Communication Complexity of Distributed High Dimensional Correlation Testing

K. R. Sahasranand and Himanshu Tyagi, *Senior Member, IEEE*

***Abstract*—We consider a two-party distributed hypothesis testing problem for correlated Gaussian random variables. For a d-dimensional random vector X and a scalar random variable Y, where X and Y are jointly Gaussian with an unknown correlation vector ρ, parties $\mathcal{P}_1$ and $\mathcal{P}_2$ observe independent copies of X and Y, respectively. The parties seek to test if their observations are correlated or not, namely they seek to test if $\|\rho\|_2$ exceeds τ or is it 0. To that end, they communicate interactively and declare the test output. We show that roughly order d/τ^2 bits of communication are sufficient and necessary for resolving the distributed correlation testing problem above. Furthermore, we establish a lower bound of roughly d^2/τ^2 bits for the communication needed for distributed estimation of ρ, implying that distributed correlation testing requires less communication than distributed estimation. Both our lower bounds for testing and estimation hold for an arbitrary d and interactive communication with shared randomness, while our distributed test requires only one-way communication with shared randomness. For the one-dimensional case, with one-way communication and with probability of one of the error-types fixed, our bounds are more refined in the dependence on the other error-type and are tight even in the constant.**

***Index Terms*—Gaussian correlation, high-dimensional statistics, hypercontractivity, hypothesis testing, interactive communication.**

I. Introduction

Parties $\mathcal{P}_1$ and $\mathcal{P}_2$ observe jointly Gaussian random variables X^n and Y^n, respectively, comprising independent and identically distributed (i.i.d.) samples (X_t, Y_t), $1 \leq t \leq n$, with $X_t \in \mathbb{R}^d$, $Y_t \in \mathbb{R}$, and such that $\mathbb{E}\left[Y_1 \mid X_1\right] = \rho^T X_1$. They communicate with each other to determine if their observations are correlated, i.e., to test if $\|\rho\|_2 \geq \tau$ or $\|\rho\|_2 = 0$. For a given probability of error requirement and an arbitrary large n, what is the minimum communication needed between the parties? (See Section II for a formal description.)

This problem of distributed correlation testing is an instance of a distributed hypothesis testing problem and it has been studied for several decades in the information theory literature starting with the seminal work [6] and closely followed by [33]. Most formulations in this literature focus on the tradeoff between the error exponent and communication rate per sample for simple binary hypothesis testing problems; see [22] for a survey. We remark that our setting differs from these earlier settings since we consider a composite hypothesis testing problem. Furthermore, we do not focus on the error exponent and allow arbitrarily large number of samples n. In particular, the error exponent can be shown to be 0 when we restrict to communication of rate 0 ($cf.$ [30]), which is an allowed regime for us since we can take as many samples as we like to minimize the communication.

The problem of distributed independence testing with multiple rounds of interactive communication is studied in [38], [39]. Similar problems with more general hypotheses or more elaborate communication models are considered in [40], [37], [23], [31]. The error exponent for the conditional independence testing problem is studied in [28], where both upper and lower bound for it are obtained. In another recent line of work [32], strong converse for testing against independence over a noisy channel is obtained using the technique in [36]. Recently, and subsequent to the publication of the initial version [29] of this paper, related problems were considered in various works. In [20], an improved upper bound on the error exponent for testing between two known positive Gaussian correlations is provided. The communication complexity of estimating one-dimensional Gaussian correlations is established in [19] and that of independence testing over discrete alphabet in the large sample regime is characterized in [9]. The tradeoff between communication complexity and sample complexity for detecting pairwise correlations is studied in [15]. A related line of recent work considers composite hypothesis testing under communication, privacy, and shared randomness constraints [3], [4], [1], [2]. However, the constraints are placed on each independent sample rather than on parties observing multiple correlated samples. In particular, none of the prior works consider our specific composite hypothesis testing problem.

Our main result is the characterization of the minimum communication needed for distributed correlation testing. Our proposed distributed test uses one-way communication, with shared randomness, and solves the d-dimensional problem by reducing it to the one-dimensional problem. This is done by multiplying the observation vectors X_ts of $\mathcal{P}_1$ with a scaled random sign vector. Specifically, for $d = 1$, our test entails the use of shared randomness to sample a vector that is close to $\mathcal{P}_1$'s overall observation $(X_1, ..., X_n)$, sending the identity of this vector to $\mathcal{P}_2$, and then $\mathcal{P}_2$ checking if its observation vector $(Y_1, ..., Y_n)$ is close to this vector as well. We show that this test requires roughly $\max\{(1/\tau^2)\log 1/\varepsilon, (1/\tau^2 - 1)\log 1/\delta\}$ bits of communication to get probabilities of false alarm and missed detection to be less than δ and ε, respectively, when n is sufficiently large. For a general d, noting that multiplying

The authors are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India. Email: {sahasranand, htyagi}@iisc.ac.in

This work was supported, in part, by a research grant from Robert Bosch Center for Cyber Physical Systems, IISc and in part by the Center for Networked Intelligence (CNI), a Cisco CSR initiative.

A preliminary version of this paper [29] was presented at the IEEE International Symposium on Information Theory, Vail, USA, 2018.

the observations with a random sign vector will yield a one-dimensional correlation testing problem with correlation roughly $\|\rho\|_2/\sqrt{d}$, we show that the d-dimensional problem can be resolved using roughly $(d/\tau^2)\max\{\log 1/\varepsilon, \log 1/\delta\}$ bits of communication.

Interestingly, we establish a lower bound that shows that the amount of one-way communication used by our protocol for $d = 1$ is optimal among all one-way communication protocols. We show this bound by using the notions of hypercontractivity and reverse hypercontractivity (*cf.* [12], [17], [10], [13], [8], [25]). We note that in [18] and [11], the notion of hypercontractivity has been used to show lower bounds in the context of agreement distillation, with and without communication respectively. We apply the hypercontractivity inequalities for obtaining a measure change between joint and product distributions in the context of hypothesis testing, a slightly different application in comparison to the contractivity properties of the Markov operator used in [18], [11]. Also, by using the tensorization property of the hypercontractivity ribbon, we extend the bound to a general d to obtain a lower bound on one-way communication of roughly $(d/\tau^2)\max\{\log 1/\varepsilon, \log 1/\delta\}$ bits.

Recently, a strong data processing inequality for interactive correlation estimation was derived in [19]. We invoke this result to show that roughly d/τ^2 bits of communication are needed even when interactive communication is allowed, rendering our proposed one-way communication protocol optimal among interactive protocols. We note that this bound is slightly weaker for one-way communication than the one obtained using hypercontractivity.

The related problem of correlation vector estimation was studied in [21]. In that work, an estimation protocol was given that uses roughly d^2/τ^2 bits of communication to estimate ρ within a mean squared error of τ^2. Clearly, directly using this estimate to test will not be communication optimal. However, a natural question arises: can we find a better distributed estimation protocol that will remain communication-optimal even for testing? We show that, in fact, d^2/τ^2 bits of communication are necessary for estimation, whereby estimate-and-test strategy is strictly suboptimal for testing. Of course, this does not rule out estimating a function of ρ and using it as a statistic for the test. Indeed, our reduction to the one-dimensional case implies that estimating $\|\rho\|_2$ and comparing it with a threshold is communication optimal (up to constants).

Our proposed test is practical. In fact, we have simulated a version with slightly different parameters than those presented in our theoretical analysis below; the empirical performance is depicted in Figure 1. A phase transition in probability of error can be seen clearly when we communicate a number of bits proportional[1] to d/τ^2.

The remainder of the paper is organized as follows. We present our problem formulation in the next section, followed by the main results in Section III. Our distributed correlation test as well as its analysis are presented in Section IV. The proof of our lower bounds for one-way communication is in Section V and for interactive communication is in Section VI. We conclude with a discussion and some extensions of our results in the final section.

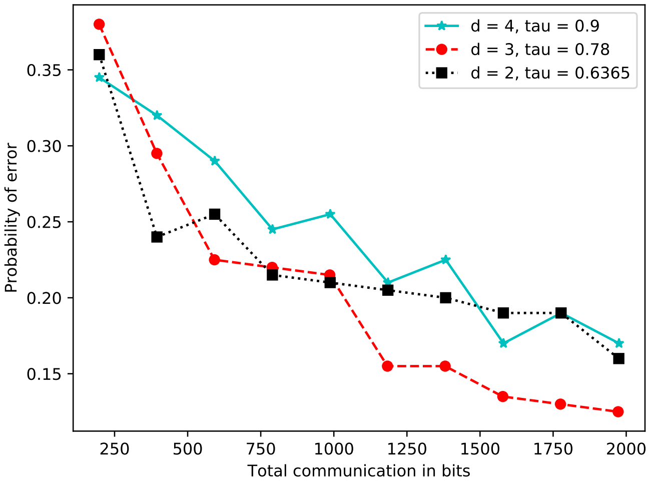

Fig. 1. Performance of the d-dimensional test for different values of d and τ^2 with $d/\tau^2 \approx 4.93$. The probability of error (in y-axis) is the average of probability of false alarm and probability of missed detection, evaluated by averaging over 100 iterations.

Notation. Random variables are denoted by capital letters such as X, Y, *etc.*; their specific realizations by the corresponding small letters such as x, y, *etc.*; and their ranges by the corresponding calligraphic forms such as $\mathcal{X}$, $\mathcal{Y}$, *etc.*. $[N]$ denotes the set of integers $\{1, 2, \ldots, N\}$. For a distribution $\mathbb{P}_\rho$ parametrized by ρ, we use $\mathbb{E}_\rho[X]$ to denote the expectation of the random variable X with respect to $\mathbb{P}_\rho$. Also, $\mathbb{E}_U$ denotes expectation with respect to U. $\mathbb{P}_{\mathcal{H}_i}(A)$ denotes the probability of event A under hypothesis $\mathcal{H}_i$. All the logarithms denoted by $\log$ are to the base 2; when needed, we use $\ln a$ to denote the natural logarithm of a. For a vector a, $a(i)$ denotes its i-th coordinate, a^T denotes its transpose, and $\|a\|_p := \left(\sum_{i=1}^d |a(i)|^p\right)^{\frac{1}{p}}$ denotes its ℓ_p-norm.

II. Problem Setup

We consider jointly Gaussian random variables $X \in \mathbb{R}^d$ and $Y \in \mathbb{R}$ with joint distribution as follows: for $\rho(i) \in [-1, 1], 1 \le i \le d$, we assume that

$$\mathbb{E}[X(i)] = 0, \quad \mathbb{E}[X(i)X(j)] = \mathbb{1}\{i = j\}, 1 \le i \le j \le d,$$

$$\mathbb{E}[Y|X] = \sum_{i=1}^{d} \rho(i)X(i), \quad \mathbb{E}[Y^2] = 1. \tag{1}$$

Note that the assumptions above imply $\mathbb{E}[Y] = 0$. Since we assume $\mathbb{E}\left[Y^2\right] = 1$, Jensen's inequality gives

$$\|\rho\|_2^2 = \mathbb{E}\left[\mathbb{E}[Y|X]^2\right] \le \mathbb{E}\left[Y^2\right] = 1.$$

Alternatively, we can describe the joint distribution of X and Y as follows:

$$Y = \rho^T X + \sqrt{1 - \|\rho\|_2^2} Z,$$

[1] As will be seen below, our proposed test uses a "median trick" to convert the one-dimensional test to a d-dimensional test. In our simulation, even the probabilities of correctness for the one-dimensional test are boosted to the desired levels by repeating the tests and using a similar "median trick".

where Z is a standard normal random variable, and X and Z are independent.

Let $(X_t, Y_t)_{t=1}^n$ denote n independent copies of (X, Y). We consider a distributed hypothesis testing problem where parties $\mathcal{P}_1$ and $\mathcal{P}_2$ observe $X^n = (X_1, ..., X_n)$ and $Y^n = (Y_1, ..., Y_n)$, respectively, and seek to resolve the following composite hypothesis testing problem:

$$\mathcal{H}_0^d : \|\rho\|_2 \geq \tau,$$
$$\mathcal{H}_1^d : \rho = 0,$$

where τ takes values in $(0, 1]$ and is known to both the parties.

