Distributed Signal Detection under Communication Constraints

author names withheld

Editor: Under Review for COLT 2020

Abstract

Independent draws from a d-dimensional spherical Gaussian distribution are distributed across users, each holding one sample. A central server seeks to distinguish between the two hypotheses: the distribution has zero mean, or the mean has ℓ_2-norm at least ε, a pre-specified threshold. However, the users can each transmit at most ℓ bits to the server. This is the problem of detecting whether an observed signal is simply white noise in a distributed setting. We study this distributed testing problem with and without the availability of a common randomness shared by the users. We design schemes with and without such shared randomness which achieve sample complexities. We then obtain lower bounds for protocols with public randomness, tight when $\ell = O(1)$. We finally conclude with several conjectures and open problems.

Keywords: Hypothesis testing, signal detection, distributed algorithms, high-dimensional distributions, Gaussian Location Model, mean testing, communication constraints

1. Introduction

We consider the problem of signal detection in Gaussian noise, a natural composite hypothesis testing problem for the Gaussian location model (GLM), defined as follows. Given a distance-separation parameter $\varepsilon \in (0, 1]$ and n i.i.d. samples from a high-dimensional spherical Gaussian $\mathcal{G}(\mu, \mathbb{I}_d)$ with unknown mean vector $\mu \in \mathbb{R}^d$, we seek to test whether $\mu = \mathbf{0}$ or $\|\mu\|_2 > \varepsilon$, namely we need to test if we are observing just noise or a signal with significant "power" in a Gaussian noise.

It is well-known that a threshold test using the squared ℓ_2 norm of the empirical estimator of μ as the statistic yields a minimax-optimal test for this question, with sample complexity $n = \Theta(\sqrt{d}/\varepsilon^2)$ – a quadratic (in the dimension) improvement over the sample complexity $\Theta(d/\varepsilon^2)$ of estimating the mean vector μ up to an accuracy ε. However, the simple statistic above requires full observation of the n samples; the question of sample optimal tests for signal detection using partial observations using a fixed number of bits per sample, to the best of our knowledge, remains largely unresolved.

In this paper, we make significant progress towards answering this question. Specifically, the samples are distributed among n users, each holding one sample. Each user can send only an ℓ-bit message to a centralized server, which upon observing these messages must decide if the signal is present or not. In particular, in the regime $\ell \ll d$, the server only has access to very little information about each of the n samples. We seek to quantify how does this communication constraint affect the sample complexity of the task?

Our results. We consider two distinct variants of this distributed setting. In the first, *private-coin* setting, the messages of the n users are fully independent: upon observing sample x, user i sends a (possibly randomized) message $y = W_i(x)$ to the central server, where W_i is a function depending on user i's private randomness only. In the second variant, the *public-coin* setting, however, the users have access to a common random seed U, which is independent of their observations; so that

the functions $W_1, \ldots, W_n$ are jointly randomized. Note that public-coin protocols are at least as powerful as private-coin ones, as they include the latter as a special case.

We establish the following results. For brevity, we refer to our hypothesis testing problem as the *Gausssian mean testing* problem.

Theorem 1 (Private-Coin Upper Bound) *For $1 \leq \ell \leq d$, there exists a private-coin protocol for distributed Gaussian mean testing under ℓ-bit communication constraints, with $n = O\left(\min\left(\frac{d^{3/2}}{\ell\varepsilon^2}, \frac{d}{\varepsilon^4}\right)\right)$ users.*

As a quick sanity check, we can verify that setting $\ell = d$ retrieves the full-observation (centralized) sample complexity $\Theta(\sqrt{d}/\varepsilon^2)$. Interestingly, for large enough values of ε (i.e., $\varepsilon \gg d^{-1/4}$), the sample complexity upper bound we obtain exhibits a regime transition at $\ell \approx \varepsilon^2\sqrt{d}$: we discuss this in more detail in Section 5.

Theorem 2 (Public-Coin Upper Bound) *For $1 \leq \ell \leq d$, there exists a public-coin protocol for distributed Gaussian mean testing under ℓ-bit communication constraints, with $n = O\left(\frac{d}{\sqrt{\ell}\varepsilon^2}\right)$ users.*

Here as well, it is immediate to see that this matches the centralized sample complexity for $\ell = d$. Although intuitively reasonable, this fact is not actually immediate, as fully communicating one sample drawn from a d-dimensional Gaussian to the server would require infinitely many bits (and not merely d); even appropriately discretized, one would expect a number of bits depending on both d and ε (e.g., $\Theta(d\log(1/\varepsilon))$. Our results imply that a much coarser discretization of these continuous observations is sufficient to achieve the optimal sample complexity.

We further conjecture the broader optimality of all above upper bounds:

Conjecture 3 (Private- and Public-Coin Optimality) *Theorems 1 and 2 are optimal. Specifically, for any $1 \leq \ell \leq d$, any private-coin (resp., public-coin) protocol for distributed Gaussian mean testing under ℓ-bit communication constraints must have $\Omega\left(\min\left(\frac{d^{3/2}}{\ell\varepsilon^2}, \frac{d}{\varepsilon^4}\right)\right)$ (resp., $\Omega\left(d/(\sqrt{\ell}\varepsilon^2)\right)$) users.*

We discuss this conjecture, and provide evidence for it, in Section 5. We are, however, able to show the following (weaker) lower bounds, which are optimal for small values of ℓ:

Theorem 4 (Public-Coin Lower Bounds) *Let $1 \leq \ell \leq \sqrt{d}/\varepsilon^2$. Any public-coin protocol for distributed Gaussian mean testing under ℓ-bit communication constraints requires $\Omega\left(\min\left(\frac{d}{\ell\varepsilon^2}, \frac{d}{\ell^2\varepsilon^4}\right)\right)$ users.*

Related work. Distributed signal detection has been widely studied in the signal processing and communication community. Primarily, previous works consider the setting where the signal is guaranteed to be one of two possibilities (i.e., a simple hypothesis testing problem), whereas we are concerned with a composite testing problem; furthermore, they focus on the asymptotic regime, aiming to characterize the error exponents as the number of draws goes to ∞. From the pioneering works of Tsitsiklis (Tsitsiklis, 1993), and Varshney (Varshney, 2012; Viswanathan and Varshney, 1997), decentralized detection has also received significant attention in the control and signal processing literature, with main focus on information structure, likelihood ratio tests, and combining local decisions for global inference. Finally, distributed inference under communication constraints was also studied in the information theory community (Ahlswede and Csiszár, 1986; Han,

1987; Han and Amari, 1998) to characterize the error exponents as a function of the communication rate. Recent results in this area have focused on more complex communication models (Wigger and Timo, 2016).

From a computer science and machine learning perspective, the problem of distributed estimation has been considered in a number of recent works (Zhang et al., 2013; Garg et al., 2014; Shamir, 2014; Braverman et al., 2016; Xu and Raginsky, 2017; Han et al., 2018a,b; Barnes et al., 2019; Cai and Wei, 2020; Acharya et al., 2019b,d), while distributed testing of distributions in the finite sample regime was considered in (Acharya et al., 2019b,d; Andoni et al., 2018; Fischer et al., 2018; Diakonikolas et al., 2019). Moreover, several recent papers have been focusing on the sample complexity of signal detection under privacy constraints (Canonne et al., 2019a; Acharya et al., 2019a) (as opposed to communication ones), or in the centralized setting without the spherical identity covariance assumption (Canonne et al., 2019b).

