \cdot g_1(\mathbf{y}_H, \mathbf{y}_T) \middle| (\mathbf{x}_H, \mathbf{y}_H) \text{ is not 'good'} \right] \\
&\quad + \Pr[(\mathbf{x}_H, \mathbf{y}_H) \text{ is 'good'}] \cdot \mathbb{E}_{\mathbf{x}_H, \mathbf{y}_H} \left[\mathbb{E}_{\mathbf{x}_T, \mathbf{y}_T} f_1(\mathbf{x}_H, \mathbf{x}_T) \cdot g_1(\mathbf{y}_H, \mathbf{y}_T) \middle| (\mathbf{x}_H, \mathbf{y}_H) \text{ is 'good'} \right] \\
&\leq \Pr[(\mathbf{x}_H, \mathbf{y}_H) \text{ is not 'good'}] \cdot 1 \\
&\quad + \Pr[(\mathbf{x}_H, \mathbf{y}_H) \text{ is 'good'}] \cdot \mathbb{E}_{\mathbf{x}_H, \mathbf{y}_H} \left[\mathbb{E}_{r_A, r_B} f_2(\mathbf{x}_H, r_A) \cdot g_2(\mathbf{y}_H, r_B) + \varepsilon \middle| (\mathbf{x}_H, \mathbf{y}_H) \text{ is 'good'} \right] \\
&= \Pr[(\mathbf{x}_H, \mathbf{y}_H) \text{ is not 'good'}] \cdot \left(1 - \mathbb{E}_{\mathbf{x}_H, \mathbf{y}_H} \left[\mathbb{E}_{r_A, r_B} f_2(\mathbf{x}_H, r_A) \cdot g_2(\mathbf{y}_H, r_B) + \varepsilon \middle| (\mathbf{x}_H, \mathbf{y}_H) \text{ is not 'good'} \right] \right) \\
&\quad + \mathbb{E}_{\mathbf{x}_H, \mathbf{y}_H} \left[\mathbb{E}_{r_A, r_B} f_2(\mathbf{x}_H, r_A) \cdot g_2(\mathbf{y}_H, r_B) + \varepsilon \right] \\
&\leq \mathbb{E}_{\mathbf{x}_H, \mathbf{y}_H} \left[\mathbb{E}_{r_A, r_B} f_2(\mathbf{x}_H, r_A) \cdot g_2(\mathbf{y}_H, r_B) \right] + 2\tau \cdot (2 - \varepsilon) + \varepsilon \\
&\leq \mathbb{E}_{\mathbf{x}_H, \mathbf{y}_H} \left[\mathbb{E}_{r_A, r_B} f_2(\mathbf{x}_H, r_A) \cdot g_2(\mathbf{y}_H, r_B) \right] + 2\varepsilon
\end{aligned}$$

Step 3 above is due to the definition of f_2 and g_2 and Theorem 6.3. The last step follows because $\tau \ll \varepsilon$, and so we can upper bound $2\tau \cdot (2 - \varepsilon) \leq \varepsilon$.

Thus, finally we choose $\varepsilon = \gamma/2$ for Theorem 6.3, and we get $\tau = \tau(\gamma)$ accordingly, thereby getting the final requirement of Lemma 6.1, that is,

$$\mathbb{E}_{\substack{(\mathbf{x}_H, \mathbf{y}_H) \sim \mu^{\otimes h} \\ (r_A, r_B) \sim \mathcal{G}(\rho)}} f_2(\mathbf{x}_H, r_A) \cdot g_2(\mathbf{y}_H, r_B) \geq \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n}} f_1(\mathbf{x}) \cdot g_1(\mathbf{y}) - \gamma$$

□

7 Simulating Correlated Gaussians

In this section, we use the technique due to Witsenhausen [Wit75] which shows that for any joint probability space $(\mathcal{A} \times \mathcal{B}, \mu)$ with maximal correlation ρ , Alice and Bob can non-interactively simulate ρ -correlated gaussians upto arbitrarily small 2-dimensional Kolmogorov distance. We obtain the following lemma.