To determine the true hypothesis, the parties communicate with each other interactively in multiple rounds. Specifically, the parties use an r-round interactive communication protocol π that comprises mappings $f_1, ..., f_r$; $\mathcal{P}_1$ and $\mathcal{P}_2$ use mappings f_i, $1 \leq i \leq r$, to communicate in odd and even rounds i, respectively. Each mapping f_i takes as input the local observation of the party, the previously seen communication, and a shared random variable V available to both the parties and outputs a binary string. Formally, denoting by C_j the random binary string sent in round j, we have

$$f_i : (X^n, C_1, ..., C_{i-1}, V) \mapsto C_i \in \{0,1\}^{\ell_i}, 1 \leq i \leq r, i \text{ odd},$$
$$f_i : (Y^n, C_1, ..., C_{i-1}, V) \mapsto C_i \in \{0,1\}^{\ell_i}, 1 \leq i \leq r, i \text{ even},$$

where ℓ_i, $1 \leq i \leq r$, denotes the length of communication in round i. The overall random communication $(C_1, ..., C_r)$ is called the transcript of the protocol and is denoted by Π. Furthermore, we denote by $|\pi|$ the length $\sum_{i=1}^r \ell_i$ of the transcript of the protocol. For simplicity, we describe our formulation below only for odd r; the case of even r can be handled similarly.

For an odd r, an *r-interactive distributed test* $T = (\pi, g)$ consists of an r-round interactive communication protocol π and a decision mapping $g(Y^n, \Pi, V) \in \{0, 1\}$. A distributed test $T = (\pi, g)$ constitutes an $(\ell, \delta, \varepsilon, \tau)$-test with observation length n if $|\pi| = \ell$ and

$$\mathbb{P}_{\mathcal{H}_0^d}\left(g\big(Y^n, \Pi, V\big) = 1\right) \leq \delta, \text{ and}$$
$$\mathbb{P}_{\mathcal{H}_1^d}\left(g\big(Y^n, \Pi, V\big) = 0\right) \leq \varepsilon.$$

Note that hypothesis $\mathcal{H}_0$ is composite, and we have used $\mathbb{P}_{\mathcal{H}_0^d}(\cdot)$ as a shorthand for $\max_{\rho : \|\rho\|_2 \geq \tau} \mathbb{P}_\rho(\cdot)$, and further, $\mathbb{P}_{\mathcal{H}_1^d}(\cdot)$ denotes $\mathbb{P}_0(\cdot)$. Also, note that shared randomness V can be used by both the interactive communication protocol π as well as the decision mapping g.

Our goal is to design a distributed test that communicates as few bits as possible, while possessing the desired probabilities of error. Formally, we seek bounds for the minimum communication for d-dimensional correlation testing, defined next.

Definition 1. Given $\delta, \varepsilon \in [0, 1]$ and $\tau \in (0, 1]$, the minimum r-round communication for d-dimensional correlation testing $C_d^r(\delta, \varepsilon, \tau)$ is the least ℓ such that there exists an $(\ell, \delta, \varepsilon, \tau)$-test $T = (\pi, g)$ with an r-round interactive communication protocol π, for all observations of length n sufficiently large.

The minimum communication for d-dimensional correlation testing $C_d(\delta, \varepsilon, \tau)$ is the infimum over $r \in \mathbb{N}$ of $C_d^r(\delta, \varepsilon, \tau)$.

While we have formulated the problem for general r, our main focus in this work is the minimum communication $C_d^1(\delta, \varepsilon, \tau)$ for one-way communication protocols. We characterize the dependence of $C_d^1(\delta, \varepsilon, \tau)$ on ε (respectively δ), up to absolute multiplicative constants and additive constants that may depend on δ (respectively ε). Furthermore, we show that the dependence on ε is optimal up to constant factors, even when additional rounds of interaction are available. We summarize our results formally in the next section. But before that we formulate the related problem of correlation estimation.

Consider the problem of estimating ρ for the joint distribution given in (1). The observation of the parties and the r-round interactive communication protocol is defined as before; as above, we define the problem only for odd r. An *r-interactive distributed estimate* is a pair $(\pi, \widehat{\rho})$ where π is an r-round interactive communication protocol and $\widehat{\rho} : (Y^n, \Pi, V) \mapsto \widehat{\rho}(Y^n, \Pi, V) \in [-1, 1]^d$.

An r-interactive distributed estimate $(\pi, \widehat{\rho})$ constitutes an (ℓ, τ)-estimate if $|\pi| \leq \ell$ and

$$\mathbb{E}_\rho\left\{\|\widehat{\rho}(Y^n, \Pi, V) - \rho\|_2^2\right\} \leq \tau^2, \tag{2}$$

where $\mathbb{E}_\rho$ denotes the expectation with respect to the distribution in (1).

Definition 2. Given $\tau \in (0, 1]$, the minimum r-round communication for d-dimensional correlation estimation $\widetilde{C}_d^r(\tau)$ is the least ℓ such that there exists an (ℓ, τ)-estimate $T = (\pi, \widehat{\rho})$ with an r-round interactive communication protocol π, for all observations of length n sufficiently large.

The minimum communication for d-dimensional correlation estimation $\widetilde{C}_d(\tau)$ is the infimum over $r \in \mathbb{N}$ of $\widetilde{C}_d^r(\tau)$.

In the next section, we will provide a lower bound for $\widetilde{C}_d(\tau)$, which establishes roughly that correlation estimation requires much more communication than correlation testing.

Remark 1 (Shared randomness)**.** In this paper, all lower bounds hold for protocols with shared randomness and all upper bounds are achieved by one-way protocols with shared randomness.

III. Main results

We have divided our results into three parts: upper bounds for $C_d^1(\delta, \varepsilon, \tau)$ achieved by our proposed scheme, a lower bound for $C_d^1(\delta, \varepsilon, \tau)$, and a lower bound for $C_d(\delta, \varepsilon, \tau)$ with $r > 1$. These parts are presented in separate sections below. The upshot of our results is that our protocol with $r = 1$ uses almost minimum communication not only among one-way communication protocols, but also among interactive protocols. Furthermore, we establish a lower bound for the correlation estimation protocol which shows that it requires strictly more communication than correlation testing. We conclude this section with a few comments on our assumptions and techniques.

A. Upper bounds for $C_d^1(\delta, \varepsilon, \tau)$

Our goal in this work is to handle high dimensional correlation testing. Interestingly, we establish a reduction which relates the high dimensional case to the $d = 1$ case. To state our general result, first we state the result for $d = 1$.

Theorem 1. *For every $\delta, \varepsilon \in (0, 1)$,*

$$C_1^1(\delta, \varepsilon, \tau) \leq \frac{1}{\tau^2}\left(\sqrt{\log\frac{1}{\varepsilon}} + \sqrt{(1-\tau^2)\log\frac{1}{\delta}}\right)^2 + \ln\left(\frac{2}{\tau^2}\left(\sqrt{\ln\frac{1}{\varepsilon}} + \sqrt{(1-\tau^2)\ln\frac{1}{\delta}}\right)^2 + 1\right) + O\left(\sqrt{\log\frac{1}{\delta}\log\frac{1}{\varepsilon}}\right).$$

To extend this result to the case of general d, we convert the d-dimensional problem to the one-dimensional problem as follows: Party $\mathcal{P}_1$ applies a random rotation (using shared randomness V) to the observed vector X to obtain $\widetilde{X}$. We show that the first coordinate $\widetilde{X}(1)$ of the resulting vector $\widetilde{X}$ and Y have correlation coefficient roughly $(\tau/\sqrt{d})$ under $\mathcal{H}_0^1$ (with high probability) and correlation coefficient 0 under $\mathcal{H}_1^1$. Using this fact (and a reduction result provided in the next section), we get the following upper bound for $C_d^1(\delta, \varepsilon, \tau)$.

Theorem 2. *There exists a positive constant $c > 0$ such that for every $\delta, \varepsilon \in (0, 1)$ we have*

$$C_d^1(\delta, \varepsilon, \tau) \leq c \cdot \frac{d}{\tau^2} \cdot \max\left\{\log\frac{1}{\delta}, \log\frac{1}{\varepsilon}\right\} + O\left(\ln\frac{d}{\tau^2}\right),$$

where the second term has no dependence on δ or ε.

B. Lower bounds for $C_d^1(\delta, \varepsilon, \tau)$

Our lower bound for the case $d = 1$ matches the upper bound of Theorem 1 up to additive terms of lower order to yield the following characterization for $C_1^1(\delta, \varepsilon, \tau)$.

Theorem 3. *For a fixed $\delta \in (0, 1/2)$ and every ε such that $\delta + \varepsilon^{\frac{1-\tau}{1+\tau}} \leq 1$, we have*[2]

$$C_1^1(\delta, \varepsilon, \tau) = \frac{1}{\tau^2}\log\frac{1}{\varepsilon} + O_\delta\left(\sqrt{\log\frac{1}{\varepsilon}}\right),$$

and for a fixed $\varepsilon \in (0, 1/2)$ and every $\delta \in (0, 1/2)$, we have

$$C_1^1(\delta, \varepsilon, \tau) = \frac{1-\tau^2}{\tau^2}\log\frac{1}{\delta} + O_\varepsilon\left(\sqrt{\log\frac{1}{\delta}}\right),$$

where the notation O_x denotes that the constant implied by O depends on x.

The proof of this result uses the notions of *hypercontractivity* and *reverse hypercontractivity* and is given in Section V.

In fact, we can relate the d-dimensional problem to the one-dimensional problem by revealing extra information to $\mathcal{P}_2$ to obtain a matching lower bound for Theorem 2, from which the next result follows.

Theorem 4. *For $0 < \tau \leq 1$, $\delta \in (0, 1/2)$, and ε such that $\delta + \varepsilon^{\frac{1-\tau}{1+\tau}} \leq 1$, we have*

$$C_d^1(\delta, \varepsilon, \tau) = \Theta\left(\frac{d}{\tau^2} \cdot \max\left\{\log\frac{1}{\varepsilon}, \log\frac{1}{\delta}\right\}\right).$$

We remark that the reduction of the general $d \geq 1$ case to the one-dimensional case used in the proof of lower bound differs from the reduction in the upper bound; we provide the proof in Section V. Nevertheless, it is interesting that we obtain tight results by relating the high dimensional setting to the one-dimensional setting.

C. Lower bounds for $r \geq 1$

Our final set of results provide lower bounds for interactive communication with shared randomness, establishing the optimality of our proposed distributed test even among interactive tests. To derive this lower bound, we use a data processing inequality from [24], which was used in a similar context in [19]. In fact, using this technique we can even derive a lower bound for the high dimensional correlation estimation problem, showing that this problem requires orderwise higher communication in comparison to correlation testing.

We begin with the result for the correlation testing problem. Note that we only prove optimality in the dependence on ε, and not on δ.

Theorem 5. *For $\delta, \varepsilon, \tau \in (0, 1)$, we have*

$$C_d(\delta, \varepsilon, \tau) \geq \frac{d}{\tau^2}\left((1-\delta)\log\frac{1}{\varepsilon} - 1\right).$$

The proof is provided in Section VI-A.

We note that while the lower bound above extends the bounds from the previous section to the interactive setting, it does not yield optimal constants for $d = 1$ and $r = 1$ unlike Theorem 3. In fact, we believe that even the lower bound in Theorem 15 yields a tight constant; the slackness in characterization of $C_d^1(\delta, \varepsilon, \tau)$ arises from our upper bound. Thus, the lower bound in Theorem 5 is weaker than those given in the previous section for $r = 1$.

Recall that the lower bounds of the previous section are derived using the concepts of hypercontractivity and reverse hypercontractivity (which appeared in the preliminary version of this paper [29]). As mentioned above, the lower bound in Theorem 5 uses a related but different idea of strong data processing inequalities derived in [19]. Using the same bound and Fano's inequality, we also obtain the following lower bound for $\widetilde{C}_d(\tau)$.