In flavor of results, the works closest to ours are (Acharya et al., 2019b,d), which consider the problem of testing goodness-of-fit of *univariate* discrete distributions under communication constraints. In particular, they establish a *strict separation* in the sample complexity between protocols with access to public randomness, and those where users can only randomize their messages independently. We also note that the aforementioned work of (Han et al., 2018b) contains lower bound for the (related) task of distributed estimation for the Gaussian Location Model (GLM), and some of their arguments rely on tools similar to ours (specifically, Lemma 5). We were not, however, able to reproduce some of their arguments, which appear to sidestep the main technical hurdle we face in our lower bounds.[1]

2. Preliminaries

Throughout the paper, given any $\varepsilon \in (0,1]$ we will denote by $\mathcal{H}_\varepsilon$ the set of alternative hypotheses, i.e., $\mathcal{H}_\varepsilon := \{\, \mathcal{G}(\mu, \mathbb{I}_d) \,:\, \|\mu\|_2 \geq \varepsilon \,\}$. We consider the set of ℓ-bit information constraints, that is, each user must use a (randomized) channel belonging to the family

$$\mathcal{W}_\ell := \{W : \mathbb{R}^d \to \{-1,1\}^\ell\} \tag{1}$$

where $\ell \in \mathbb{N}$. Note that $\mathcal{W}_\ell$ is closed under convex combinations, so that $\overline{\mathcal{W}_\ell} = \mathcal{W}_\ell$.

Useful Result from Probability. We will heavily rely on the following measure change bound, which follows from e.g., Gibbs variational principle, and whose proof is provided for completeness in Appendix C.[2]

Lemma 5 (Subgaussian Measure Change Bound) *For $X \sim \mathcal{G}(\mathbf{0}, \mathbb{I}_d)$ and any function $a(X)$,*

$$\frac{\|\mathbb{E}[a(X)X]\|_2^2}{\mathbb{E}[a(X)]^2} \leq 2\frac{\mathbb{E}_P[a(X)\log a(X)]}{\mathbb{E}_P[a(X)]} + 2\log\frac{1}{\mathbb{E}_P[a(X)]}.$$

In particular, for $a(X) = \mathbb{1}\{X \in A\}$, $\frac{\|\mathbb{E}[a(X)X]\|_2^2}{\mathbb{E}[a(X)]^2} \leq 2\log\frac{1}{P(A)}$.

Remark 6 *In fact, a stronger bound follows from Talagrand's transportation cost-information inequality: $\mathbb{E}_Q\left[\|X\|_2^2\right] \leq \frac{2}{\log e}\log\frac{1}{P(A)}$, where $Q \ll P$ is the conditional distribution of P on A.*

1. In more detail, the argument in (Han et al., 2018b, Appendix C) handles the non-linearities inside an expectation by integrating a Taylor expansion, an approach which is not applicable in this setting.
2. We note that this lemma appears to be equivalent to one of the key "geometric inequalities" underlying Han et al. (2018b); specifically, their Lemma 14.

Note that it is customary to denote the right-side of the bound by $\mathrm{Ent}(a)$, namely, the lemma above can be restated as $\frac{\|\mathbb{E}[a(X)X]\|_2^2}{\mathbb{E}[a(X)]^2} \le 4\,\mathrm{Ent}(a)$.

2.1. Lower bound framework

We recall the following notions from (Acharya et al., 2018), which we will use extensively. Throughout we assume that the family of channels $\mathcal{W}$ consists of channels $W\colon \mathcal{X} \to \mathcal{Y}$ where the input alphabet is $\mathcal{X}$ ($\mathbb{R}^d$ for us) and the output alphabet $\mathcal{Y}$ ($\{0,1\}^\ell$ for us) is finite, and the family of alternative distributions $\mathcal{P}$ used to prove lower bounds and referred to as a *perturbed family*, can be parameterized as $\{\, \mathbf{p}_z \,:\, z \in \mathcal{Z} \,\}$. Let $\mathbf{p}^W$ and $\mathbf{q}^W$, respectively, be the output distributions for a channel W with input distributions $\mathbf{p}$ and $\mathbf{q}$ (that is, $\mathbf{p}^W(y) = \mathbb{E}_{\mathbf{p}}[W(y \mid X)]$ for $y \in \mathcal{Y}$, where the expectation is over $X \sim \mathbf{p}$). Then, for $\delta_z(x) := (\mathbf{p}_z(x) - \mathbf{p}(x))/\mathbf{p}(x)$, we have that the normalized perturbation for the induced perturbed family is given by

$$\delta_Z^W(y) = \frac{1}{\mathbf{p}^W(y)} \cdot \mathbb{E}_{\mathbf{p}}[\delta_Z(X) W(y \mid X)]\,, \qquad y \in \mathcal{Y}\,. \tag{2}$$

Definition 7 ((Acharya et al., 2018, Definition IV.3)) *Consider a perturbed family $\mathcal{P} = \{\mathbf{p}_z : z \in \mathcal{Z}\}$ and a family of channels $\mathcal{W}$. The* induced chi-square fluctuation *of $\mathcal{P}$ for $W \in \mathcal{W}$ is given by*

$$\chi^2\left(W \mid \mathcal{P}\right) := \mathbb{E}_Z\left[\left\|\delta_Z^W\right\|_2^2\right],$$

where Z is distributed uniformly over $\mathcal{Z}$. The n-fold induced decoupled chi-square fluctuation *of $\mathcal{P}$ for $W^n \in \mathcal{W}^n$ is given by*

$$\chi^{(2)}\left(W^n \mid \mathcal{P}\right) := \log \mathbb{E}_{ZZ'}\left[\exp\left(\sum_{i=1}^n \left\langle \delta_Z^{W_i}, \delta_{Z'}^{W_i} \right\rangle\right)\right].$$

When the distribution ζ of Z is not uniform, we replace $\mathcal{P}$ with $\mathcal{P}_\zeta$ in our notation.

Note that, for any given perturbed family and channel, (n times) the induced chi-square fluctuation is an upper bound on the induced decoupled fluctuation. The following is a straightforward variant of (Acharya et al., 2018, Definition IV.4, Lemmata IV.7 and IV.9), obtained by adapting the definition of "ε-perturbed family" to our purpose:

Definition 8 *For a family of channels $\mathcal{W}$, the* (n,ε)-minmax decoupled chi-square fluctuation for $\mathcal{W}$ *is given by*

$$\overline{\chi}^{(2)}(\mathcal{W}^n, \varepsilon) := \inf_{\mathcal{P}_\varepsilon} \sup_{W^n \in \mathcal{W}^n} \chi^{(2)}\left(W^n \mid \mathcal{P}_\varepsilon\right),$$

and the (n,ε)-maxmin decoupled chi-square fluctuation for $\mathcal{W}$ *is given by*

$$\underline{\chi}^{(2)}(\mathcal{W}^n, \varepsilon) := \sup_{W^n \in \mathcal{W}^n} \inf_{\mathcal{P}_\varepsilon} \chi^{(2)}\left(W^n \mid \mathcal{P}_\varepsilon\right),$$

where the infimum is over all ε-perturbed families $\mathcal{P}_\varepsilon$, i.e., all families $\mathcal{P}_\varepsilon \subseteq \mathcal{H}_\varepsilon$, where recall that $\mathcal{H}_\varepsilon := \{\, \mathcal{G}(\mu, \mathbb{I}_d) \,:\, \|\mu\|_2 \ge \varepsilon \,\}$.