Lemma 7.1 (Witsenhausen's rounding). *Let $(\mathcal{A} \times \mathcal{B}, \mu)$ be a joint probability space, and let $\rho = \rho(\mathcal{A}, \mathcal{B}; \mu)$ be its maximal correlation. Let $\zeta > 0$ be any given parameter. Then, there exists $w \stackrel{\text{def}}{=} w((\mathcal{A} \times \mathcal{B}, \mu), \zeta) \in \mathbb{N}$, such that the following holds:*

For all functions $f_2 : \mathcal{A}^h \times \mathbb{R} \rightarrow [-1, 1]$ and $g_2 : \mathcal{B}^h \times \mathbb{R} \rightarrow [-1, 1]$ having the following special form: there exist functions $f'_2 : \mathcal{A}^h \rightarrow \mathbb{R}$ and $g'_2 : \mathcal{B}^h \rightarrow \mathbb{R}$ such that,

$$f_2(\mathbf{x}, r) = \begin{cases} 1 & r \geq f'_2(\mathbf{x}) \\ -1 & r < f'_2(\mathbf{x}) \end{cases} \quad \text{and} \quad g_2(\mathbf{y}, r) = \begin{cases} 1 & r \geq g'_2(\mathbf{y}) \\ -1 & r < g'_2(\mathbf{y}) \end{cases}$$

there exist functions $f_3 : \mathcal{A}^{h+w} \rightarrow [-1, 1]$ and $g_3 : \mathcal{B}^{h+w} \rightarrow [-1, 1]$, such that,

$$\left| \mathbb{E}_{\mathbf{x} \sim \mu_{\mathcal{A}}^{\otimes (h+w)}} f_3(\mathbf{x}) - \mathbb{E}_{\substack{\mathbf{x} \sim \mu_{\mathcal{A}}^{\otimes h} \\ r_A \sim \mathcal{N}(0,1)}} [f_2(\mathbf{x}, r_A)] \right| \leq \zeta \quad \text{and} \quad \left| \mathbb{E}_{\mathbf{y} \sim \mu_{\mathcal{B}}^{\otimes (h+w)}} g_3(\mathbf{y}) - \mathbb{E}_{\substack{\mathbf{x} \sim \mu_{\mathcal{B}}^{\otimes h} \\ r_B \sim \mathcal{N}(0,1)}} [g_2(\mathbf{y}, r_B)] \right| \leq \zeta$$

and,

$$\left| \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes (h+w)}} [f_3(\mathbf{x}) \cdot g_3(\mathbf{y})] - \mathbb{E}_{\substack{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes h} \\ (r_A, r_B) \sim \mathcal{G}(\rho)}} [f_2(\mathbf{x}, r_A) \cdot g_2(\mathbf{y}, r_B)] \right| \leq \zeta$$

In particular, one may take $w = O\left(\frac{1+\rho}{\alpha \cdot (1-\rho)^3 \cdot \zeta^2}\right)$, where $\alpha \stackrel{\text{def}}{=} \alpha(\mu)$ is the minimum non-zero probability in μ .

The main idea in obtaining the functions f_3 and g_3 is the technique of Witsenhausen [Wit75], of simulating ρ -correlated gaussians from many copies of $(\mathcal{A} \times \mathcal{B}, \mu)$.

Lemma 7.2 (Simulating gaussians [Wit75]). *Let $(\mathcal{A} \times \mathcal{B}, \mu)$ be a joint probability space, and let $\rho = \rho(\mathcal{A}, \mathcal{B}; \mu)$ be its maximal correlation. Let $\zeta > 0$ be any given parameter. Then, there exists $w \stackrel{\text{def}}{=}} w((\mathcal{A} \times \mathcal{B}, \mu), \zeta) \in \mathbb{N}$, such that the following holds,*

For all $\nu_1, \nu_2 \in [-1, +1]$, there exist functions $P_{\nu_1} : \mathcal{A}^w \rightarrow [-1, 1]$ and $Q_{\nu_2} : \mathcal{B}^w \rightarrow [-1, 1]$ such that $|\mathbb{E}[P_{\nu_1}(\mathbf{x})] - \nu_1| \leq \zeta/2$, $|\mathbb{E}[Q_{\nu_2}(\mathbf{y})] - \nu_2| \leq \zeta/2$ and

$$|\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes w}} [P_{\nu_1}(\mathbf{x})Q_{\nu_2}(\mathbf{y})] - \bar{\Gamma}_\rho(\nu_1, \nu_2)| \leq \zeta$$

In particular, one may take $w = O\left(\frac{1+\rho}{\alpha \cdot (1-\rho)^3 \cdot \zeta^2}\right)$, where $\alpha \stackrel{\text{def}}{=} \alpha(\mu)$.

Proof. Since $\rho = \rho(\mathcal{A}, \mathcal{B}; \mu)$, we have that there exist functions $f : \mathcal{A} \rightarrow \mathbb{R}$ and $g : \mathcal{B} \rightarrow \mathbb{R}$ such that $\mathbb{E}_{x \sim \mu_A} f(x) = \mathbb{E}_{y \sim \mu_B} g(y) = 0$, $\text{Var}(f) = \text{Var}(g) = 1$ and $\mathbb{E}_{(x, y) \sim \mu} [f(x) \cdot g(y)] = \rho$.