Theorem 6. *There exists a constant $c > 0$, such that for every $\tau \in (0, 1)$,*

$$\widetilde{C}_d(\tau) \geq \frac{c\, d^2}{\tau^2}.$$

The proof is provided in Section VI-B.

In fact, the lower bound above is tight too, and matches the upper bound attained by the distributed estimate proposed in [21]. The lower bound above establishes that a simple estimate-and-test approach using the estimate in [21] or other

[2] With an abuse of notation, the $O(x)$ notation for the additive error denotes that the upper and lower bounds differ by at most an $O(x)$ term.

estimates will not be able to attain the optimal $O(d/\tau^2)$ communication needed for correlation testing.

D. Comments on assumptions and techniques

Our proposed one-way communication scheme is related to the scheme in [18] where communication for common randomness generation ($cf.$ [7]) was considered. We draw on the heuristic connection between independence testing and common randomness generation highlighted in [35], [34] to adapt the scheme of [18] to devise a distributed correlation test. In particular, the scheme in [18] uses a correlated sampling idea to find a common vector U that is close to both X^n and Y^n. The index of this common vector constitutes the generated common randomness; the communication used is of lower length than the common randomness since X_t and Y_t, $1 \leq t \leq n$, are correlated. In contrast, for our problem, it is not *a priori* known whether X_t and Y_t are correlated or not. Nonetheless, we use a similar correlated sampling scheme as in [18], with an important difference – we may not find a common vector any more. In fact, we declare independence ($\mathcal{H}_1$) if no common vector is found, and "correlatedness" ($\mathcal{H}_0$) otherwise. Note that since the random variables need not be correlated, the worst-case communication length is that for the independence case.

We remark that for $d = 1$, the problem is, in effect, to decide whether the given vectors X^n and Y^n are orthogonal or have an inner product greater than a known quantity. For this case, it is possible to use a simple scheme that quantizes each value X_t to its sign $\mathbb{1}\{X_t \geq 0\}$ and uses the known sample complexity results for independence testing for the collocated case ($cf.$ [5]). This, too, will result in a scheme that requires $O(1/\tau^2)$ bits of communication. However, we noted in [29], where we study the communication complexity of one-dimensional independence testing, that our proposed scheme uses communication that is a constant factor lower than this baseline scheme. In particular, for X_t and Y_t with correlation $\rho > 0$, the correlation between the bits $\mathbb{1}\{X_t \geq 0\}$ and $\mathbb{1}\{Y_t \geq 0\}$ is $\frac{2}{\pi}\arcsin(\rho) \leq \rho$. On the other hand, our scheme requires a much larger n than this baseline scheme for $d = 1$.

Finally, note that we have chosen the distribution to be Gaussian just for convenience. Since we allow the number of samples to be arbitrarily large, even when X_t and Y_t are not Gaussian, we can replace subset of samples with their sample means and use the central limit theorem (Berry-Esseen approximation) to do similar calculations as those presented in this paper. Now we present our scheme and its analysis.

IV. Our scheme and its analysis

Our general scheme is obtained by first relating the d-dimensional correlation testing problem to the one-dimensional correlation testing problem, and then relating the one-dimensional problem to its one-sided version. We develop a test for this one-sided problem first and then, in steps, convert it to a test for the d-dimensional problem in separate subsections below.

A. One-sided correlation test

Consider the following one-sided variant of the correlation testing problem with $d = 1$:

$$\mathcal{H}_0^+ : \rho \geq \tau,$$
$$\mathcal{H}_1^1 : \rho = 0,$$

where $\tau \in (0, 1]$ is known to both parties. We present a 1-interactive distributed test for this problem; namely, we present a test using one-way communication from $\mathcal{P}_1$ to $\mathcal{P}_2$.

Specifically, fix parameters $r > 0$, $\theta \leq \tau$, and $k \in \mathbb{N}$. Throughout this section, for brevity, with a slight abuse of notation we denote by $X = (X_1, ..., X_n) \in \mathbb{R}^n$ and $Y = (Y_1, ..., Y_n) \in \mathbb{R}^n$, respectively, the observation of $\mathcal{P}_1$ and $\mathcal{P}_2$, where $(X_t, Y_t)_{t=1}^n$ are generated i.i.d. from the distribution in (1). Furthermore, for two vectors u and v in $\mathbb{R}^n$, we denote $u \cdot v := u^T v$.

1) Using the shared randomness, parties generate an $n \times 2^k$ matrix U consisting of i.i.d. uniform $\{-1, +1\}$-valued entries U_{ij}, $1 \leq i \leq n$, $1 \leq j \leq 2^k$.
2) Let U_j denote the j-th column of U. $\mathcal{P}_1$ finds the *least* index $j \in [2^k]$ such that
$$U_j \cdot X \geq r\sqrt{n},$$
and sends the k-bit representation of j to $\mathcal{P}_2$. If no such j is found, declare $\mathcal{H}_1^1$.
3) $\mathcal{P}_2$, upon receiving j, declares $\mathcal{H}_0^+$ if
$$U_j \cdot Y \geq \theta \cdot r\sqrt{n}.$$

The next result captures the performance of our proposed distributed test for an optimized choice of parameters r, θ, and k.

Theorem 7. *For $\delta, \varepsilon \in (0, 1)$, $\tau \in (0, 1)$, an appropriate choice of $\theta \leq \tau$, and for all n sufficiently large, the 1-interactive test proposed above satisfies*

$$\mathbb{P}_{\mathcal{H}_0^+}\left[\textit{Declare } \mathcal{H}_1^1\right] \leq \delta \ \textit{ and } \ \mathbb{P}_{\mathcal{H}_1^1}\left[\textit{Declare } \mathcal{H}_0^+\right] \leq \varepsilon, \quad (3)$$

when r is set as follows:

$$r^2 = \frac{2\ln 2}{\tau^2}\left(\sqrt{\log\frac{1}{\varepsilon} + \log\ln\frac{3}{\delta} + 1} + \sqrt{(1-\tau^2)\log\frac{3}{\delta}}\right)^2,$$

and the communication length k satisfies

$$k = \left\lceil \log\frac{1}{Q(r)} + \log\ln\frac{3}{\delta}\right\rceil,$$

where $Q(\cdot)$ denotes the complementary cumulative distribution function of a standard Gaussian random variable.

Remark 2 (Proof outline)**.** At a high-level, the distributed test proposed above first generates a "random codebook" by generating independent copies of U and then uses a correlated sampling scheme to find a codeword that is "close" to both X and Y (in Euclidean distance). If no such vector is found in the codebook, our scheme declares independence. An error of type-I may occur if no such vector is found in spite of X and Y being correlated. In the first part of our proof, we derive a lower bound for the probability of finding such a common vector when X and Y are correlated, which is the same as an upper bound for the error of type-I. In the other direction, an error of type-II may occur if we find such a common vector for independent X and Y; our second step is to derive an upper bound for the probability of this event.

In both these derivations, we define "closeness" in terms of appropriately chosen thresholds for distance, which are different for X and Y. Specifically, in our scheme, $\mathcal{P}_1$ finds the index of a codeword that is within a fixed Euclidean distance of X, shares this index with $\mathcal{P}_2$, which then tests if the same codeword is within (another) fixed Euclidean distance of Y. Thus, our analysis outlined above, in effect, evaluates the probabilities of random high-dimensional vectors being close in Euclidean distance, a well-studied calculation. Our final step is to carefully choose values for these thresholds that give our stated performance bound.

Proof. We present the proof in steps.

1) *Lower bound for the probability of correctness under $\mathcal{H}_0^+$*: We have

$$\begin{aligned}&\mathbb{P}_{\mathcal{H}_0^+}\left[\text{Declare } \mathcal{H}_0^+\right]\\ &= \sum_{j=1}^{2^k} \mathbb{P}_{\mathcal{H}_0^+}\Big(U_l \cdot X < r\sqrt{n} \text{ for all } l \le j-1,\\ &\qquad U_j \cdot X \ge r\sqrt{n}, U_j \cdot Y \ge \theta \cdot r\sqrt{n}\Big),\end{aligned}$$

where U_0 is set to be 0. We approximate the right-side using the Berry-Esseen theorem ($cf.$ [16]) for a fixed realization $X = x$. The event in the summands above, conditioned on $X = x$, is an intersection of two events, whose probabilities we analyze below:

(i) *(The rejection sampling event $\{U_l \cdot X < r\sqrt{n}$ for all $l \le j-1, U_j \cdot X \ge r\sqrt{n}\}$)* Noting that $U_j \cdot x = \sum_{i=1}^n U_{ij}x_i$ is a sum of independent random variables, the Berry-Esseen theorem yields

$$\begin{aligned}&\mathbb{P}_{\mathcal{H}_0^+}\big(U_l \cdot X < r\sqrt{n} \text{ for all } l \le j-1,\\ &\qquad U_j \cdot X \ge r\sqrt{n}\big| X = x\big)\\ &= \left[\mathbb{P}_{\mathcal{H}_0^+}\left(\sum_{i=1}^n U_{i1}x_i < r\sqrt{n}\middle| X = x\right)\right]^{j-1}\\ &\qquad \mathbb{P}_{\mathcal{H}_0^+}\left(\sum_{i=1}^n U_{ij}x_i \ge r\sqrt{n}\middle| X = x\right)\\ &\ge \left(1 - Q\left(\frac{r\sqrt{n}}{\sqrt{\sum_{i=1}^n x_i^2}}\right) - c_0\frac{\sum_{i=1}^n |x_i|^3}{(\sum_{i=1}^n x_i^2)^{\frac{3}{2}}}\right)^{j-1}\\ &\qquad \left(Q\left(\frac{r\sqrt{n}}{\sqrt{\sum_{i=1}^n x_i^2}}\right) - c_0\frac{\sum_{i=1}^n |x_i|^3}{(\sum_{i=1}^n x_i^2)^{\frac{3}{2}}}\right),\end{aligned}$$

where c_0 is a constant.

(ii) *(The correlated sampling event $\{U_j \cdot Y \ge \theta \cdot r\sqrt{n}\}$)* We analyze the probability of this second event conditioned on the first. Specifically, note that under $\mathcal{H}_0^+$, for each $i \in [n]$ we have $\mathbb{E}[Y_i|X_i] = \rho X_i$ with $\rho \ge \tau$. It follows that for a fixed realization $X = x$ and $U = u$, the random variables $U_{ij}Y_i$, $1 \le i \le n$, are independent with distribution $\mathcal{N}(\rho x_i, 1-\rho^2)$ for every $j \in [2^k]$. Note that for $u_j \cdot x \ge r\sqrt{n}$, we have

$$\mathbb{E}\left[U_j \cdot Y \mid U = u, X = x\right] = \rho(u_j \cdot x) \ge \rho r\sqrt{n}.$$

Therefore, for every u and x such that $u_j \cdot x \ge r\sqrt{n}$ and $u_l \cdot x < r\sqrt{n}$ for all $l \le j-1$, we have

$$\begin{aligned}&\mathbb{P}_{\mathcal{H}_0^+}\left(U_j \cdot Y \ge \theta r\sqrt{n}\middle| U = u, X = x\right)\\ &\ge Q\left(\frac{r\left(\theta - \rho\right)}{\sqrt{1-\rho^2}}\right) \ge Q\left(\frac{r\left(\theta - \tau\right)}{\sqrt{1-\tau^2}}\right),\end{aligned}$$

where the final bound holds since $Q(a)$ is decreasing in a and the function $f(a) = (\theta - a)/\sqrt{1-a^2}$ is non increasing in a for $a \ge \theta$; specifically, this bound uses our assumption that $\theta \le \tau$.