Lemma 9 (Minmax decoupled chi-square fluctuation bound for testing) *For $0 < \varepsilon < 1$ and a reference distribution $\mathbf{p}$, the sample complexity $n = n(d,\varepsilon)$ of distinguishing $\mathbf{p}$ from $\mathcal{H}_\varepsilon$ using $\mathcal{W}$ for public-coin protocols must satisfy*

$$\overline{\chi}^{(2)}(\mathcal{W}^n, \varepsilon) \ge c,$$

for some $c > 0$ depending only on the probability of error.

Lemma 10 (Maxmin decoupled chi-square fluctuation bound for testing) *For $0 < \varepsilon < 1$ and a reference distribution $\mathbf{p}$, the sample complexity $n = n(d, \varepsilon)$ of distinguishing $\mathbf{p}$ from $\mathcal{H}_\varepsilon$ using $\mathcal{W}$ for private-coin protocols must satisfy*

$$\underline{\chi}^{(2)}(\overline{\mathcal{W}}^n, \varepsilon) \geq c,$$

for some $c > 0$ depending only on the probability of error (where $\overline{\mathcal{W}}$ denotes the convex hull of $\mathcal{W}$).

Expanding the above definitions, we can get more convenient forms for the quantities at play: for every $W \in \mathcal{W}$,

$$\chi^2\left(W \mid \mathcal{P}_\varepsilon\right) = \sum_{y \in \mathcal{Y}} \mathbb{E}_Z\left[\frac{\mathbb{E}_X[\delta_Z(X) W(y \mid X)]^2}{\mathbb{E}_X[W(y \mid X)]}\right] = \sum_{y \in \mathcal{Y}} \frac{\mathbb{E}_{XX'}[\mathbb{E}_Z[\delta_Z(X)\delta_Z(X')] W(y \mid X) W(y \mid X')]}{\mathbb{E}_X[W(y \mid X)]} \tag{3}$$

where X is drawn according to $\mathbf{p}$. Similarly,

$$\chi^{(2)}\left(W^n \mid \mathcal{P}_\varepsilon\right) = \log \mathbb{E}_{ZZ'}\left[\exp\left(\sum_{i=1}^n \left\langle \delta_Z^{W_i}, \delta_{Z'}^{W_i} \right\rangle\right)\right], \tag{4}$$

where

$$\left\langle \delta_Z^{W_i}, \delta_{Z'}^{W_i} \right\rangle = \sum_{y \in \mathcal{Y}} \frac{\mathbb{E}_X[\delta_Z(X) W(y \mid X)] \mathbb{E}_X[\delta_{Z'}(X) W(y \mid X)]}{\mathbb{E}_X[W(y \mid X)]}. \tag{5}$$

3. Upper Bounds

3.1. Private-Coin Upper Bounds

In this section, we prove Theorem 1, by providing two algorithms, each corresponding to one the two terms in the upper bound.

3.1.1. UPPER BOUND OF $O(d^{3/2}/\ell\varepsilon^2)$ USING DISTRIBUTED SIMULATION

We first show a reduction from Gaussian mean testing to the analogously defined binary product mean testing problem described below.

Distributed binary product mean testing. Given samples from a product distribution $\mathbf{p}$ over $\{-1, 1\}^d$, distinguish between $\mathbf{p}$ being the uniform distribution over $\{-1, 1\}^d$ and $\mathbf{p}$ having mean vector of ℓ_2 norm at least ε. It is well known (see, e.g., (Canonne et al., 2017)) that the sample complexity of solving this problem in the centralized setting is $\Theta(\sqrt{d}/\varepsilon^2)$, and that this problem is equivalent to testing whether $\mathbf{p}$ is ε-far from uniform in total variation (statistical) distance. Here, we consider this problem in the distributed setting where the samples are distributed over n users, and each user can send only ℓ bits to the central server. As in the Gaussian case, our goal is to characterize the sample complexity (number of users) required.

The rationale for considering this related problem stems from our next lemma, which shows that an upper bound for *binary product* mean testing implies an upper bound for the Gaussian testing problem.

Lemma 11 *Fix $\mu \in \mathbb{R}^d$. Let $X \sim \mathcal{G}(\mu, \mathbb{I}_d)$, and $Y \in \{-1, 1\}^d$ be defined by $Y_i = \operatorname{sgn}(X_i)$ for all $i \in [d]$. Then Y follows a product distribution with mean vector $\nu \in \mathbb{R}^d$, such that*

- *If* $\mu = \mathbf{0}_d$*, then* $\nu = \mathbf{0}_d$*;*
- *further,* $\|\nu\|_2 \geq \min\{\|\mu\|_2, \sqrt{d/2}\}/\sqrt{2}$.

(This reduction was first communicated by the authors of the present paper to the authors of (Canonne et al., 2019a), where it was used in the context of differentially private testing. We include a self-contained proof in Appendix A for completeness.) As a corollary, we obtain the following.

Corollary 12 *If there is a private- (resp. public-) coin protocol for distributed binary mean testing with* $n(d, \varepsilon)$ *users, then there exists a a private- (resp. public-) coin protocol for distributed Gaussian mean testing with* $n(d, \varepsilon/\sqrt{2})$ *users.*

Proof User j can convert their sample $\mathbf{X}_j$ into $\mathbf{Y}_j$ by taking term-wise signs, and then solve the binary distribution mean detection problem over the $\mathbf{Y}_j$'s. ■

Theorem 13 *There exists a private-coin protocol for distributed Gaussian mean testing with* $n = O\left(\frac{d^{3/2}}{\varepsilon^2 \ell}\right)$ *users.*

Proof The proof follows the idea of "distributed simulation" introduced in (Acharya et al., 2019c). Assume without loss of generality that ℓ divides d, and d/ℓ divides n. We partition the n users into groups of size d/ℓ where the ith group consists of users $\frac{d}{\ell} \cdot i + 1, \ldots, \frac{d}{\ell} \cdot i + \frac{d}{\ell}$ for $i \in \{0, 1, \ldots (\ell n/d - 1)\}$. For $1 \leq j \leq d/\ell$, the jth user within each group sends the signs of the coordinates $(j-1)\ell + 1, \ldots, j\ell$ of its sample. Since the samples are i.i.d., and the coordinates are independent within each sample, the d bits transmitted in total any given group of users are distributed as $\mathbf{Y} = \mathrm{sgn}(\mathbf{X})$. Therefore, we can *simulate* one sample of $\mathbf{Y}$ at the server from the messages of every group of d/ℓ users. Now, since there is a centralized procedure that uses $O(\sqrt{d}/\varepsilon^2)$ samples, with a total of $O(\sqrt{d}/\varepsilon^2 \cdot d/\ell)$ users, we can solve the distributed binary mean detection problem, and by Theorem 12 this implies the result. ■