We define $F(\mathbf{x}) = \frac{\sum_{i=1}^w f(x_i)}{\sqrt{w}}$ and $G(\mathbf{y}) = \frac{\sum_{i=1}^w g(y_i)}{\sqrt{w}}$. And define P_{ν_1} and Q_{ν_2} as follows,

$$P_{\nu_1}(\mathbf{x}) = \begin{cases} 1 & F(\mathbf{x}) \leq \Phi^{-1}\left(\frac{1+\nu_1}{2}\right) \\ -1 & \text{otherwise} \end{cases} \quad \text{and} \quad Q_{\nu_2}(\mathbf{y}) = \begin{cases} 1 & G(\mathbf{y}) \leq \Phi^{-1}\left(\frac{1+\nu_2}{2}\right) \\ -1 & \text{otherwise} \end{cases}$$

We apply Lemma 2.22 for the pair of random variables $(f(x), g(y))$ with parameter ζ being $\zeta/4$, to obtain the appropriate w . It easily follows that, $|\mathbb{E}[P_{\nu_1}(\mathbf{x})] - \nu_1| \leq \zeta/2$ and $|\mathbb{E}[Q_{\nu_2}(\mathbf{y})] - \nu_2| \leq \zeta/2$ and

$$|\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes w}} [P_{\nu_1}(\mathbf{x})Q_{\nu_2}(\mathbf{y})] - \bar{\Gamma}_\rho(\nu_1, \nu_2)| \leq \zeta$$

□

We are now ready to prove Lemma 7.1.

Proof of Lemma 7.1. Given $(\mathcal{A} \times \mathcal{B}, \mu)$ and ζ , we obtain w as in Lemma 7.2. Given functions f_2 and g_2 , of the said form, we construct functions $f_3 : \mathcal{A}^{h+w} \rightarrow [-1, 1]$ and $g_3 : \mathcal{B}^{h+w} \rightarrow [-1, 1]$ by invoking Lemma 7.2 for every assignment to the first h variables with parameter ζ . In particular for every $\mathbf{x}_1 \in \mathcal{A}^h, \mathbf{x}_2 \in \mathcal{A}^w$, we define $f_3(\mathbf{x}_1, \mathbf{x}_2) = P_{f_2'(\mathbf{x}_1)}(\mathbf{x}_2)$. Similarly, for $\mathbf{y}_1 \in \mathcal{B}^h, \mathbf{y}_2 \in \mathcal{B}^w$, we define $g_3(\mathbf{y}_1, \mathbf{y}_2) = Q_{g_2'(\mathbf{y}_1)}(\mathbf{y}_2)$.

This gives us that $|\mathbb{E}[f_3(\mathbf{x})] - \mathbb{E}[f_2(\mathbf{x}, r_A)]| \leq \zeta/2$ and $|\mathbb{E}[g_3(\mathbf{y})] - \mathbb{E}[g_2(\mathbf{y}, r_B)]| \leq \zeta/2$ and,

$$\left| \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes (h+w)}} [f_3(\mathbf{x}) \cdot g_3(\mathbf{y})] - \mathbb{E}_{\substack{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes h} \\ (r_A, r_B) \sim \mathcal{G}(\rho)}} [f_2(\mathbf{x}, r_A) \cdot g_2(\mathbf{y}, r_B)] \right| \leq \zeta$$

Thus, we have f_3 and g_3 as desired. □

8 Putting it all together!

In this section we finally use all the lemmas we have developed to prove Theorem 3.1.

Proof of Theorem 3.1. Given $(\mathcal{A} \times \mathcal{B}, \mu)$ and $\delta > 0$ and functions $f : \mathcal{A}^n \rightarrow [-1, 1]$ and $g : \mathcal{B}^n \rightarrow [-1, 1]$, we wish to apply Lemma 6.1 with parameter $\gamma = \delta/3$ followed by Lemma 7.1 with parameter $\zeta = \delta/3$. Lemma 6.1 will dictate a value $\tau = \tau((\mathcal{A} \times \mathcal{B}, \mu), \gamma)$. We wish to apply the Joint regularity lemma (Lemma 5.1), with this parameter τ , which will dictate a value of