Upon combining the bounds above, denoting $\sigma_n(X) = \sqrt{\sum_{i=1}^n X_i^2}$ and $\beta_n(X) = c_0\sum_{i=1}^n |X_i|^3/\sigma_n^3(X)$, we obtain

$$\begin{aligned}&\mathbb{P}_{\mathcal{H}_0^+}\left[\text{Declare } \mathcal{H}_0^+\right]\\ &\ge Q\left(\frac{r\left(\theta - \tau\right)}{\sqrt{1-\tau^2}}\right)\mathbb{E}\left[\frac{Q\left(\frac{r\sqrt{n}}{\sigma_n(X)}\right) - \beta_n(X)}{Q\left(\frac{r\sqrt{n}}{\sigma_n(X)}\right) + \beta_n(X)}\times\right.\\ &\left.\left(1 - \left(1 - Q\left(\frac{r\sqrt{n}}{\sigma_n(X)}\right) - \beta_n(X)\right)^{2^k}\right)\right].\end{aligned}$$

Using the law of large numbers and the inequality $1 - a \le e^{-a}$, for every $\eta > 0$ and all n sufficiently large, we get

$$\begin{aligned}&\mathbb{P}_{\mathcal{H}_0^+}\left[\text{Declare } \mathcal{H}_0^+\right]\\ &\ge (1-\eta)\left(1 - e^{-2^k Q(r)}\right)Q\left(\frac{r(\theta - \tau)}{\sqrt{1-\tau^2}}\right)\\ &\ge 1 - e^{-2^k Q(r)} - Q\left(\frac{r(\tau - \theta)}{\sqrt{1-\tau^2}}\right) - \eta, \qquad (4)\end{aligned}$$

where we used the bound $(1-x)(1-y)(1-z) \ge 1 - (x+y+z)$ for $x, y, z \in (0,1)$.

2) *Upper bound for the probability of error under $\mathcal{H}_1^1$*: We derive a bound for the probability of declaring $\mathcal{H}_0^+$ when $\mathcal{H}_1^1$ is true, which holds for every fixed realization u of the random codebook U. Since $\sum_{i=1}^n u_{ij}X_i$ is a sum of n independent standard Gaussian random variables, we have

$$\mathbb{P}_{\mathcal{H}_1^1}\left(\sum_{i=1}^n u_{ij}X_i \ge r\sqrt{n}\middle| U = u\right) = Q(r),$$

and similarly,

$$\mathbb{P}_{\mathcal{H}_1^1}\left(\sum_{i=1}^n u_{ij}Y_i \ge \theta r\sqrt{n}\middle| U = u\right) = Q(\theta r).$$

Therefore,

$$\mathbb{P}_{\mathcal{H}_1^1}\left(\text{Declare } \mathcal{H}_0^+\right) \le \mathbb{E}_U\left[\sum_{j=1}^{2^k}\mathbb{P}_{\mathcal{H}_1^1}\left(\sum_{i=1}^n U_{ij}X_i \ge r\sqrt{n}\right)\right.$$

$$\mathbb{P}_{\mathcal{H}_1^1}\left(\sum_{i=1}^{n} U_{ij}Y_i \geq \theta r\sqrt{n}\right)\Bigg]$$
$$\leq 2^k\, Q(r)\, Q(\theta r). \tag{5}$$

3) *Choice of the optimal parameters satisfying the error requirements:* To satisfy the error condition (3), by (4) and (5) it suffices to set $\eta = \delta/3$ and choose r, θ, and k to satisfy the following:

$$\ln\frac{3}{\delta} \leq 2^k Q(r) \leq 2\ln\frac{3}{\delta}, \tag{6}$$
$$Q\left(\frac{r(\tau-\theta)}{\sqrt{1-\tau^2}}\right) \leq \frac{\delta}{3}, \tag{7}$$
$$Q(\theta r) \leq \frac{\varepsilon}{2\ln\frac{3}{\delta}}. \tag{8}$$

Using Chernoff bound $Q(x) \leq e^{-x^2/2}$, for conditions (7) and (8) it suffices to have

$$\frac{1-\tau^2}{(\tau-\theta)^2}\cdot\log\frac{3}{\delta} \leq \frac{r^2}{2\ln 2},$$
$$\frac{1}{\theta^2}\left(\log\frac{1}{\varepsilon} + \log\ln\frac{3}{\delta} + 1\right) \leq \frac{r^2}{2\ln 2}.$$

Therefore, the least value of k is given by an r that satisfies

$$\frac{r^2}{2\ln 2} = \min_{\theta\leq\tau}\max\left\{\frac{a}{\theta^2}, \frac{b}{(\tau-\theta)^2}\right\},$$

where $a = \left(\log\frac{1}{\varepsilon} + \log\ln\frac{3}{\delta} + 1\right)$ and $b = (1-\tau^2)\log\frac{3}{\delta}$. The optimal θ^* for the problem on the right-side is given by

$$\theta^* = \frac{\tau\sqrt{a}}{\sqrt{b}+\sqrt{a}},$$

whereby our optimal choice of r^2 is

$$\frac{r^2}{2\ln 2} = \frac{1}{\tau^2}\left(\sqrt{a}+\sqrt{b}\right)^2$$
$$= \frac{1}{\tau^2}\left(\sqrt{\log\frac{1}{\varepsilon} + \log\ln\frac{3}{\delta} + 1} + \sqrt{(1-\tau^2)\log\frac{3}{\delta}}\right)^2.$$

Thus, by (6), we can satisfy (3) if we set[2] $k = \left\lceil \log\frac{1}{Q(r)} + \log\ln\frac{3}{\delta}\right\rceil$ for r given above.

□

B. Distributed correlation test for $d = 1$

We now extend the one-sided test above to a test for $d = 1$. We present a general reduction which will allow us to use any 1-interactive distributed test for the one-sided problem (not just the one above) for the (two-sided) correlation testing problem with $d = 1$.

Lemma 8 (Two-sided to one-sided)**.** *For $\delta \in (0,1)$, $\varepsilon \in (0,1/2)$, $\tau \in (0,1)$, and $\ell \in \mathbb{N}$, suppose that $T^+ = T^+(X^n, Y^n)$ is an 1-interactive $(\ell,\delta,\varepsilon,\tau)$-test for the one-sided correlation testing problem. Then, we can find a 1-interactive $(\ell,\delta,2\varepsilon,\tau)$-test for the correlation testing problem with $d = 1$.*

Proof. We begin by noting that $T^-(X^n, Y^n) = T^+(X^n, -Y^n)$ is an $(\ell,\delta,\varepsilon,\tau)$-test for the following alternative one-sided problem:

$$\mathcal{H}_0^- : \rho \leq -\tau,$$
$$\mathcal{H}_1^1 : \rho = 0.$$

Note that the communication protocol for T^+ and T^- is the same; the corresponding decision mappings g^+ and g^- differ. In particular, $g^-(Y^n,\Pi,V) = g^+(-Y^n,\Pi,V)$, and let π be the common communication protocol for T^+ and T^-. Consider the following 1-interactive distributed test $T = (\pi, g)$ for the correlation testing problem.

1) Parties execute the communication protocol π.
2) Use decision mapping

$$g(Y^n,\Pi,V) = \min\{g^+(Y^n,\Pi,V), g^-(Y^n,\Pi,V)\}.$$

For this test, we can verify that

$$\begin{aligned}
&\mathbb{P}_{\mathcal{H}_1^1}\left[g(Y^n,\Pi,V) = 0\right]\\
&= \mathbb{P}_{\mathcal{H}_1^1}\left[g^+(Y^n,\Pi,V) = 0 \text{ or } g^-(Y^n,\Pi,V) = 0\right]\\
&\leq \mathbb{P}_{\mathcal{H}_1^1}\left[g^+(Y^n,\Pi,V) = 0\right] + \mathbb{P}_{\mathcal{H}_1^1}\left[g^-(Y^n,\Pi,V) = 0\right]\\
&\leq 2\varepsilon.
\end{aligned}$$

Furthermore, under $\mathcal{H}_0^1$,

$$\begin{aligned}
&\mathbb{P}_{\mathcal{H}_0^1}\left[g(Y^n,\Pi,V) = 1\right]\\
&= \mathbb{P}_{\mathcal{H}_0^1}\left[g^+(Y^n,\Pi,V) = g^-(Y^n,\Pi,V) = 1\right]\\
&\leq \max\Big\{\mathbb{P}_{\mathcal{H}_0^+}\left[g^+(Y^n,\Pi,V) = g^-(Y^n,\Pi,V) = 1\right],\\
&\qquad \mathbb{P}_{\mathcal{H}_0^-}\left[g^+(Y^n,\Pi,V) = g^-(Y^n,\Pi,V) = 1\right]\Big\}\\
&\leq \delta,
\end{aligned}$$

which shows that T constitutes an $(\ell,\delta,2\varepsilon,\tau)$-test. □

Lemma 8, Theorem 7, and the well-known bound $Q(x) \geq \frac{x}{\sqrt{2\pi}(x^2+1)}e^{-x^2/2}$ yield Theorem 1.

C. Proof of Theorem 2

Finally, now that we have a correlation test for $d = 1$, we complete the proof of Theorem 2 to obtain a test for general d. We begin by making a simple observation akin to the "median trick" in randomized algorithms.

Lemma 9. *For $\alpha, \beta, \tau \in (0,1)$ with $\alpha + \beta < 1$, suppose that we have an r-interactive (ℓ,α,β,τ)-test for the d-dimensional correlation testing problem. Then, for every $\delta, \varepsilon \in (0,1)$, we can obtain an r-interactive $(m\ell,\delta,\varepsilon,\tau)$-test for the d-dimensional correlation testing problem whenever*

$$m \geq \frac{2}{(1-\beta-\alpha)^2}\max\left\{\ln\frac{1}{\delta}, \ln\frac{1}{\varepsilon}\right\}.$$

Proof. We provide proof only for odd r; even r can be handled similarly. Consider an r-interactive distributed test $T = (\pi, g)$ that satisfies

$$\mathbb{P}_{\mathcal{H}_0^d}\left(g(Y^n,\Pi,V) = 1\right) \leq \alpha,$$

[2] In our analysis, we cannot set k higher than this either.

$$\mathbb{P}_{\mathcal{H}_1^d}\left(g(Y^n,\Pi,V)=0\right)\le\beta,$$

where π is a communication protocol of length ℓ. To construct the desired test, we repeat the test above m times independently. Specifically, we first apply the test above to m independent copies of (X^n, Y^n, V) to obtain transcripts $\Pi_1, ..., \Pi_m$. Note that the resulting communication protocol is still an r-round protocol, with length $m\ell$. Denote by $V_1, ..., V_m$ the independent copies of the shared randomness used for the protocol. Further, denote by D_i the output $g(Y_{n(i-1)+1}^{ni}, \Pi_i, V_i)$, $1 \le i \le m$, for the i-th copy of the test. Consider the new decision mapping g^m given by

$$g^m(Y^{nm},\Pi^m,V^m)=\mathbb{1}\left\{\sum_{i=1}^{m}D_i>mt\right\},$$

for a fixed $\alpha < t < 1-\beta$. Note that $D_1, ..., D_m$ are independent bits and by our assumption about T, satisfy

$$\begin{aligned}&\mathbb{P}_{\mathcal{H}_0^d}\left(D_i=1\right)\le\alpha,\\&\mathbb{P}_{\mathcal{H}_1^d}\left(D_i=1\right)\ge1-\beta,\end{aligned}$$

for every $1 \le i \le m$. Therefore, by Hoeffding's inequality,

$$\begin{aligned}\mathbb{P}_{\mathcal{H}_0^d}\left(g^m(Y^{nm},\Pi^m,V^m)=1\right)&=\mathbb{P}_{\mathcal{H}_0^d}\left(\sum_{i=1}^{m}D_i>mt\right)\\&\le e^{-2m(t-\alpha)^2},\end{aligned}$$

and similarly,

$$\begin{aligned}\mathbb{P}_{\mathcal{H}_1^d}\left(g^m(Y^{nm},\Pi^m,V^m)=0\right)&=\mathbb{P}_{\mathcal{H}_1^d}\left(\sum_{i=1}^{m}D_i\le mt\right)\\&\le e^{-2m(1-\beta-t)^2}.\end{aligned}$$

In particular, by setting $m \ge \frac{2}{(1-\beta-\alpha)^2}\max\left\{\ln\frac{1}{\delta},\ln\frac{1}{\varepsilon}\right\}$ and $t=(1-\beta+\alpha)/2$, we obtain the desired test. □

Thus, it suffices to construct a distributed test with constant probability of error. We do that in the result below by using a 1-interactive distributed test for $d=1$. Our test uses a randomized construction; to facilitate its analysis, we note the following fact.