3.1.2. UPPER BOUND OF $O(d/\varepsilon^4)$ USING NORMS OF SAMPLES

Our second algorithm is motivated by the following simple observation: when X is drawn from $\mathcal{G}(\mu, \mathbb{I}_d)$, we have $\mathbb{E}[\|X\|_2^2] = d + \|\mu\|_2^2$, which should allow us to separate $\mu = 0$ from $\|\mu\|_2 \geq \varepsilon$. We propose an algorithm that uses the ℓ_2 norm of the samples to achieve the aforementioned $O(d/\varepsilon^4)$ bound on the number of users. Specifically, we describe in Algorithm 1 an algorithm where each user sends a single bit (namely, $\ell = 1$) denoting whether the norm of their sample is larger than or smaller than a pre-specified suitable threshold. Analyzing this algorithm establishes the following theorem:

Theorem 14 *There is a private-coin protocol for distributed Gaussian mean testing with* $\ell = 1$ *bit of communication and* $n = O(d/\varepsilon^4)$ *users.*

Proof The proof of the theorem boils down to the following lemma, which shows that the chosen thresholding provides a good proxy for distinguishing the two cases.

Lemma 15 *For* $\varepsilon \in (0, 1)$*, if* $\|\mu\|_2 > \varepsilon$ *then* Y_i *in Algorithm 1 satisfies*

$$\Pr_{\mu}[\, Y_i = 1 \,] - \Pr_{0}[\, Y_i = 1 \,] = \Omega(\varepsilon^2/\sqrt{d}).$$

Algorithm 1: One-bit Gaussian mean testing without shared randomness

Input: n samples $\mathbf{X}_1, \ldots, \mathbf{X}_n \in \mathbb{R}^d$, $\mathbf{X}_j$ at user j

`/* The threshold may not be optimal (eventually losing constant factors), but it makes the analysis simpler. */`

User j sends $Y_j \leftarrow \mathbb{1}\left\{ \|\mathbf{X}_j\|_2^2 > d \right\}$.

if $\frac{1}{n}\sum_{j=1}^n Y_j < \frac{\Gamma(d/2,d/2)}{\Gamma(d/2)} + c\varepsilon^2/\sqrt{d}$ `// c > 0 is an absolute constant.`

then

  return accept

else

  return reject

Proof We note that, for each $j \in [n]$, $\|\mathbf{X}_j\|_2^2$ follows a chi-squared distribution with d degrees of freedom and non-centrality parameter $\|\mu\|_2^2$. Consider such a r.v. $Z \sim \chi^2_d(\|\mu\|_2^2)$, and denote by Q the Marcum Q-Function with parameters $d/2, \sqrt{d}$: $Q(x) = \text{Marcum}_{d/2}(x, \sqrt{d})$, so that

$$\Pr[\, Z > d\,] = Q(\|\mu\|_2)\,.$$

In particular, Q is increasing.

When $\|\mu\|_2 = 0$, we have $\Pr[\, Z > d\,] = Q(0) = \frac{\Gamma(d/2,d/2)}{\Gamma(d/2)} \asymp 1$. On the other hand, when $\|\mu\|_2 \geq \varepsilon$, $\Pr[\, Z > d\,] \geq Q(\varepsilon)$ and thus, for small ε using properties of the Marcum Q-Function, we get

$$\Pr[\, Z > d\,] = Q(0) + (1 + o(1))\frac{\varepsilon^2}{2}Q''(0) = Q(0) + \Theta(\varepsilon^2/\sqrt{d})\,,$$

establishing the lemma. ■

With this lemma in hand, the proof of correctness of the protocol is immediate, as a number of users (i.e., bits) quadratic in the inverse of the gap is sufficient to distinguish between the two cases. ■

To extend this result to the case where each user can send ℓ bits, we could consider dividing the coordinates into ℓ blocks, with each user then thresholds the norm of their sample within each block and sending the signs obtained. However, our analysis shows that the sample complexity of this approach is still $O(d/\varepsilon^4)$. In fact, we can show that *any* estimator based on degree-two moments will have sample complexity $\Omega(d/\varepsilon^4)$. Specifically, let $\mathbf{Z}_j = (\mathbf{X}_{j1}^2, \ldots, \mathbf{X}_{jd}^2)$ denote the vector whose entries are the squares of the entries of the samples $\mathbf{X}_j$'s. We prove that even with *full* access to $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ no algorithm can solve distributed Gaussian mean testing with fewer than $O(d/\varepsilon^4)$ users.

Theorem 16 *Given access only to $\mathbf{Z}_1, \ldots \mathbf{Z}_n$, the Gaussian mean testing problem cannot be solved unless $n = \Omega(d/\varepsilon^4)$, regardless of the value of ℓ.*

The proof is deferred to Appendix A. Again, note that the theorem above holds even in the general (centralized case), and shows that no scheme that only uses this information can achieve a sample complexity better than $\Omega(d/\varepsilon^4)$. This can be seen as one of the pieces of evidence that the d/ε^4 term in our upper bound is, in fact, optimal.

3.2. Public-Coin Upper Bound

Theorem 17 *There is a public-coin protocol for distributed Gaussian mean testing under ℓ-bit communication constraints with $O\left(\frac{d}{\varepsilon^2\sqrt{\ell}}\right)$ users.*

Proof The users use their shared randomness to jointly choose a rotation matrix R u.a.r.

Algorithm 2: One bit Gaussian mean testing with shared randomness

Input: n samples $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{R}^d$, $\mathbf{X}_i$ at user i, R a random rotation matrix available to all users

1 At user i, let $Y_{i,j} = \mathbb{I}\{(R \cdot \mathbf{X}_i)_j > 0\}$ for $j \in [\ell]$. Send $Y_i := (Y_{i,1}, \dots, Y_{i,\ell})$

2 Use these binary product distributions to test using the product distribution uniformity testing algorithm of Canonne et al. (2017).

By properties of Gaussian distributions, note that if $X \sim \mathcal{G}(\mu, \mathbb{I}_d)$, $RX \sim \mathcal{G}(R\mu, \mathbb{I}_d)$; further, $R\mu$ is a uniform distribution on the sphere of radius $\|\mu\|_2$.

Lemma 18 *Let $Z := (Z_1, \dots, Z_d) \sim \mathcal{G}(0, \mathbb{I}_d)$. Then the distribution of $R\mu$ (over the choice of R) is the same as the distribution of $(W_1, \dots, W_d) := \|\mu\|_2 \cdot \frac{(Z_1,\dots,Z_d)}{\|Z\|_2}$.*

This implies the following lemma.