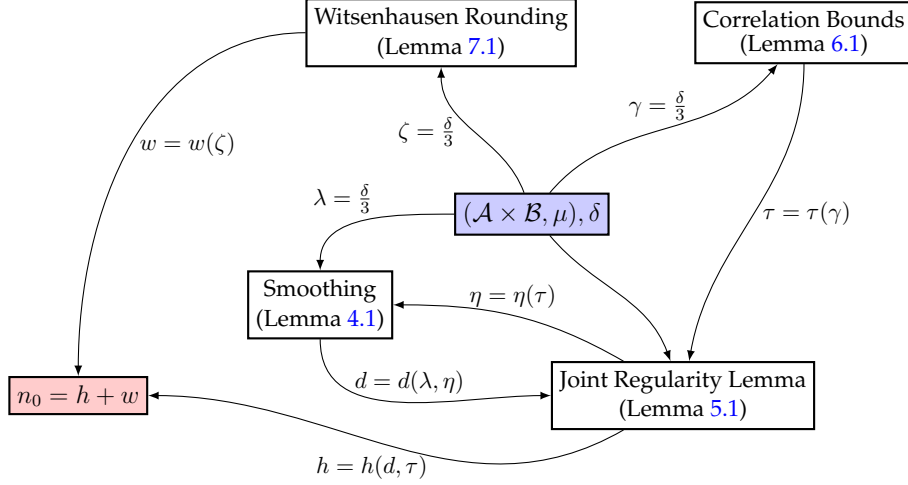


Figure 3: Dependency of parameters in the proof of Theorem 3.1

$\eta = \eta(\tau)$. Using this value of η , and $\lambda = \delta/3$, we apply the Smoothing lemma (Lemma 4.1), which will dictate a value of $d = d((\mathcal{A} \times \mathcal{B}, \mu), \lambda, \eta)$. We use this d to feed into the joint regularity lemma (Lemma 5.1), to obtain a value of h . The final value of n_0 is the sum of $h((\mathcal{A} \times \mathcal{B}, \mu), d, \tau)$ given by the joint regularity lemma (Lemma 5.1) and $w((\mathcal{A} \times \mathcal{B}, \mu), \zeta)$ given by Witsenhausen's rounding procedure (Lemma 7.1). This dependency of parameters is pictorially described in Figure 3 (the dependencies on $(\mathcal{A} \times \mathcal{B}, \mu)$ are suppressed, for sake of clarity). It can be shown by putting everything together that $n_0 = \exp\left(\text{poly}\left(\frac{1}{\delta}, \frac{1}{1-\rho}, \log\left(\frac{1}{\alpha}\right)\right)\right)$.

Once we have all the parameters set, we are now able to apply them to any pair of functions $f : \mathcal{A}^n \rightarrow [-1, 1]$ and $g : \mathcal{B}^n \rightarrow [-1, 1]$. In particular, we proceed as described in the overview (Section 3).

- Step 1:** We apply Lemma 4.1 to functions f and g with parameters λ and η as obtained above. This gives us a degree d and functions f_1 and g_1 , such that, $\sum_{|\sigma|>d} \widehat{f}(\sigma)^2 < \eta$ and $\sum_{|\sigma|>d} \widehat{g}(\sigma)^2 < \eta$.
- Step 2:** We apply the joint regularity lemma (Lemma 5.1) on functions f_1 and g_1 , with parameters d and τ as obtained above (note that, the conditions involving η are satisfied, because we chose precisely this η to be given to the Smoothing lemma). This gives us a subset $H \subseteq [n]$ such that $|H| \leq h$ and with high probability over restrictions to this subset H , the restricted versions of both f_1 and g_1 have all individual influences to be at most τ .
- Step 3:** We apply the correlation bounds result (Lemma 6.1) to functions f_1 and g_1 (note that all the conditions involving τ are satisfied already because we chose precisely this τ to be given to the joint regularity lemma). This gives us functions $f_2 : \mathcal{A}^h \times \mathbb{R} \rightarrow [-1, 1]$ and $g_2 : \mathcal{B}^h \times \mathbb{R} \rightarrow [-1, 1]$ of the form: there exist functions $f'_2 : \mathcal{A}^h \rightarrow \mathbb{R}$ and $g'_2 : \mathcal{B}^h \rightarrow \mathbb{R}$ such that,

$$f_2(\mathbf{x}, r) = \begin{cases} 1 & r \geq f'_2(\mathbf{x}) \\ -1 & r < f'_2(\mathbf{x}) \end{cases} \quad \text{and} \quad g_2(\mathbf{y}, r) = \begin{cases} 1 & r \geq g'_2(\mathbf{y}) \\ -1 & r < g'_2(\mathbf{y}) \end{cases}$$

- Step 4:** Functions f_2 and g_2 are exactly in the form for which Lemma 7.1 is applicable, which we use with parameters ζ as obtained above. This gives us functions $f_3 : \mathcal{A}^{h+w} \rightarrow [-1, 1]$ and $g_3 : \mathcal{B}^{h+w} \rightarrow [-1, 1]$.