Lemma 10. *For $R=\frac{1}{\sqrt{d}}W$ with W a random vector consisting of i.i.d. Rademacher entries, for every vector $x\in\mathbb{R}^d$ we have,*

$$\mathbb{P}\left(\left(R^Tx\right)^2\ge\frac{\|x\|_2^2}{2d}\right)\ge\frac{1}{28}.$$

Proof. The proof uses the Paley-Zygmund inequality. Specifically, denote by Z the random variable R^Tx. Then,

$$\begin{aligned}\mathbb{E}\left[Z^2\right]&=\mathbb{E}\left[\left(\sum_{j=1}^{d}R_jx_j\right)^2\right]\\&=\mathbb{E}\left[\frac{1}{d}\|x\|_2^2+\sum_{i=1}^{d}\sum_{j=1}^{d}R_iR_jx_ix_j\mathbb{1}\{j\neq i\}\right]\\&=\frac{1}{d}\left\|x\right\|_2^2,\end{aligned}$$

where the last step follows from the fact that entries of R are independent with zero-mean. Next, we consider $\mathbb{E}\left[Z^4\right]$. Note that the only terms in the expansion of $\left(\sum_{i=1}^d R_ix_i\right)^4$ that have nonzero mean are those which have only even powers of entries of R. In particular, these are terms of the form $R_i^4x_i^4$ and $R_i^2R_j^2X_i^2X_j^2$ with distinct i, j. Therefore, we have

$$\begin{aligned}\mathbb{E}\left[Z^4\right]&=\frac{1}{d^2}\sum_{i=1}^{d}x_i^4+\binom{4}{2}\frac{1}{d^2}\sum_{i=1}^{d}\sum_{j=1}^{d}x_i^2x_j^2\mathbb{1}\{i\neq j\}\\&\le\frac{1}{d^2}\left(\|x\|_4^4+6\|x\|_2^4\right)\\&\le\frac{7\|x\|_2^4}{d^2},\end{aligned}$$

where the final inequality uses $\|x\|_4\le\|x\|_2$. Therefore, by the Paley-Zygmund inequality, for $\nu\in(0,1)$,

$$\mathbb{P}\left(Z^2>\nu\mathbb{E}\left[Z^2\right]\right)\ge(1-\nu)^2\frac{\mathbb{E}\left[Z^2\right]^2}{\mathbb{E}\left[Z^4\right]}\ge\frac{(1-\nu)^2}{7}.$$

The claim follows by setting $\nu=1/2$. □

We are now in a position to complete the proof of Theorem 2. We use the distributed test for $d=1$ from Theorem 1 to build a test for a general d. Specifically, we replace the d-dimensional observations $X_1, ..., X_n$ of $\mathcal{P}_1$ with one-dimensional $\widetilde{X}_1, ..., \widetilde{X}_n$ given by $\widetilde{X}_t = R^TX_t$, $1\le t\le n$, where R is a random vector generated as in Lemma 10. Note that $(\widetilde{X}_t, Y_t)_{t=1}^n$ are i.i.d. with

$$\begin{aligned}\mathbb{E}\left[Y_1\widetilde{X}_1\mid R\right]&=\mathbb{E}\left[\left(\rho^TX_1+\sqrt{1-\|\rho\|_2^2}Z_1\right)\left(R^TX_1\right)\mid R\right]\\&=\rho^T\mathbb{E}\left[X_1X_1^T\right]R\\&=\rho^TR.\end{aligned}$$

Thus, by Lemma 10,

$$\mathrm{P}_R\left(\left\{r:\left|\mathbb{E}\left[Y_1\widetilde{X}_1|R=r\right]\right|\ge\frac{\|\rho\|_2}{\sqrt{2d}}\right\}\right)\ge\frac{1}{28}.$$

Denoting $\mathcal{G}:=\left\{r:\left|\mathbb{E}\left[Y_1\widetilde{X}_1|R=r\right]\right|\ge\frac{\|\rho\|_2}{\sqrt{2d}}\right\}$ and $\widetilde{\rho}(r):=\left|\mathbb{E}\left[Y_1\widetilde{X}_1\mid R=r\right]\right|$, for every $r\in\mathcal{G}$ we have

$$\begin{aligned}&\widetilde{\rho}(r)\ge\tau/\sqrt{2d}\text{ under }\mathcal{H}_0^d,\\&\widetilde{\rho}(r)=0\text{ under }\mathcal{H}_1^d.\end{aligned}$$

Also, in the test we construct for the d-dimensional case, we invoke a 1-interactive $(\ell, 1/56, 1/112, \tau/\sqrt{2d})$-test T_1 for the one-dimensional correlation testing problem $\widetilde{\rho}(r)\ge\tau/\sqrt{2d}$ versus $\widetilde{\rho}(r)=0$ with

$$\ell\le\frac{cd}{\tau^2},$$

for an appropriate constant c, as guaranteed by Theorem 1.

Next, consider the test for $\mathcal{H}_0^d$ versus $\mathcal{H}_1^d$ that samples R from shared randomness executes the aforementioned test T_1 for $\mathcal{H}_0^d$ versus $\mathcal{H}_1^d$ the one-dimensional problem $\widetilde{\rho}(R)\ge\tau/\sqrt{2d}$ versus $\widetilde{\rho}(R)=0$. We make the observation that $\widetilde{\rho}(R)=0$ *almost surely* for R, when $\rho=0$. Thus, the missed detection probability for the one-dimensional test remains unchanged. However, a false alarm may be raised when $R \notin \mathcal{G}$ or when the one-dimensional test raises a false alarm. It follows that for this test

$$\mathbb{P}_{\mathcal{H}_0^d}\left(\text{Declare } \mathcal{H}_1^d\right) \leq \frac{1}{56} + \mathbb{P}\left(R \notin \mathcal{G}\right) \leq \frac{55}{56},$$

and

$$\mathbb{P}_{\mathcal{H}_1^d}\left(\text{Declare } \mathcal{H}_0^d\right) \leq \frac{1}{112},$$

whereby it constitutes a $(cd/\tau^2, 55/56, 1/112, \tau)$-test for the d-dimensional correlation testing problem.

Thus, we have obtained our desired test with constant probability of error guarantees. Theorem 2 follows by using this test along with Lemma 9.

V. Proof of lower bounds for $r = 1$

We begin by deriving lower bounds for the one-dimensional problem. Our lower bounds involve the notions of hypercontractivity and reverse hypercontractivity (*cf.* [8], [26], [13], [25]), which we define first.

For $1 \leq q \leq p < \infty$, a pair of random variables (X, Y) is (p, q)-hypercontractive if for all $\mathbb{R}$-valued functions f of X and g of Y,

$$\mathbb{E}\left[|f(X)g(Y)|\right] \leq \|f(X)\|_{p'}\|g(Y)\|_q,$$

where $p' = p/(p-1)$ is the Hölder conjugate of p. Similarly, for $1 \geq q > p$, a pair of random variables (X, Y) is (p, q)-reverse hypercontractive if for all $\mathbb{R}$-valued functions f of X and g of Y,

$$\mathbb{E}\left[|f(X)g(Y)|\right] \geq \|f(X)\|_{p'}\|g(Y)\|_q.$$

The set of all (p, q) for which (X, Y) is (p, q)-hypercontractive and (p, q)-reverse hypercontractive, respectively, are called the hypercontractivity ribbon and the reverse hypercontractivity ribbon of (X, Y). In particular, we use the following characterization obtained by setting f and g to be indicator functions. For $\mathcal{A} \times \mathcal{B}$, a measurable subset of $\mathbb{R}^n \times \mathbb{R}^n$, for $1 \leq q \leq p < \infty$,

$$\mathbb{P}[\mathcal{A} \times \mathcal{B}] \leq \mathbb{P}[\mathcal{A}]^{1/p'}\mathbb{P}[\mathcal{B}]^{1/q},$$

and for $1 \geq q > p$,

$$\mathbb{P}[\mathcal{A} \times \mathcal{B}] \geq \mathbb{P}[\mathcal{A}]^{1/p'}\mathbb{P}[\mathcal{B}]^{1/q}.$$

The following *tensorization* property of hypercontractivity and reverse hypercontractivity ribbons are well known.

Lemma 11 (Tensorization [26] [25])**.** *For $p \geq 1$, define*

$$q_p(X, Y) = \inf\{q : (X, Y) \textit{ is } (p, q)\textit{-hypercontractive}\},$$

and $r_p(X, Y) = q_p(X, Y)/p$. If $(X_i, Y_i)_{i=1}^n$ are independent, then

$$r_p(X^n, Y^n) = \max_{1 \leq i \leq n} r_p(X_i, Y_i).$$

Furthermore, for $p \leq 1$, define

$$q_p(X, Y) = \sup\{q : (X, Y) \textit{ is } (p, q)\textit{-reverse hypercontractive}\},$$

and $s_p(X, Y) = q_p(X, Y)/p$. If $(X_i, Y_i)_{i=1}^n$ are independent, then

$$s_p(X^n, Y^n) = \max_{1 \leq i \leq n} s_p(X_i, Y_i).$$

We use the notions of hypercontractivity and reverse hypercontractivity to obtain the change of measure bounds between the joint distribution and the independent distribution, which in turn lead to the following two lower bounds for $C_1^1(\delta, \varepsilon, \tau)$. Specifically, we note that the acceptance region corresponding to one-way communication corresponds to a union of disjoint rectangle sets. We use hypercontractivity to relate the measures of rectangle sets under the joint distribution corresponding to $|\rho| > \tau$ and the product distribution corresponding to $\rho = 0$, which in turn leads to the required lower bound.

Theorem 12 (Lower bound 1)**.** *Given $\delta, \varepsilon \in (0, 1)$ and (p, q) such that $1 \leq p' \leq q \leq p$ and (X, Y) is (p, q)-hypercontractive, the minimum one-way communication for one-dimensional correlation testing $C_1^1(\delta, \varepsilon, \tau)$ is bounded below as*

$$C_1^1(\delta, \varepsilon, \tau) \geq \frac{p}{q}\log\frac{1}{\varepsilon} - p\log\frac{1}{1-\delta}. \tag{9}$$

Proof. For $1 \leq q \leq p$, suppose that (X, Y) is (p, q)-hypercontractive, which by Lemma 11 implies that (X^n, Y^n) is (p, q)-hypercontractive. Furthermore, assume that $p' \leq q$ which is the same as $q' \leq p$. Then, for any subset $\mathcal{A} \subset \mathcal{X}^n$ and $\mathcal{B} \subset \mathcal{Y}^n$, we have

$$\mathrm{P}_{X^nY^n}\left(\mathcal{A} \times \mathcal{B}\right) \leq \mathrm{P}_{X^n}\left(\mathcal{A}\right)^{\frac{1}{p'}}\mathrm{P}_{Y^n}\left(\mathcal{B}\right)^{\frac{1}{q}}. \tag{10}$$