Lemma 19 *If $\|\mu\|_2 \geq \varepsilon$, and $1 \leq \ell \leq d$, then $\Pr_R\left[\sum_{i=1}^{\ell} W_i^2 > \frac{\varepsilon^2\ell}{20d}\right] > \frac{1}{2}$.*

Proof With probability at least $3/4$, we have $\|Z\|_2^2 < 2d$ by Gaussian concentration. Further, since $\sum_{i=1}^{\ell} Z_i^2$ follows a χ^2_ℓ distribution, $\Pr\left[\sum_{i=1}^{\ell} Z_i^2 \geq \frac{\ell}{10}\right] \geq \frac{3}{4}$ for every $\ell \geq 1$. Recalling Lemma 18, this yields the result by a union bound over these two events. ■

Let $\pi_\ell \colon \mathbb{R}^d \to \mathbb{R}^\ell$ denote the projection on the first ℓ coordinates. For a rotation matrix $R \in \mathbb{R}^{d\times d}$, and a user $j \in [n]$, let $\mathbf{X}'_j := \pi_\ell(R\mathbf{X}_j) \in \mathbb{R}^\ell$ be the random variable obtained by projecting $R\mathbf{X}_j$ to its first ℓ coordinates. Then $\mathbf{X}'_1, \dots, \mathbf{X}'_n$ are i.i.d. r.v.'s distributed according to $\mathcal{G}(\pi_\ell(R\mu), \mathbb{I}_d)$. In particular, for every $j \in [n]$ and $i \in [\ell]$, we have

$$p_i(R) := \frac{1}{2}\operatorname{Erfc}\left(-\frac{1}{\sqrt{2}}(R\mu)_i\right) = \Pr_{\mathbf{X}}\left[\mathbf{X}'_{j,i} > 0\right] = \Pr[Y_{j,i} = 1]$$

and similar to the proof of Lemma 11, standard properties of Erfc ensure that, for every $i \in [\ell]$, $\left(p_i(R) - \frac{1}{2}\right)^2 \geq \frac{1}{8}\min\left((R\mu)_i^2, \frac{1}{2}\right)$. From the above and Lemma 19, we get that, with probability at least $1/2$ over the choice of R, $\sum_{i=1}^{\ell}\left(p_i(R) - \frac{1}{2}\right)^2 \geq \frac{1}{8}\min\left(\ell, \sum_{i=1}^{\ell}(R\mu)_i^2\right) \geq \frac{1}{160} \cdot \frac{\ell\varepsilon^2}{d}$.

Therefore, we have n i.i.d. samples $Y_1, \dots, Y_n$ from a product distribution $P(R)$ over $\{0,1\}^\ell$ with mean vector $p(R) \in [0,1]^\ell$, such that the following holds over the choice of R: (i) if $\mu = 0$, then with probability one $P(R)$ is the uniform distribution over $\{0,1\}^\ell$, i.e., $p(R) = \mathbf{u} := (1/2, \dots, 1/2)^\ell$; (ii) if $\|\mu\|_2 \geq \varepsilon$, then with probability at least $1/2$ we have $\|\mathbf{p} - \mathbf{u}\|_2^2 \geq \frac{1}{160} \cdot \frac{\ell\varepsilon^2}{d}$.

Invoking e.g., the algorithm of (Canonne et al., 2017, Theorem 4.1) (looking at the soundness guarantee in terms of the ℓ_2 norm between mean vectors, (Canonne et al., 2017, Claims 4.2 and 4.3)), this readily implies an upper bound of

$$O\left(\sqrt{\ell}/\left(\varepsilon\sqrt{\ell/d}\right)^2\right) = O\left(d/(\sqrt{\ell}\varepsilon^2)\right)$$

samples to distinguish between (i) and (ii) with probability at least $9/10$.

To summarize, over the choice of R and the $n = O(d/(\sqrt{\ell}\varepsilon^2))$ i.i.d. samples $\mathbf{X}_1, \dots, \mathbf{X}_n$, we get that (i) if $\mu = \mathbf{0}$, then the algorithm outputs accept with probability at least $9/10$; (ii) if $\|\mu\|_2 \geq \varepsilon$, then the algorithm outputs accept with probability at most $1/2 + 1/2 \cdot 1/10 = 11/20$. Finally, repeating a constant number of times independently (choosing a new R every time), thus paying only a constant factor in the number n of users required, allows us to obtain probability $9/10$ of being correct in both cases. ■

4. Lower Bounds

In view of the lower bound framework outlined in Section 2.1, in order to establish lower bounds on the sample complexity of distributed Gaussian mean testing, we are left with the following tasks. (1) Choose a convenient perturbed family $\mathcal{P}_\varepsilon \subseteq \mathcal{H}_\varepsilon$, i.e., a set $\mathcal{P}_\varepsilon = \{ \mathbf{p}_z : z \in \mathcal{Z} \}$ (along with a distribution ζ on $\mathcal{Z}$) such that, for all z, (i) $\mathbf{p}_z = \mathcal{G}(\mu_z, \mathbb{I}_d)$, and (ii) $\|\mu_z\|_2 \geq \varepsilon$.[3] (2) Fixing any channel $W \in \mathcal{W}_\ell$ (resp., $W^n \in \mathcal{W}_\ell^n$), upper bound $\chi^2(W \mid \mathcal{P}_\varepsilon)$ (resp., $\chi^{(2)}(W^n \mid \mathcal{P}_\varepsilon)$) as a function of d, n, ε, ℓ, before (3) invoking Lemma 9 or Lemma 10 to conclude.

Given this roadmap, two natural candidates for the perturbed family come to mind. The first consists of d Gaussians whose mean vector deviates from $\mathbf{0}$ in exactly one large "hidden coordinate:"

$$\mathcal{P}_\varepsilon^{\mathrm{hs}} := \{ \mathbf{p}_z = \mathcal{G}(\mu_z, \mathbb{I}_d) : \mu_z = \varepsilon \mathbf{e}_z, z \in [d] \} \tag{6}$$

(this corresponds to $\mathcal{Z} = [d]$, and the superscript stands for "hide-and-seek," name coined in Shamir (2014)). A natural choice for the distribution ζ is then the uniform distribution over $\mathcal{Z} = [d]$.