Note that, $\mathbb{E} f = \mathbb{E} f_1 = \mathbb{E} f_2$ and $|\mathbb{E} f_3 - \mathbb{E} f_2| \leq \zeta = \delta/3$ and similarly $\mathbb{E} g = \mathbb{E} g_1 = \mathbb{E} g_2$ and

$|\mathbb{E} g_3 - \mathbb{E} g_2| \leq \zeta = \delta/3$. Moreover, we have from Lemmas 7.1, 6.1 and 4.1 that,

$$\begin{aligned}
\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes (h+w)}} [f_3(\mathbf{x}) \cdot g_3(\mathbf{y})] &\geq \mathbb{E}_{\substack{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes h} \\ (r_A, r_B) \sim \mathcal{G}(\rho)}} [f_2(\mathbf{x}) \cdot g_2(\mathbf{y})] - \zeta \\
&\geq \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n}} [f_1(\mathbf{x}) \cdot g_1(\mathbf{y})] - \gamma - \zeta \\
&\geq \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n}} [f(\mathbf{x}) \cdot g(\mathbf{y})] - \lambda - \gamma - \zeta \\
&= \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n}} [f(\mathbf{x}) \cdot g(\mathbf{y})] - \delta
\end{aligned}$$

Hence, taking $\tilde{f} = f_3$ and $\tilde{g} = g_3$, proves Theorem 3.1. \square

8.1 Generalizing to arbitrary binary targets

We now give a proof sketch of Theorem 2.3. Even though this is not a black-box application of Theorem 2.5, it follows the same proof steps. We highlight the main differences in this section.

We consider two cases, (I) $\mathbb{E}[UV] \geq \mathbb{E}[U] \cdot \mathbb{E}[V]$ and (II) $\mathbb{E}[UV] \leq \mathbb{E}[U] \cdot \mathbb{E}[V]$.

Case (I): $\mathbb{E}[UV] \geq \mathbb{E}[U] \cdot \mathbb{E}[V]$

We need to modify the GAP-BAL-MAX-INNER-PRODUCT problem 2.6, by replacing the conditions on $|\mathbb{E}[f(\mathbf{x})]|$ by $|\mathbb{E}[f(\mathbf{x})] - \mathbb{E}[U]|$, and similarly replacing the conditions on $|\mathbb{E}[g(\mathbf{y})]|$ by $|\mathbb{E}[g(\mathbf{y})] - \mathbb{E}[V]|$ and replacing ρ by $\mathbb{E}[UV]$. The reduction between GAP-NON-INT-SIM and GAP-BAL-MAX-INNER-PRODUCT works in almost exactly the same way.

It is easy to see that using the main technical theorem 3.1 and following the same proof as of Theorem 2.5, we also get decidability for GAP-NON-INT-SIM($(\mathcal{A} \times \mathcal{B}, \mu), (\mathcal{U} \times \mathcal{V}, \nu), \delta$).

Case (II): $\mathbb{E}[UV] \leq \mathbb{E}[U] \cdot \mathbb{E}[V]$

As in the previous case, we need to modify the GAP-BAL-MAX-INNER-PRODUCT problem 2.6, by replacing the conditions on $|\mathbb{E}[f(\mathbf{x})]|$ by $|\mathbb{E}[f(\mathbf{x})] - \mathbb{E}[U]|$, and similarly replacing the conditions on $|\mathbb{E}[g(\mathbf{y})]|$ by $|\mathbb{E}[g(\mathbf{y})] - \mathbb{E}[V]|$. The condition on $\mathbb{E}[f(\mathbf{x})g(\mathbf{y})]$ will however change as, $\mathbb{E}[f(\mathbf{x})g(\mathbf{y})] \leq \mathbb{E}[UV] + \delta$ in case (i) vs. $\mathbb{E}[f(\mathbf{x})g(\mathbf{y})] \geq \mathbb{E}[UV] + 4\delta$ in case (ii). The reduction between GAP-NON-INT-SIM and GAP-BAL-MAX-INNER-PRODUCT works in almost exactly the same way.