We begin by considering a deterministic test where the shared randomness U is constant. Specifically, given a deterministic $(\ell, \delta, \varepsilon, \tau)$-test $T = (f, g)$, denoting[3] $L = 2^\ell$, let $\mathcal{A}_i = f^{-1}(i)$ for $i = 1, ..., L$. Then, $\{\mathcal{A}_1, ..., \mathcal{A}_L\}$ constitutes a partition of $\mathcal{X}^n$. Further, let $\mathcal{B}_i$ denote the set $\{\mathbf{y} \in \mathcal{Y}^n : g(\mathbf{y}, i) = 0\}$, namely the set of $\mathbf{y}$ where $\mathcal{P}_2$ declares $\mathcal{H}_0^1$ upon receiving i from $\mathcal{P}_1$. It follows that

$$\begin{aligned}1 - \delta &\leq \sum_{i=1}^{L}\mathrm{P}_{X^nY^n}\left(\mathcal{A}_i \times \mathcal{B}_i\right)\\ &\leq \sum_{i=1}^{L}\mathrm{P}_{X^n}\left(\mathcal{A}_i\right)^{\frac{1}{p'}}\mathrm{P}_{Y^n}\left(\mathcal{B}_i\right)^{\frac{1}{q}},\end{aligned}$$

where the previous inequality uses (10). Upon bounding the right-side using Hölder's inequality, we get

$$\begin{aligned}1 - \delta &\leq \sum_{i=1}^{L}\left(\mathrm{P}_{X^n}\left(\mathcal{A}_i\right)\mathrm{P}_{Y^n}\left(\mathcal{B}_i\right)\right)^{\frac{1}{q}}\mathrm{P}_{X^n}\left(\mathcal{A}_i\right)^{\frac{1}{p'}-\frac{1}{q}}\\ &\leq \left(\sum_{i=1}^{L}\mathrm{P}_{X^n}\left(\mathcal{A}_i\right)\mathrm{P}_{Y^n}\left(\mathcal{B}_i\right)\right)^{\frac{1}{q}}\left(\sum_{i=1}^{L}\mathrm{P}_{X^n}\left(\mathcal{A}_i\right)^{q'\left(\frac{1}{p'}-\frac{1}{q}\right)}\right)^{\frac{1}{q'}}\\ &\leq \varepsilon^{\frac{1}{q}}\left(\sum_{i=1}^{L}\mathrm{P}_{X^n}\left(\mathcal{A}_i\right)^{q'\left(\frac{1}{p'}-\frac{1}{q}\right)}\right)^{\frac{1}{q'}},\end{aligned}$$

[3] With a slight abuse of notation, we denote the one-way communication protocol by a mapping f. Further, we assume that f and g are measurable.

where the previous inequality uses the requirement $\mathbb{P}_{\mathcal{H}_1^1}\left(\text{Declare } \mathcal{H}_0^1\right) \le \varepsilon$. Noting that $q'(1/p'-1/q) = 1-q'/p$, the assumption $q' \le p$ and Hölder's inequality imply

$$\sum_{i=1}^{L} \mathrm{P}_{X^n}\left(\mathcal{A}_i\right)^{q'\left(\frac{1}{p'}-\frac{1}{q}\right)} \le L^{\frac{q'}{p}}.$$

Combining the bounds above, we get

$$1-\delta \le \varepsilon^{\frac{1}{q}} L^{\frac{1}{p}},$$

which completes the proof.

When shared randomness is available, we follow the procedure above for the deterministic test obtained by conditioning on the shared randomness V; let $(\mathcal{A}_i^V, \mathcal{B}_i^V)$, $1 \le i \le L$, denote the counterpart of $(\mathcal{A}_i, \mathcal{B}_i)$ above for shared randomness V. Proceeding as before, we have

$$\begin{aligned}1-\delta &\le \mathbb{E}_V\left[\left(\sum_{i=1}^{L} \mathrm{P}_{X^n}\left(\mathcal{A}_i^V\right)\mathrm{P}_{Y^n}\left(\mathcal{B}_i^V\right)\right)^{\frac{1}{q}}\right.\\ &\qquad\left.\left(\sum_{i=1}^{L} \mathrm{P}_{X^n}\left(\mathcal{A}_i^V\right)^{q'\left(\frac{1}{p'}-\frac{1}{q}\right)}\right)^{\frac{1}{q'}}\right]\\ &\le \mathbb{E}_V\left[\left(\sum_{i=1}^{L} \mathrm{P}_{X^n}\left(\mathcal{A}_i^V\right)\mathrm{P}_{Y^n}\left(\mathcal{B}_i^V\right)\right)^{\frac{1}{q}}\right] L^{\frac{1}{p}}.\end{aligned}$$

It follows from Jensen's inequality that

$$\begin{aligned}1-\delta &\le \mathbb{E}_V\left[\sum_{i=1}^{L} \mathrm{P}_{X^n}\left(\mathcal{A}_i^V\right)\mathrm{P}_{Y^n}\left(\mathcal{B}_i^V\right)\right]^{\frac{1}{q}} L^{\frac{1}{p}}.\\ &\le \varepsilon^{\frac{1}{q}} L^{\frac{1}{p}},\end{aligned}$$

which completes the proof of Theorem 12. □

Theorem 13 (Lower bound 2)**.** *Let $\delta, \varepsilon \in (0,1)$ and (p,q) be such that $1 \ge q \ge 0 \ge q' \ge p$ and (X,Y) is (p,q)-reverse hypercontractive. Then,*

$$C_1^1(\delta,\varepsilon,\tau) \ge \frac{p}{q}\log\frac{1}{1-\varepsilon} - p\log\frac{1}{\delta}. \tag{11}$$

Proof. For $1 \ge q > p$, suppose that (X,Y) is (p,q)-reverse hypercontractive, which with Lemma 11 implies that (X^n, Y^n) is (p,q)-reverse hypercontractive. Furthermore, assume that $q' \ge p$. Then, for any subset $\mathcal{A} \subset \mathcal{X}^n$ and $\mathcal{B} \subset \mathcal{Y}^n$, for $0 \le \theta \le 1$ we have

$$\mathrm{P}_{X^nY^n}\left(\mathcal{A}\times\mathcal{B}\right)^{\theta} \ge \mathrm{P}_{X^n}\left(\mathcal{A}\right)^{\theta\left(\frac{p-1}{p}\right)}\mathrm{P}_{Y^n}\left(\mathcal{B}\right)^{\theta\frac{1}{q}}. \tag{12}$$

We only provide a proof for deterministic tests; the extension to the case when shared randomness is used can be completed as in the proof of Theorem 12. Given a deterministic $(\ell,\delta,\varepsilon,\tau)$-test $T=(f,g)$, let $\mathcal{A}_i = f^{-1}(i)$ for $i=1,...,L=2^{\ell}$, and let $\mathcal{B}_i$ denote the set $\{\mathbf{y}\in\mathcal{Y}^n : g(\mathbf{y},i)=1\}$. It follows that

$$\begin{aligned}1-\varepsilon &\le \sum_{i=1}^{L}\mathrm{P}_{X^n}\left(\mathcal{A}_i\right)\mathrm{P}_{Y^n}\left(\mathcal{B}_i\right)\\ &\le \sum_{i=1}^{L}\mathrm{P}_{X^n}\left(\mathcal{A}_i\right)^{1-\theta\left(\frac{p-1}{p}\right)}\mathrm{P}_{Y^n}\left(\mathcal{B}_i\right)^{1-\frac{\theta}{q}}\mathrm{P}_{X^nY^n}\left(\mathcal{A}_i\times\mathcal{B}_i\right)^{\theta},\end{aligned}$$

where the previous inequality uses (12). Upon bounding the right-side using Hölder's inequality, we get

$$\begin{aligned}1-\varepsilon &\le \left(\sum_{i=1}^{L}\left(\mathrm{P}_{X^n}\left(\mathcal{A}_i\right)^{1-\theta\left(\frac{p-1}{p}\right)}\mathrm{P}_{Y^n}\left(\mathcal{B}_i\right)^{1-\frac{\theta}{q}}\right)^{\frac{1}{1-\theta}}\right)^{1-\theta}\\ &\qquad\left(\sum_{i=1}^{L}\mathrm{P}_{X^nY^n}\left(\mathcal{A}_i\times\mathcal{B}_i\right)\right)^{\theta}\\ &\le \left(\sum_{i=1}^{L}\mathrm{P}_{X^n}\left(\mathcal{A}_i\right)^{1+\left(\frac{\theta}{p(1-\theta)}\right)}\mathrm{P}_{Y^n}\left(\mathcal{B}_i\right)^{\frac{q-\theta}{q(1-\theta)}}\right)^{1-\theta}\delta^{\theta},\end{aligned}$$

where the previous inequality uses the requirement $\mathbb{P}_{\mathcal{H}_0^1}\left(\text{Declare } \mathcal{H}_1^1\right) \le \delta$. Choosing $\theta = q$, the assumption $q' \ge p$ together with Hölder's inequality implies

$$\left(\sum_{i=1}^{L}\mathrm{P}_{X^n}\left(\mathcal{A}_i\right)^{1+\frac{q}{p(1-q)}}\right)^{1-q} \le L^{\frac{-q}{p}}.$$

Combining the bounds above, we get

$$1-\varepsilon \le \delta^{q} L^{\frac{-q}{p}},$$

which completes the proof. □

To obtain tight lower bounds for one-dimensional X and Y jointly Gaussian, we need to optimize our lower bounds over the entire hypercontractivity and reverse hypercontractivity ribbon. We rely on the following characterizations of the hypercontractivity and the reverse hypercontractivity ribbons.

Lemma 14 (*cf.* [17],[27, Theorem 11.23])**.** *Let X and Y be one-dimensional with joint distribution given by* (1)*. For $1 \le q \le p$, (X,Y) is (p,q)-hypercontractive if and only if*

$$\frac{q-1}{p-1} \ge \rho^2. \tag{13}$$

Furthermore, for $1 \ge q \ge p$, (X,Y) is (p,q)-reverse hypercontractive if and only if

$$\frac{1-q}{1-p} \ge \rho^2. \tag{14}$$

Theorem 3 is obtained by maximizing the right-sides of (9) and (11), respectively, over the set of (p,q) satisfying (13) and (14); the upper bound is from Theorem 1. It suffices to show the lower bound for any fixed distribution in $\mathcal{H}_0$ and we pick $\mathbb{P}_{\tau}$.

Proof of Theorem 3. Assume first that $\varepsilon^{\frac{1-\tau}{1+\tau}} \le 1-\delta$. Using the characterization in Lemma 14, (X,Y) with joint distribution $\mathbb{P}_{\tau}$ is (p,q)-hypercontractive for any p and q satisfying

$$\begin{aligned}p &= 1+w,\\ q &= 1+\tau^2 w,\end{aligned}$$

for any $w \ge 0$. Inserting this choice of (p,q) in the lower bound of Theorem 12, we get for any $(\ell,\delta,\varepsilon,\tau)$-test that

$$\ell \ge \frac{1+w}{1+\tau^2 w}\log\frac{1}{\varepsilon} - (1+w)\log\frac{1}{1-\delta}.$$

For brevity, we let $\xi := \log(1-\delta)/\log\varepsilon$; our assumption $\varepsilon^{\frac{1-\tau}{1+\tau}} \leq 1-\delta$ is equivalent to $\xi \leq (1-\tau)/(1+\tau)$. To obtain the tightest lower bound, we maximize $(1+w)/(1+\tau^2 w) - \xi(1+w)$ over $w \geq 0$. The maximum is obtained at w^* given by

$$w^* = \frac{1}{\tau^2}\left(\sqrt{\frac{1-\tau^2}{\xi}} - 1\right),$$

provided $\xi \leq 1-\tau^2$, which holds since $1-\tau^2 \geq (1-\tau)/(1+\tau)$. Furthermore, the corresponding optimal p^* and q^* satisfy $p^{*\prime} \leq q^*$ if and only if

$$\tau^2 \leq \left(\sqrt{\frac{1-\tau^2}{\xi}} - 1\right)^2,$$

which is satisfied when $\xi \leq (1-\tau)/(1+\tau)$. Thus,

$$\begin{aligned}\ell &\geq \left(\log\frac{1}{\varepsilon}\right)\left(\frac{1+w^*}{1+\tau^2 w^*} - \xi(1+w^*)\right)\\ &= \frac{1}{\tau^2}\left(\sqrt{\log\frac{1}{\varepsilon}} - \sqrt{\left(1-\tau^2\right)\log\frac{1}{1-\delta}}\right)^2. \qquad (15)\end{aligned}$$

The first part of Theorem 3 follows from (15) and Theorem 1.