The second consists of 2^d Gaussians whose mean vector deviates from $\mathbf{0}$ by $\pm\varepsilon/\sqrt{d}$ in all coordinates:

$$\mathcal{P}_\varepsilon^{\mathrm{local}} := \left\{ \mathbf{p}_z = \mathcal{G}(\mu_z, \mathbb{I}_d) : \mu_z = \frac{\varepsilon}{\sqrt{d}} z, z \in \{-1,1\}^d \right\} \tag{7}$$

(this corresponds to $\mathcal{Z} = \{-1,1\}^d$, and the superscript stands for "locally perturbed"). A natural choice for the distribution ζ is then the uniform distribution over $\mathcal{Z} = \{-1,1\}^d$.[4]

4.1. Heuristic argument, and the challenges ahead.

For the sake of intuition, suppose for now that instead of trying to establish a lower bound on the task of distributed Gaussian mean testing, we were considering the related problem of *binary* mean testing, i.e., with product distributions over $\{-1,1\}^d$ instead of spherical Gaussians over $\mathbb{R}^d$. As seen in the upper bound section, this problem is tightly connected to the Gaussian one (as an algorithm for the binary case implies one for the Gaussian case). We can define the analogue of $\mathcal{P}_\varepsilon^{\mathrm{hs}}$ for the binary case, $\mathcal{P}_\varepsilon^{\mathrm{hs}'} := \{ \mathbf{p}_z \textit{ product on } \{-1,1\}^d : \mu_z = \varepsilon \mathbf{e}_z, z \in [d] \}$. The corresponding perturbation (with respect to the uniform distribution $\mathbf{p}$ on $\{-1,1\}^d$) can then easily be seen to be, for $z \in [d]$,

$$\forall x \in \{-1,1\}^d, \quad \delta_z(x) = (\mathbf{p}_z(x) - \mathbf{p}(x))/\mathbf{p}(x) = \prod_{i=1}^d (1 + \varepsilon x_i (\mu_z)_i) - 1 = \varepsilon x_z$$

3. Instead of (ii), a weaker condition would actually suffice, namely, that $\Pr_{Z\sim\zeta}[\|\mu_Z\|_2 \geq \varepsilon] \geq 9/10$. However, we will not require this in this section.

4. Using the remark from the previous footnote, one could also decide to instead define $\mathcal{P}_\varepsilon^{\mathrm{local}}$ for $z \in \mathbb{R}$, along with the prior ζ being itself a standard Gaussian. Although this simplifies some computations and complicates some others, the two choices are essentially equivalent.

and therefore, for any channel $W \in \mathcal{W}_\ell$, we get by unrolling the definition of δ_Z^W in Eq. (3) that

$$\begin{aligned}\chi^2\left(W \mid \mathcal{P}_\varepsilon^{\mathrm{hs}'}\right) &= \frac{\varepsilon^2}{d}\sum_{y\in\mathcal{Y}}\sum_{j=1}^{d}\frac{\mathbb{E}_{XX'}\left[X_jX'_jW(y \mid X)W(y \mid X')\right]}{\mathbb{E}_X[W(y \mid X)]} = \frac{\varepsilon^2}{d}\sum_{y\in\mathcal{Y}}\frac{\|\mathbb{E}_X[XW(y \mid X)]\|_2^2}{\mathbb{E}_X[W(y \mid X)]} \\ &\le \frac{4\varepsilon^2}{d}\sum_{y\in\mathcal{Y}}\mathbb{E}_X[W(y \mid X)]\log\frac{1}{\mathbb{E}_X[W(y \mid X)]} \qquad \text{(by Lemma 5)} \\ &\le \frac{4\varepsilon^2}{d}\log|\mathcal{Y}| = (4\log 2)\frac{\varepsilon^2\ell}{d}\,,\end{aligned}$$

which by Lemma 9 immediately implies a lower bound of $\Omega(d/(\ell\varepsilon^2))$ on the number of users required for (public-coin) distributed binary mean testing, respectively. This in particular allows us to very easily rederive the non-interactive version of the $\Omega(d/(\ell\varepsilon^2))$ lower bound of Shamir (Shamir, 2014, Theorem 2) on distributed (or memory-bounded) binary mean estimation.

However, this is considering the *binary* version of the problem, while we are interested in the *Gaussian* case. Mimicking the above argument, we get that, for said Gaussian setting, the perturbation with regard to the reference distribution $\mathbf{p}$ (which is the standard Gaussian $\mathcal{G}(\mathbf{0}, \mathbb{I}_d)$) obtained from any $\mathbf{p}_z \in \mathcal{P}_\varepsilon^{\mathrm{hs}}$ is given by

$$\forall x \in \mathbb{R}^d, \quad \delta_z(x) = (\mathbf{p}_z(x) - \mathbf{p}(x))/\mathbf{p}(x) = \frac{e^{-\|x-\mu_z\|_2^2/2}}{e^{-\|x\|_2^2/2}} - 1 = e^{-\frac{1}{2}\varepsilon^2}e^{\varepsilon x_z} - 1\,.$$

for all $z \in [d]$. Pursuing the same line of reasoning, we get that

$$\chi^2\left(W \mid \mathcal{P}_\varepsilon^{\mathrm{hs}}\right) = \mathbb{E}_Z\left[\left\|\delta_Z^W\right\|_2^2\right] = \frac{1}{d}\sum_{j=1}^{d}\sum_{y\in\mathcal{Y}}\frac{\mathbb{E}_X[(e^{-\frac{1}{2}\varepsilon^2}e^{\varepsilon X_j} - 1)W(y \mid X)]^2}{\mathbb{E}_X[W(y \mid X)]} \tag{8}$$

This does look promising: since $e^{-\frac{1}{2}\varepsilon^2}e^{\varepsilon X_j} - 1 \approx \varepsilon X_j$ for small ε, a heuristic argument would lead to $\chi^2\left(W \mid \mathcal{P}_\varepsilon^{\mathrm{hs}}\right) \approx \frac{\varepsilon^2}{d}\sum_{y\in\mathcal{Y}}\frac{\|\mathbb{E}_X[XW(y|X)]\|_2^2}{\mathbb{E}_X[W(y|X)]}$, from which we would recover the same bound as in the binary case. The issue, unfortunately, is that this heuristic argument relies on a first-order Taylor expansion of $e^{-\frac{1}{2}\varepsilon^2}e^{\varepsilon X_j} - 1$ *within the expectation*, even though X_j (a standard univariate Gaussian) is unbounded: however tempting this line of reasoning seems, one cannot take this route. Indeed, this non-linearity of the quantity to analyze is the crux of the difficulty. (We discuss some related conjectures in Section 5.)

4.2. Our results.

Our first lower bound applies to all public-coin protocols as long as ℓ is not too large (roughly, $\ell \ll \sqrt{d}$, and in particular is optimal for constant values of ℓ.

Lemma 20 (General Lower Bound) *For any channel $W \in \mathcal{W}_\ell$, and $\ell \le \sqrt{d}/\varepsilon^2$, we have that*

$$\chi^2\left(W \mid \mathcal{P}_\varepsilon^{\mathrm{local}}\right) = O\left(\max\left(\frac{\varepsilon^2\ell}{d}, \frac{\varepsilon^4\ell^2}{d}\right)\right).$$

Proof [Sketch] Fix any W. In view of bounding $\chi^2\left(W \mid \mathcal{P}_\varepsilon^{\mathrm{local}}\right) = \sum_{y\in\mathcal{Y}}\mathbb{E}_Z\left[\frac{\mathbb{E}_X[\delta_Z(X)W(y|X)]^2}{\mathbb{E}_X[W(y|X)]}\right]$, we start by noting that, since $\|z\|_2^2 = d$ for all $z \in \{-1,1\}^d$, δ_z is given by

$$\delta_z(x) = e^{-\frac{\varepsilon^2}{2}}e^{\frac{\varepsilon}{\sqrt{d}}\langle x,z\rangle} - 1, \qquad x \in \mathbb{R}^d\,.$$