The main difference in this case however is that, we want each of the steps to ‘increase’ correlation by a small amount as opposed to ‘decrease’ the correlation. In particular, the main condition in Theorem 3.1 will change as follows,

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n_0}} [\tilde{f}(\mathbf{x}) \cdot \tilde{g}(\mathbf{y})] \leq \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n}} [f(\mathbf{x}) \cdot g(\mathbf{y})] + \delta$$

The steps of Smoothing (Lemma 4.1) and Joint Regularity (Lemma 5.1) and Witsenhausen rounding (Lemma 7.1) don’t need any modification as they approximately preserve the correlation in both directions. However, in the step of applying Correlation Bounds (Lemma 6.1), we need to use the lower bound of $\underline{\Gamma}_\rho(\cdot, \cdot)$ instead of the upper bound of $\overline{\Gamma}_\rho(\cdot, \cdot)$. In particular, the lemma will change slightly resulting in functions such that,

$$\mathbb{E}_{\substack{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes h} \\ (r_A, r_B) \sim \mathcal{G}(\rho)}} [f_2(\mathbf{x}, r_A) \cdot g_2(\mathbf{x}, r_B)] \leq \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mu^{\otimes n}} [f_1(\mathbf{x}) \cdot g_1(\mathbf{y})] + \gamma$$

Additionally, f_2 and g_2 will have the following special form: there exist functions $f'_2 : \mathcal{A}^h \rightarrow \mathbb{R}$ and $g'_2 : \mathcal{B}^h \rightarrow \mathbb{R}$ such that,

$$f_2(\mathbf{x}, r) = \begin{cases} 1 & r \geq f'_2(\mathbf{x}) \\ -1 & r < f'_2(\mathbf{x}) \end{cases} \quad \text{and} \quad g_2(\mathbf{y}, r) = \begin{cases} -1 & r \geq g'_2(\mathbf{y}) \\ 1 & r < g'_2(\mathbf{y}) \end{cases}$$

This structural difference in f_2 and g_2 affects the Witsenhausen Rounding step (Lemma 7.1) slightly, but it is easy to see that the same proof strategy works.

It is also easy to see that using this modified main theorem (analog of Theorem 3.1) and following the same proof steps as of Theorem 2.5, we also get decidability for GAP-NON-INT-SIM($(\mathcal{A} \times \mathcal{B}, \mu), (\mathcal{U} \times \mathcal{V}, \nu), \delta$) in this case.

9 Open Questions

In this work, we proved computable bounds on the non-interactive simulation of any 2×2 distribution. We now conclude with some interesting open questions.

The running time of our algorithm is at least doubly-exponential in the input size⁸. It would be very interesting to understand the computational complexity of the non-interactive simulation problem. We point out that the question of generating the best DSBS can be thought of as a tensored version of the following “MIN-BIPARTITE-BISECTION” problem: We are given a weighted bipartite graph $G = (L \cup R, E)$, and we wish to find a subset S of $L \cup R$ such that $S \cap L$ roughly contains half the vertices of L , and $S \cap R$ roughly contains half the vertices of R , while minimizing the total weight of edges crossing the cut (S, \bar{S}) . While it follows from [RST12] that MIN-BIPARTITE-BISECTION is hard to approximate, the same is not necessarily true about its tensored version.

Another interesting open question is to generalize our decidability results to larger alphabets, which seems to require new technical ideas. Indeed, our proof of Theorems 1.1 and 1.2 relied on the fact that for (X, Y) being correlated random Gaussians, the maximum possible agreement of any pair of ± 1 -valued functions $f(X)$ and $g(Y)$ is at most that of two appropriate dictator threshold functions $F(X_1)$ and $G(Y_1)$ where F only depends on the marginals of f (i.e., the probability that f takes the values -1 and $+1$), and similarly G only depends on the marginals of g . The analogous statement for the ternary case is not true. Namely, let $f(X), g(Y) \in \{0, 1, 2\}$, and assume that the marginals of f are $(1/3, 1/3, 1/3)$. Then, depending on whether the marginals of g are $(1/3, 1/3, 1/3)$ or $(1/2, 1/2, 0)$, the largest agreement of (f, g) would be achieved by very different functions f , assuming the “Standard Simplex Conjecture” (see [IM12] and Proposition 2.10 of [HMN15]). This example shows that in the ternary case Alice cannot replace f by a function of a very small number of copies without taking the marginals of Bob’s function g into account, and this is a major obstacle in generalizing our approach for proving Theorems 1.1 and 1.2 to larger alphabets.

Yet another interesting open question is to generalize our computability results to more than two players, which also seems to require new technical ideas.

Finally, it will be very interesting to see if these techniques could apply to other ‘tensorized’ problems. The most relevant problems seem to be (i) deciding a quantum version of our problem, namely that of local state transformation of quantum entanglement [Bei12, DB13] and (ii) approximately computing the entangled value of a 2-prover 1-round game ([KKM⁺11]; also see the open problem [ope]).