To get the second part of Theorem 3, we obtain a replacement for (15) using the reverse hypercontractivity part of Lemma 14. Specifically, (X,Y) with joint distribution $\mathbb{P}_\tau$ is (p,q)-reverse hypercontractive for any p and q satisfying

$$\begin{aligned}p &= 1-w,\\ q &= 1-\tau^2 w,\end{aligned}$$

for any $\frac{1}{\tau^2} \geq w \geq 0$ since q must be greater than or equal to 0. Inserting this choice of (p,q) in the lower bound of Theorem 13, we get for any $(\ell,\delta,\varepsilon,\tau)$-test that

$$\ell \geq \frac{1-w}{1-\tau^2 w}\log\frac{1}{1-\varepsilon} - (1-w)\log\frac{1}{\delta}.$$

We maximize the right-side of the above inequality subject to $w \leq \frac{1}{\tau^2}$. The maximum is obtained at w^* given by

$$w^* = \frac{1}{\tau^2}\left(1 - \sqrt{\frac{(1-\tau^2)\log\frac{1}{1-\varepsilon}}{\log\frac{1}{\delta}}}\right).$$

Note that $w^* \leq \frac{1}{\tau^2}$ is satisfied for every δ and ε, and the additional assumption $\varepsilon^{\frac{1-\tau}{1+\tau}} \leq 1-\delta$ of the first part of Theorem 3 is not required for the second part. Thus,

$$\begin{aligned}\ell &\geq \frac{1-w^*}{1-\tau^2 w^*}\log\frac{1}{1-\varepsilon} - (1-w^*)\log\frac{1}{\delta}\\ &= \frac{1}{\tau^2}\left(\sqrt{\log\frac{1}{1-\varepsilon}} - \sqrt{\left(1-\tau^2\right)\log\frac{1}{\delta}}\right)^2,\end{aligned}$$

which together with Theorem 1 yields the second part of Theorem 3. □

Finally, we exploit tensorization property in Lemma 11 to provide a matching lower bound for Theorem 2 in the result below.

Theorem 15. *For $0 < \tau \leq 1$,*

1) *for $\delta \in (0,1)$ with ε such that $\delta + \varepsilon^{\frac{1-\tau}{1+\tau}} \leq 1$, we have*

$$C_d^1(\delta,\varepsilon,\tau) \geq \frac{d}{\tau^2}\left(\sqrt{\log\frac{1}{\varepsilon}} - \sqrt{\left(1-\frac{\tau^2}{d}\right)\log\frac{1}{1-\delta}}\right)^2;$$

2) *for $\delta, \epsilon \in (0,1)$,*

$$C_d^1(\delta,\varepsilon,\tau) \geq \frac{d}{\tau^2}\left(\sqrt{\log\frac{1}{1-\varepsilon}} - \sqrt{\left(1-\frac{\tau^2}{d}\right)\log\frac{1}{\delta}}\right)^2.$$

Proof. We consider a different problem where the observation of $\mathcal{P}_1$ remains the same but we provide more information to $\mathcal{P}_2$. Specifically, $\mathcal{P}_1$ observes i.i.d. copies of $X = (X(1),\ldots,X(d))$ and $\mathcal{P}_2$ observes i.i.d. copies of $Y = (Y(1),\ldots,Y(d))$ where for $i = 1,\ldots,d$,

$$\mathbb{E}[Y(i)|X] = \rho(i)X(i).$$

Note that in our original problem the observation of $\mathcal{P}_2$ are i.i.d. copies of $Y(1)+\ldots+Y(d)$. With this modified observation for $\mathcal{P}_2$, we consider the hypothesis testing problem of $\mathcal{H}_0^d$ versus $\mathcal{H}_1^d$ as before. Denote by $\widetilde{C}_d^1(\delta,\varepsilon,\tau)$ the minimum ℓ such that we can find a 1-interactive $(\ell,\delta,\varepsilon,\tau)$-test for this modified problem. Since the observation for the former problem can be obtained from the latter problem as well, we have

$$C_d^1(\delta,\varepsilon,\tau) \geq \widetilde{C}_d^1(\delta,\varepsilon,\tau).$$

Furthermore, with $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$, the proof of Theorem 12 applies to the modified problem as well, and we obtain the following bound:

$$\widetilde{C}_d^1(\delta,\varepsilon,\tau) \geq \frac{p}{q}\log\frac{1}{\varepsilon} - p\log\frac{1}{1-\delta},$$

where (X^n,Y^n) is (p,q)-hypercontractive. By Lemma 11 and Lemma 14, we can parameterize p and q as

$$\begin{aligned}p &= 1+w,\\ q &= 1+\rho_{\max}^2 w,\end{aligned}$$

with $w \geq 0$ and $\rho_{\max}^2 := \max_{i=1}^d \rho(i)^2$. Proceeding as in the proof of Theorem 3, we get

$$\widetilde{C}_d^1(\delta,\varepsilon,\tau) \geq \frac{1}{\rho_{\max}^2}\left(\sqrt{\log\frac{1}{\varepsilon}} - \sqrt{\left(1-\rho_{\max}^2\right)\log\frac{1}{1-\delta}}\right)^2.$$

Note that we can choose any ρ such that $\|\rho\|_2 \geq \tau$. Among all such ρs, the minimum value of $\rho_{\max}$ is attained by ρ with $\rho(i)^2 = \tau^2/d$. Using this value for ρ, we get

$$\widetilde{C}_d^1(\delta,\varepsilon,\tau) \geq \frac{d}{\tau^2}\left(\sqrt{\log\frac{1}{\varepsilon}} - \sqrt{\left(1-\frac{\tau^2}{d}\right)\log\frac{1}{1-\delta}}\right)^2,$$

which completes the proof of the first part of Theorem 15. The proof of the second part is completed similarly by using the tensorization property of the reverse hypercontractivity ribbon. □

VI. Lower bounds for testing and estimation with interactive communication

The following bound was derived in [19] using a strong data processing inequality; the statement follows from [19, Theorem 7.1 and 7.2].[4]

Lemma 16. *(see [19]) For $\rho \in [-1,1]^d$ and any interactive communication protocol π with inputs X^n and Y^n for parties $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively, we have*

$$D(\mathbb{P}_\rho \| \mathbb{P}_0) \leq \rho_{\max}^2 |\pi|,$$

where $\rho_{\max}^2 = \max_{i\in[d]} \rho(i)^2$, $\mathbb{P}_\rho$ denotes the distribution in (1) *and $D(P||Q)$ denotes the Kullback-Leibler (KL) divergence between distributions P and Q.*

We use this lemma to prove Theorems 5 and 6.

A. Proof of Theorem 5 - Lower Bound for Testing

Let $T = (\pi, g)$ constitute an $(\ell, \delta, \varepsilon, \tau)$-test. Denote by P the distribution of (Y^n, Π, V) under $\mathcal{H}_0^d$ and by Q the distribution of (Y^n, Π, V) under $\mathcal{H}_1^d$.

Then,

$$p_0 \triangleq P\left(g(Y^n, \Pi, V) = 0\right) \geq 1 - \delta,$$
$$q_0 \triangleq Q\left(g(Y^n, \Pi, V) = 0\right) \leq \varepsilon.$$

Then, by the data processing inequality applied using the channel $\mathbb{1}\{g(Y^n, \Pi, V) = 0\}$, we have

$$\begin{aligned} D(P\|Q) &\geq D(p_0||q_0) \\ &\geq p_0 \log \frac{1}{q_0} - 1 \\ &\geq (1-\delta)\log\frac{1}{\varepsilon} - 1, \end{aligned} \tag{16}$$

where $D(p_0||q_0)$ denotes the KL divergence between probability mass functions $(p_0, 1-p_0)$ and $(q_0, 1-q_0)$, and we have used the bound $h(p_0) \leq 1$ where $h(\cdot)$ denotes the binary entropy function. Furthermore, by Lemma 16 we have $D(P\|Q) \leq \rho_{\max}^2 \ell$, which with the previous bound gives

$$\ell \geq \frac{1}{\rho_{\max}^2}\left((1-\delta)\log\frac{1}{\varepsilon} - 1\right).$$

To obtain the tightest possible bound, we choose ρ such that $\rho(i)^2 = \tau^2/d$ for every $1 \leq i \leq d$, which yields

$$\ell \geq \frac{d}{\tau^2}\left[(1-\delta)\log\frac{1}{\varepsilon} - 1\right].$$

B. Proof of Theorem 6 - Lower Bound for Estimation

We provide lower bounds for the estimation error using Fano's method. Using the *Gilbert-Varshamov construction* (see, for instance, [14, Problem 5.5]) we can find $m \geq 2^{d(1-h(1/4))} \geq 2^{d/6}$ vectors $u_1, ..., u_m \in \{-1,+1\}^d$ such that their Hamming distance $d_H(u_i, u_j) \geq d/8$ for every $i \neq j$. For every $\Delta > 0$, the vectors $\rho_i := \frac{\Delta}{\sqrt{d}} \cdot u_i$, $1 \leq i \leq m$ satisfy

$$\min_{i,j\in[m]:i\neq j} \|\rho_i - \rho_j\|_2^2 = \frac{4\Delta^2 d_H(u_i,u_j)}{d} \geq \frac{\Delta^2}{2},$$
$$\max_i \rho_j(i)^2 = \frac{\Delta^2}{d}, \quad \forall 1 \leq j \leq m.$$

Consider an r-interactive (ℓ, τ)-estimate $(\pi, \widehat{\rho})$. We use the estimator $\widehat{\rho}$ to resolve between the hypotheses $\mathcal{H}_j, j \in [m]$ where under $\mathcal{H}_j$, $X \in \mathbb{R}^d$ and $Y \in \mathbb{R}$ are jointly Gaussian and

$$\mathbb{E}_{\rho_j}[Y|X] = \sum_{i=1}^d \rho_j(i) X(i).$$

Consider the test that declares[5] $\mathcal{H}_j$ if $\|\widehat{\rho} - \rho_j\|_2^2 < \Delta^2/8$; the output is unique since $\|\rho_i - \rho_j\|_2^2 \geq \Delta^2/2$ for every $i \neq j$. By Markov's inequality, the probability of error for this test under $\mathbb{P}_{\rho_j}$ is bounded above by

$$\mathbb{P}_{\rho_j}\left(\|\widehat{\rho} - \rho_j\|_2^2 \geq \Delta^2/8\right) \leq \frac{8}{\Delta^2}\mathbb{E}_{\rho_j}\left[\|\widehat{\rho} - \rho_j\|_2^2\right].$$

Therefore, denoting by P_e^* the minimum average probability of error for this hypothesis testing problem under uniform prior on the hypotheses, we get from (2) that

$$\begin{aligned} \tau^2 &\geq \max_{j\in[m]} \frac{\Delta^2}{8}\mathbb{P}_{\rho_j}\left(\|\widehat{\rho} - \rho_j\|_2^2 \geq \Delta^2/8\right) \\ &\geq \frac{\Delta^2}{8m}\sum_{i=1}^m \mathbb{P}_{\rho_j}\left(\|\widehat{\rho} - \rho_j\|_2^2 \geq \Delta^2/8\right) \\ &\geq \frac{\Delta^2}{8}P_e^*. \end{aligned}$$

By Fano's inequality, we have

$$P_e^* \geq 1 - \frac{C(W)+1}{\log m}, \tag{17}$$

where W denotes the channel with input $j \in \{1, ..., m\}$ and output (Y^n, Π, V) with distribution corresponding to the correlation ρ_j between X^n and Y^n, and $C(W)$ denotes the capacity of channel W. Recall the well-known bound

$$C(W) \leq \min_Q \max_j D(W(\cdot|j)\|Q).$$

We use this bound with Q chosen to be the distribution of (Y^n, Π, V) when the correlation between X and Y is $\rho = 0$.