"Since", then, for $d \gg 1$, $\varepsilon \ll 1$, and any fixed z, x, $\delta_z(x) \approx \frac{\varepsilon}{\sqrt{d}} \langle x, z \rangle$, simple manipulations yield

$$\chi^2\left(W \mid \mathcal{P}_\varepsilon^{\text{local}}\right) \approx \frac{\varepsilon^2}{d} \sum_{y \in \mathcal{Y}} \mathbb{E}_Z\left[\frac{\langle \mathbb{E}_X[XW(y \mid X)], Z\rangle^2}{\mathbb{E}_X[W(y \mid X)]}\right] = \frac{\varepsilon^2}{d} \sum_{y \in \mathcal{Y}} \left\| \frac{\mathbb{E}_X[XW(y \mid X)]}{\mathbb{E}_X[W(y \mid X)]} \right\|_2^2$$

which we can then upper bound by $O(\varepsilon^2 \ell / d)$ as in the previous discussion by invoking Lemma 5. Of course, the above intuition suffers the same issue – the linear approximation cannot be taken in the expectation, and handling this non-linearity is the technical crux of the proof, whose details are given in Appendix B. Roughly speaking, to handle this non-linearity, we instead expand the square $\mathbb{E}_X[\delta_Z(X) W(y \mid X)]^2$ to bring the outer expectation (over Z) inside, leaving us with a term $\mathbb{E}_{XX'}[\mathbb{E}_Z[\delta_Z(X)\delta_Z(X')] W(y \mid X) W(y \mid X')]$. To handle the inner $\mathbb{E}_Z[\delta_Z(X)\delta_Z(X')]$, we expand the product $\delta_Z(X)\delta_Z(X')$ (based on the expression of δ_Z), and compute their expectation with respect to Z to obtain three non-linear terms now only depending on X. The rest of the argument involves their (infinite) series expansion, and carefully obtained bounds reminiscent of the "level-k inequalities" of Lee (2019), to bound the expectation of the non-linear terms of this expansion. ■

By invoking Lemma 9, this implies Theorem 4.

Our next two results are lower bounds against *restricted* class of public-coin protocols, and in particular apply to the ones underlying our public-coin upper bounds (Theorem 2). In more detail, recall that our public-coin upper bound is achieved by (after applying a rotation chosen u.a.r. from the public randomness) the channel

$$\forall x \in \mathbb{R}^d, \forall y \in \{-1, 1\}^\ell, \quad W(y \mid x) = \prod_{i=1}^{\ell} \mathbb{1}\{y_i x_i \geq 0\} \in \{0, 1\} \tag{9}$$

i.e., the $W(y \mid \cdot)$'s partition the domain into 2^ℓ orthants corresponding to the first ℓ bits of the sample x. For this class of public-coin protocols, we can establish the following bounds on the two perturbed families we consider. For the first, we are able to obtain the following bound:

Lemma 21 (Restricted Lower Bound I) *For the channel $W \in \mathcal{W}_\ell$ given in (9), we have*

$$\chi^2\left(W \mid \mathcal{P}_\varepsilon^{\text{hs}}\right) = O\left(\frac{\varepsilon^2 \ell}{d}\right).$$

(Note that this bound is tight, as the distributed mean testing problem restricted to the hide-and-seek family $\mathcal{P}_\varepsilon^{\text{hs}}$ *can* be solved with $O(\frac{d}{\varepsilon^2 \ell})$ users.) The proof of this lower bound turns out to be resfreshingly simple, as we are able to leverage the product structure from (9).

Proof For this particular channel, we can easily and explicitly compute

$$\mathbb{E}_X\left[(e^{-\frac{1}{2}\varepsilon^2} e^{\varepsilon X_j} - 1) W(y \mid X)\right] = \begin{cases} \frac{1}{2^\ell} \text{Erf}\left(\frac{y_j \varepsilon}{\sqrt{2}}\right) & \text{if } j \leq \ell \\ 0 & \text{if } j > \ell \end{cases}$$

Plugging this into the definition of $\chi^2\left(W \mid \mathcal{P}_\varepsilon^{\text{hs}}\right)$ yields, recalling that for this W we have $\mathbb{E}_X[W(y \mid X)] = 1/2^\ell$ for all $y \in \mathcal{Y}$,

$$\chi^2\left(W \mid \mathcal{P}_\varepsilon^{\text{hs}}\right) = \frac{1}{d} \sum_{j=1}^{\ell} \sum_{y \in \mathcal{Y}} \frac{1}{2^\ell} \text{Erf}\left(\frac{y_j \varepsilon}{\sqrt{2}}\right) = \frac{1}{d} \sum_{j=1}^{\ell} \sum_{y \in \mathcal{Y}} \frac{\frac{2}{\pi}(y_j \varepsilon)^2 + o(\varepsilon^2)}{2^\ell} = \frac{2}{\pi}(1 + o(1)) \cdot \frac{\ell \varepsilon^2}{d}$$

which concludes the proof. ■

By considering the second family (the locally perturbed one), we can further show that our analysis of the protocol of Theorem 2 (which we conjecture to be optimal) cannot be improved upon, at least in the regime $\ell \ll d^{2/3}$. The proof is deferred to Appendix B in the interest of space.

Lemma 22 (Restricted Lower Bound II) *For the channel $W \in \mathcal{W}_\ell$ given in (9), we have that*

$$\chi^{(2)}\left(W^n \mid \mathcal{P}_\varepsilon^{\mathrm{local}}\right) = O\left(\max(n^2, n\ell) \cdot \frac{\varepsilon^4 \ell}{d^2}\right).$$

Indeed, by invoking Lemma 9, this readily implies that using the class of protocol satisfying Eq. (9), distributed Gaussian mean testing requires $n = \Omega(d/(\varepsilon^2\sqrt{\ell}), d^2/(\varepsilon^4\ell^2))$ – matching the bound from Theorem 2 as long as $\ell \leq d^{2/3}/\varepsilon^{4/3}$. While the two lower bounds above only apply to restricted classes of protocols, we see them as evidence for a general bound, a conjecture we discuss next.

5. Connections and Conjectures

Finally, in this section we provide evidence for our main conjecture (Theorem 3), and discuss some of the technical issues that stand in the way of proving it. Although the problem we consider here is inherently high-dimensional, each user receiving a full d-dimensional vector drawn from the unknown Gaussian, there appear to be striking similarities between this question and that of *univariate discrete goodness-of-fit*. Specifically, recent work (Acharya et al., 2018, 2019c) has shown that, in the same communication-constrained setting as ours, the task of testing whether an unknown distribution over a domain of size k is (i) uniform, or (ii) ε-far from uniform in total variation distance, requires $n = \Theta(k^{3/2}/(2^\ell\varepsilon^2))$ users for private-coin protocols, and $n = \Theta(k/(\sqrt{2^\ell}\varepsilon^2))$ for public-coin ones (further, this is also sufficient). This is a striking parallel to the task considered in this paper, for which we show upper bounds of respectively $n = O(d^{3/2}/(\ell\varepsilon^2))$ and $n = O(d/(\sqrt{\ell}\varepsilon^2))$; moreover, the lower bounds from (Acharya et al., 2018) follow the same framework as we use, with the crucial difference that the perturbations they consider are *linear* (while the non-linearity in our setting, as discussed earlier, is the key technical hurdle we face).