10 Acknowledgments

We thank Sudeep Kamath for explaining to us the state-of-the-art results in the information theory community, with regards to the problem of non-interactive simulation. We also thank Boaz Barak, Mohammad Bavarian and Mohsen Ghaffari for helpful discussions. We thank Matthew Coudron and Robin Kothari for pointing us to the related problems in the quantum literature.

⁸For constant values of δ and ρ , the running time is doubly-exponential in $2^{\text{poly}(\log m)}$. Here we think of the input as a bipartite graph with m edges. This follows because $\alpha \sim 1/m$.

References

- [AC93] Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography. part i: secret sharing. *IEEE Transactions on Information Theory*, 39(4), 1993. [1](#)
- [AC98] Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography. ii. cr capacity. *Information Theory, IEEE Transactions on*, 44(1):225–240, 1998. [1](#)
- [AH11] Per Austrin and Johan Håstad. Randomly supported independence and resistance. *SIAM Journal on Computing*, 40(1):1–27, 2011. [4](#), [9](#)
- [AL06] Noga Alon and Eyal Lubetzky. The shannon capacity of a graph and the independence numbers of its powers. *Information Theory, IEEE Transactions on*, 52(5):2172–2176, 2006. [2](#)
- [BDK05] W. Bryc, A. Dembo, and A. Kagan. On the maximum correlation coefficient. *Theory of Probability and its Applications*, 49(1):132–138, 2005. [10](#)
- [Bei12] Salman Beigi. A new quantum data processing inequality. *CoRR*, abs/1210.1689, 2012. [2](#), [24](#)
- [BG15] Salman Beigi and Amin Gohari. On the duality of additivity and tensorization. *arXiv preprint arXiv:1502.00827*, 2015. [2](#)
- [BGI14] Mohammad Bavarian, Dmitry Gavinsky, and Tsuyoshi Ito. On the role of shared randomness in simultaneous communication. In *Automata, Languages, and Programming*, pages 150–162. Springer, 2014. [1](#)
- [Bor85] Christer Borell. Geometric bounds on the ornstein-uhlenbeck velocity process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 70(1):1–13, 1985. [3](#)
- [BR86] Rabindra N Bhattacharya and Ramaswamy Ranga Rao. *Normal approximation and asymptotic expansions*, volume 64. SIAM, 1986. [11](#)
- [BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 748–757. IEEE, 2011. [2](#)
- [BS94] Gilles Brassard and Louis Salvail. Secret-key reconciliation by public discussion. In *advances in Cryptology—EUROCRYPT’93*, pages 410–423. Springer, 1994. [1](#)
- [BS15] Mark Braverman and Jon Schneider. Information complexity is computable. *arXiv preprint arXiv:1502.02971*, 2015. [2](#)
- [CDS08] Eric Chitambar, Runyao Duan, and Yaoyun Shi. Tripartite entanglement transformations and tensor rank. *Physical review letters*, 101(14):140502, 2008. [1](#)
- [CGMS14] Clement Canonne, Venkat Guruswami, Raghu Meka, and Madhu Sudan. Communication with imperfectly shared randomness. *ITCS*, 2014. [1](#)
- [CN00] Imre Csiszár and Prakash Narayan. Common randomness and secret key generation with a helper. *Information Theory, IEEE Transactions on*, 46(2):344–366, 2000. [1](#)
- [DB13] Payam Delgosha and Salman Beigi. Impossibility of local state transformation via hypercontractivity. *CoRR*, abs/1307.2747, 2013. [2](#), [24](#)
- [DB14] Payam Delgosha and Salman Beigi. Impossibility of local state transformation via hypercontractivity. *Communications in Mathematical Physics*, 332(1):449–476, 2014. [1](#)
- [DSTW10] Ilias Diakonikolas, Rocco A Servedio, Li-Yang Tan, and Andrew Wan. A regularity lemma, and low-weight approximators, for low-degree polynomial threshold functions. In *Computational Complexity (CCC), 2010 IEEE 25th Annual Conference on*, pages 211–222. IEEE, 2010. [3](#), [8](#), [12](#), [14](#), [15](#), [16](#)
- [Geb41] Hans Gebelein. Das statistische problem der korrelation als variations-und eigenwertproblem und sein zusammenhang mit der ausgleichsrechnung. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 21(6):364–379, 1941. [1](#), [10](#)