Then, by Lemma 16 we have

$$D(W(\cdot|j)\|Q) \leq \max_{1\leq i\leq d} \rho_j(i)^2 \ell = \frac{\Delta^2 \ell}{d}.$$

Combining the bounds above yields

$$\tau^2 \geq \frac{\Delta^2}{8}\left(1 - \frac{6(\Delta^2\ell/d + 1)}{d}\right).$$

In particular, for $d \geq 12$, setting $\Delta^2\ell/d^2 = 1/24$ gives $\ell \geq \frac{d^2}{768\tau^2}$. Note that for $d < 12$, for an appropriate constant c, the lower bound $\ell \geq cd^2/\tau^2$ holds since we already have an

[4] [19, Theorem 7.1 and 7.2] hold for interactive protocols with shared randomness and are stated for one-dimensional X and Y. However, the result "tensorizes" (see [19, Lemma 9.3]) and gives the general form in Lemma 16.

[5] In the remainder of this proof, with an abuse of notation, we denote the random variable $\widehat{\rho}(Y^n, \Pi, V)$ by $\widehat{\rho}$.

$\Omega(d/\tau^2)$ lower bound for the testing problem. This completes the proof. □

VII. Extensions and discussion

We conclude with a discussion on various extensions of our result, and state some of these extensions without proof (the proofs are very similar to the others in the paper).

First, we note that while the hypercontractivity based lower bound yields a tight dependence on δ or ε separately in Theorem 3, it does not characterize the joint dependence on δ and ε simultaneously. Interestingly, when we allow $\delta > 1/2$ and have ε small, such a joint characterization is possible. Specifically, for $d = 1$, consider the simple binary hypothesis problem of correlation ρ versus correlation 0. The test we use in Theorem 7 for resolving between $\mathcal{H}_0^+$ and $\mathcal{H}_1^1$ with a different choice for θ and r yields a joint characterization of one-way minimum communication needed for this problem (see [29]). Interestingly, the overall communication is below $(1/\rho^2)\max\{\log 1/\varepsilon, \log 1/\delta\}$.

Theorem 17. *For $d = 1$, $0 < \rho \le 1$, $\delta \in (1/2, 1)$, and ε such that $\delta + \varepsilon^{\frac{1-\tau}{1+\tau}} \le 1$, the minimum one-way communication needed to test if correlation is ρ or 0 is given by*

$$\frac{1}{\rho^2} \cdot \left(\sqrt{\log\frac{1}{\varepsilon}} - \sqrt{\left(1-\rho^2\right)\log\frac{1}{1-\delta}}\right)^2 + O\left(\sqrt{\log\frac{1}{\varepsilon}\log\frac{1}{1-\delta}}\right).$$

In another direction, we can consider the simple binary hypothesis testing problem of $\rho = \rho_0$ versus $\rho = \rho_1$, where $1 > \rho_0 > \rho_1 > 0$. Once again, by modifying the parameters for the test used in Theorem 7, we get a generalization of our results for $d = 1$ to the case $\rho_1 > 0$. Specifically, in this case, the probability of error requirements as in (4) and (5) yield

$$1 - e^{2^k Q(r)} - Q\left(\frac{r(\rho_0 - \theta)}{\sqrt{1-\rho_0^2}}\right) - \eta \ge 1 - \delta,$$
$$2^{k+1} Q(r) Q\left(\frac{r(\theta - \rho_1)}{\sqrt{1-\rho_1^2}}\right) \le \epsilon.$$

Proceeding in a similar manner as our earlier analysis and upon setting $\theta \in (\rho_1, \rho_0)$ and

$$r^2 = \frac{2\ln 2}{(\rho_0 - \rho_1)^2}\left(\sqrt{(1-\rho_1^2)\log\frac{1}{\varepsilon} + \log\ln\frac{3}{\delta} + 1} + \sqrt{(1-\rho_0^2)\log\frac{3}{\delta}}\right)^2,$$

we obtain the following result.

Theorem 18. *For $d = 1$, $\delta, \varepsilon \in (0, 1)$, $0 < \rho_1 < \rho_0 < 1$, we can find a distributed test for $\rho = \rho_0$ versus $\rho = \rho_1$ that uses one-way communication of less than*

$$\frac{1}{(\rho_0 - \rho_1)^2}\left(\sqrt{(1-\rho_1^2)\log\frac{1}{\varepsilon}} + \sqrt{(1-\rho_0^2)\log\frac{1}{\delta}}\right)^2 + O\left(\sqrt{\log\frac{1}{\varepsilon} + \log\ln\frac{1}{\delta}}\sqrt{\log\frac{1}{\delta}}\right) \textit{ bits.}$$

We note that [20] derived an upper bound for the error exponent for this problem when communication length per sample is fixed. While the result there was stated for error exponent, the main bound [20, Equation (48)] shows that the one-way communication needed for testing $\rho = \rho_0$ versus $\rho = \rho_1$ must exceed

$$\left(\frac{(1-\rho_1)^2}{(\rho_0 - \rho_1)^2} - 1\right)\left(\max\left\{(1-\delta)\log\frac{1}{\varepsilon}, (1-\varepsilon)\log\frac{1}{\delta}\right\} - 1\right),$$

which almost matches the communication requirement for our scheme. However, we do not account for the number of samples in our scheme, and it may not attain the upper bound on the error exponent in [20].

Acknowledgements

The authors thank Ofer Shayevitz and Uri Hadar for useful discussions and Prathamesh Mayekar for suggesting the use of median trick in the proof of upper bound for general d. Thanks are also due to the associate editor and the reviewers whose comments improved the presentation of the paper.

References

[1] J. Acharya, C. Canonne, C. Freitag, and H. Tyagi, "Test without trust: Optimal locally private distribution testing," in *Proceedings of Machine Learning Research*, ser. Proceedings of Machine Learning Research, vol. 89, 16–18 Apr 2019, pp. 2067–2076.

[2] J. Acharya, C. L. Canonne, Y. Han, Z. Sun, and H. Tyagi, "Domain compression and its application to randomness-optimal distributed goodness-of-fit," in *Conference on Learning Theory*. PMLR, 2020, pp. 3–40.

[3] J. Acharya, C. L. Canonne, and H. Tyagi, "Inference under information constraints: Lower bounds from chi-square contraction," *Proceedings of Machine Learning Research vol*, vol. 99, pp. 1–15, 2019.

[4] ——, "Inference under information constraints II: Communication constraints and shared randomness," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7856–7877, 2020.

[5] J. Acharya, C. Daskalakis, and G. Kamath, "Optimal testing for properties of distributions," in *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., 2015, pp. 3591–3599.

[6] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, no. 4, pp. 533–542, July 1986.

[7] ——, "Common randomness in information theory and cryptography–part II: CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 225–240, January 1998.

[8] R. Ahlswede and P. Gacs, "Spreading of sets in product spaces and hypercontraction of the markov operator," *Ann. Probab.*, vol. 4, no. 6, pp. 925–939, December 1976.

[9] A. Andoni, T. Malkin, and N. S. Nosatzki, "Two party distribution testing: Communication and security," in *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[10] W. Beckner, "Inequalities in Fourier analysis," *Ann. of Math.*, vol. 102, no. 1, pp. 159–182, July 1975.

[11] A. Bogdanov and E. Mossel, "On extracting common random bits from correlated sources," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6351–6355, Oct 2011.

[12] A. Bonami, "Etudes des coefficients Fourier des fonctiones de $L^p(G)$," *Ann. Inst. Fourier*, vol. 20, no. 2, pp. 335–402, 1970.

[13] C. Borell, "Positivity improving operators and hypercontractivity," *Mathematische Zeitschrift*, no. 180, pp. 225–234, 1982.

[14] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless channels. 2nd edition*. Cambridge University Press, 2011.

[15] Y. Dagan and O. Shamir, "Detecting correlations with little memory and communication," in *Conference On Learning Theory*, 2018, pp. 1145–1198.

[16] R. Durrett, *Probability: theory and examples*. Cambridge university press, 2019, vol. 49.

[17] L. Gross, "Logarithmic sobolev inequalities," *American Journal of Mathematics*, vol. 97, no. 4, pp. 1061–1083, 1975.

[18] V. Guruswami and J. Radhakrishnan, "Tight bounds for communication-assisted agreement distillation," in *Proceedings of the 31st Conference on Computational Complexity*, 2016, pp. 6:1–6:17.

[19] U. Hadar, J. Liu, Y. Polyanskiy, and O. Shayevitz, "Communication complexity of estimating correlations," in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, 2019, pp. 792–803.

[20] ——, "Error exponents in distributed hypothesis testing of correlations," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 2674–2678.

[21] U. Hadar and O. Shayevitz, "Distributed estimation of gaussian correlations," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5323–5338, 2019.

[22] T. S. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, October 1998.

[23] G. Katz, P. Piantanida, and M. Debbah, "Collaborative distributed hypothesis testing with general hypotheses," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 1705–1709.

[24] J. Liu, P. Cuff, and S. Verdú, "Secret key generation with limited interaction," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7358–7381, November 2017.

[25] E. Mossel, K. Oleszkiewicz, and A. Sen, "On reverse hypercontractivity," *Geometric and Functional Analysis*, vol. 23, no. 3, pp. 1062–1097, June 2013.

[26] C. Nair, "Equivalent formulations of hypercontractivity using information measures," *Proceedings of International Zürich Seminar on Communications*, 2014.

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

[28] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6282–6303, 2012.

[29] K. R. Sahasranand and H. Tyagi, "Extra samples can reduce the communication for independence testing," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 2316–2320.

[30] H. M. H. Shalaby and A. Papamarcou, "Multiterminal detection with zero-rate data compression," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 254–267, March 1992.

[31] S. Sreekumar and D. Gündüz, "Distributed hypothesis testing over discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2044–2066, 2019.

[32] ——, "Strong converse for testing against independence over a noisy channel," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 1283–1288.

[33] T.S.Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 759–772, November 1987.

[34] H. Tyagi and S. Watanabe, "Converses for secret key agreement and secure computing," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4809–4827, September 2015.

[35] ——, "A bound for multiparty secret key agreement and implications for a problem of secure computing," in *EUROCRYPT*, 2014, pp. 369–386.

[36] ——, "Strong converse using change of measure arguments," *IEEE Transactions on Information Theory*, vol. 66, no. 2, pp. 689–703, 2019.

[37] M. Wigger and R. Timo, "Testing against independence with multiple decision centers," in *International Conference on Signal Processing and Communications (SPCOM)*, June 2016, pp. 1–5.

[38] Y. Xiang and Y. H. Kim, "Interactive hypothesis testing with communication constraints," in *50th Annual Allerton Conference on Communication, Control, and Computing*, October 2012, pp. 1065–1072.

[39] ——, "Interactive hypothesis testing against independence," in *2013 IEEE International Symposium on Information Theory*, July 2013, pp. 2840–2844.

[40] W. Zhao and L. Lai, "Distributed testing against independence with multiple terminals," in *52nd Annual Allerton Conference on Communication, Control, and Computing*, September 2014, pp. 1246–1251.

K. R. Sahasranand received his B.Tech in Computer Science and Engineering from Amrita University, Kerala in 2009 and worked as a Scientist/Engineer at Indian Space Research Organisation, Bangalore, 2010-'11. He completed his Master's in Electrical Communication Engineering from Indian Institute of Science, Bangalore in 2015 where he is currently a PhD student. His research interests include information theory, detection and estimation, and signal processing.

Himanshu Tyagi received the B.Tech. degree in electrical engineering and the M.Tech. degree in communication and information technology from the Indian Institute of Technology, Delhi, India, in 2007, and the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2013. From 2013 to 2014, he was a Postdoctoral Researcher with the Information Theory and Applications (ITA) Center, University of California at San Diego, San Diego, CA, USA. Since January 2015, he has been a Faculty Member with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore. He received the Indian National Science Academy (INSA) Medal for Young Scientist 2020. His research interests include information theory and its application in cryptography, statistics, machine learning, and computer science. Also, he is interested in communication and automation for city-scale systems.