- [GK73] Peter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973. [1](#)
- [GKS16] Badih Ghazi, Pritish Kamath, and Madhu Sudan. Communication complexity of permutation-invariant functions. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1902–1921, 2016. [1](#)
- [GW95] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995. [2](#), [10](#)
- [Hir35] Hermann O Hirschfeld. A connection between correlation and contingency. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 31, pages 520–524. Cambridge Univ Press, 1935. [1](#), [10](#)
- [HMN15] Steven Heilman, Elchanan Mossel, and Joe Neeman. Standard simplices and pluralities are not the most noise stable. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, January 11-13, 2015*, page 255, 2015. [24](#)
- [IM12] Marcus Isaksson and Elchanan Mossel. Maximally stable gaussian partitions with discrete applications. *Israel Journal of Mathematics*, 189(1):347–396, 2012. [24](#)
- [KA12] Sudeep Kamath and Venkat Anantharam. Non-interactive simulation of joint distributions: The hirschfeld-gebelein-rényi maximal correlation and the hypercontractivity ribbon. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pages 1057–1064. IEEE, 2012. [2](#), [5](#)
- [KA15] Sudeep Kamath and Venkat Anantharam. On non-interactive simulation of joint distributions. *arXiv preprint arXiv:1505.00769*, 2015. [1](#), [2](#), [4](#), [5](#)
- [Kam15] Sudeep Kamath. Personal communication. 2015. [2](#)
- [KKM⁺11] Julia Kempe, Hirotada Kobayashi, Keiji Matsumoto, Ben Toner, and Thomas Vidick. Entangled games are hard to approximate. *SIAM J. Comput.*, 40(3):848–877, 2011. [2](#), [24](#)
- [KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O’Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM Journal on Computing*, 37(1):319–357, 2007. [11](#)
- [Lov79] László Lovász. On the shannon capacity of a graph. *Information Theory, IEEE Transactions on*, 25(1):1–7, 1979. [1](#)
- [Mau93] Ueli M Maurer. Secret key agreement by public discussion from common information. *Information Theory, IEEE Transactions on*, 39(3):733–742, 1993. [1](#)
- [MO04] Elchanan Mossel and Ryan O’Donnell. Coin flipping from a cosmic source: On error correction of truly random bits. *arXiv preprint math/0406504*, 2004. [1](#)
- [MOO05] Elchanan Mossel, Ryan O’Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 21–30. IEEE, 2005. [3](#), [12](#), [18](#), [19](#)
- [MOR⁺06] Elchanan Mossel, Ryan O’Donnell, Oded Regev, Jeffrey E Steif, and Benny Sudakov. Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse bonami-beckner inequality. *Israel Journal of Mathematics*, 154(1):299–336, 2006. [1](#)
- [MORS10] Kevin Matulef, Ryan O’Donnell, Ronitt Rubinfeld, and Rocco A Servedio. Testing halfspaces. *SIAM Journal on Computing*, 39(5):2004–2047, 2010. [11](#)
- [Mos10] Elchanan Mossel. Gaussian bounds for noise correlation of functions. *Geometric and Functional Analysis*, 19(6):1713–1756, 2010. [3](#), [12](#), [13](#), [18](#), [19](#)
- [Nie99] Michael A Nielsen. Conditions for a class of entanglement transformations. *Physical Review Letters*, 83(2):436, 1999. [1](#)

- [ope] OpenQIPProblemsWiki - All the Bell Inequalities. <http://qip.itp.uni-hannover.de/qiproblems/1>. Accessed: 2016-07-12. [2](#), [24](#)
- [Rén59] Alfréd Rényi. On measures of dependence. *Acta mathematica hungarica*, 10(3-4):441–451, 1959. [1](#), [10](#)
- [RST12] Prasad Raghavendra, David Steurer, and Madhur Tulsiani. Reductions between expansion problems. In *Computational Complexity (CCC), 2012 IEEE 27th Annual Conference on*, pages 64–73. IEEE, 2012. [24](#)
- [RW05] Renato Renner and Stefan Wolf. Simple and tight bounds for information reconciliation and privacy amplification. In *Advances in cryptology-ASIACRYPT 2005*, pages 199–216. Springer, 2005. [1](#)
- [Sha56] Claude E Shannon. The zero error capacity of a noisy channel. *Information Theory, IRE Transactions on*, 2(3):8–19, 1956. [1](#)
- [Wit75] Hans S Witsenhausen. On sequences of pairs of dependent random variables. *SIAM Journal on Applied Mathematics*, 28(1):100–113, 1975. [1](#), [2](#), [3](#), [10](#), [13](#), [20](#), [21](#)
- [Wol07] Pawel Wolff. Hypercontractivity of simple random variables. *Studia Mathematica*, 180(3):219–236, 2007. [9](#)
- [Wyn75] Aaron D. Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, 1975. [1](#)