This could seem to be merely a coincidence, albeit an intriguing one, without the following additional observations. In both ours (Gaussian mean testing in ℓ_2) and theirs (univariate discrete goodness-of-fit in total variation), the underlying protocols work by estimating the ℓ_2 norm of a high-dimensional vector: either the mean of the Gaussian, or the probability vector of the probability mass function. Now, our private-coin upper bound does have an $O(d/\varepsilon^4)$ term (with one bit of communication), which the univariate setting does not appear to present; however, this can be explained by the fact that in our case, each user gets a d-dimensional sample, and there can actually compute an estimator of the ℓ_2 norm of the target vector. Given a single sample, in the univariate case, this is not possible (no unbiased estimator of the ℓ_2 norm of the density function exists); however, it is not hard to observe that allowing even *two* samples per user in the discrete univariate case enables a private-coin protocol with $n = O(k/\varepsilon^4)$ users, and also one bit of communication![5]

These similarities, along with our (limited) lower bounds (Lemmas 20, 21 and 22, and Theorem 16) and the seemingly intrinsic ε^4/d term which appears in the proof of our lower bound (see (17)) provide strong evidence in support of Theorem 3. However, in order to establish these lower bounds for the distributed Gaussian mean testing problem *via* our techniques, one would need an analogue of Lemma 5 for *non-linear* integrands, which we have not been able to obtain so far.

5. Namely, this protocol simply asks that each of the n users sends the bit indicating whether their two independent samples fall unto the same value of the domain $[k]$, i.e., if they observed a "collision."

References

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. Inference under information constraints I: lower bounds from chi-square contraction. *CoRR*, abs/1812.11476, 2018. In submission. Full version of Acharya et al. (2019b).

Jayadev Acharya, Clément L Canonne, Cody Freitag, and Himanshu Tyagi. Test without trust: Optimal locally private distribution testing. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 2067–2076. PMLR, 16–18 Apr 2019a. URL http://proceedings.mlr.press/v89/acharya19b.html.

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. Inference under information constraints: Lower bounds from chi-square contraction. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 3–17, Phoenix, USA, 25–28 Jun 2019b. PMLR. URL http://proceedings.mlr.press/v99/acharya19a.html.

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. Inference under information constraints II: communication constraints and shared randomness. *CoRR*, abs/1905.08302, 2019c. In submission. Full version of Acharya et al. (2019d).

Jayadev Acharya, Clément L Canonne, Canonne, and Himanshu Tyagi. Communication-constrained inference and the role of shared randomness. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 30–39, Long Beach, California, USA, 09–15 Jun 2019d. PMLR. URL http://proceedings.mlr.press/v97/acharya19a.html.

Rudolf Ahlswede and Imre Csiszár. Hypothesis testing with communication constraints. *IEEE Transactions on Information Theory*, 32(4):533–542, July 1986.

Alexandr Andoni, Tal Malkin, and Negev S. Nosatzki. Two party distribution testing: Communication and security. *ArXiV*, abs/1811.04065, November 2018.

Leighton Pate Barnes, Yanjun Han, and Ayfer Özgür. Learning distributions from their samples under communication constraints. *CoRR*, abs/1902.02890, 2019. URL http://arxiv.org/abs/1902.02890.

Mark Braverman, Ankit Garg, Tengyu Ma, Huy L. Nguyen, and David P. Woodruff. Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In *Symposium on Theory of Computing Conference, STOC'16*, pages 1011–1020. ACM, 2016.

T. Tony Cai and Hongji Wei. Distributed gaussian mean estimation under communication constraints: Optimal rates and communication-efficient algorithms. *CoRR*, abs/2001.08877, 2020.

Clément L. Canonne, Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. Testing bayesian networks. In Satyen Kale and Ohad Shamir, editors, *Proceedings of the 2017 Conference on*

Learning Theory, volume 65 of *Proceedings of Machine Learning Research*, pages 370–448, Amsterdam, Netherlands, 07–10 Jul 2017. PMLR. URL http://proceedings.mlr.press/v65/canonne17a.html.

Clément L. Canonne, Gautam Kamath, Audra McMillan, Jonathan Ullman, and Lydia Zakynthinou. Private identity testing for high-dimensional distributions. *CoRR*, abs/1905.11947, 2019a.

Clément L. Canonne, Xi Chen, Gautam Kamath, Amit Levi, and Erik Waingarten. Random restrictions of high-dimensional distributions and uniformity testing with subcube conditioning. *CoRR*, abs/1911.07357, 2019b.

Ilias Diakonikolas, Themis Gouleakis, Daniel M. Kane, and Sankeerth Rao. Communication and memory efficient testing of discrete distributions. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 1070–1106, Phoenix, USA, 25–28 Jun 2019. PMLR. URL http://proceedings.mlr.press/v99/diakonikolas19a.html.

Orr Fischer, Uri Meir, and Rotem Oshman. Distributed uniformity testing. In Calvin Newport and Idit Keidar, editors, *Proceedings of the 2018 ACM Symposium on Principles of Distributed Computing, PODC 2018, Egham, United Kingdom, July 23-27, 2018*, pages 455–464. ACM, 2018. URL https://dl.acm.org/citation.cfm?id=3212772.

Ankit Garg, Tengyu Ma, and Huy L. Nguyen. On communication cost of distributed statistical estimation and dimensionality. In *Advances in Neural Information Processing Systems 27*, pages 2726–2734, 2014.

Te Sun Han. Hypothesis testing with multiterminal data compression. *IEEE Transactions on Information Theory*, 33(6):759–772, November 1987.

Te Sun Han and Shun-Ichi Amari. Statistical inference under multiterminal data compression. *IEEE Transactions on Information Theory*, 44(6):2300–2324, October 1998.

Yanjun Han, Pritam Mukherjee, Ayfer Özgür, and Tsachy Weissman. Distributed statistical estimation of high-dimensional and nonparametric distributions with communication constraints, February 2018a. URL http://ita.ucsd.edu/workshop/18/files/abstract/abstract_2352.txt. Talk given at ITA 2018.

Yanjun Han, Ayfer Özgür, and Tsachy Weissman. Geometric lower bounds for distributed parameter estimation under communication constraints. In *Proceedings of the 31st Conference on Learning Theory, COLT 2018*, volume 75 of *Proceedings of Machine Learning Research*, pages 3163–3188. PMLR, 2018b.

Chin Ho Lee. Fourier bounds and pseudorandom generators for product tests. In *Computational Complexity Conference*, volume 137 of *LIPIcs*, pages 7:1–7:25. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2019.

Ohad Shamir. Fundamental limits of online and distributed algorithms for statistical learning and estimation. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 163–171. Curran Associates, Inc., 2014.

John N. Tsitsiklis. Decentralized detection. In H. V. Poor and J. B. Thomas, editors, *Advances in Statistical Signal Processing*, volume 2, pages 297–344. JAI Press, 1993.

Pramod K Varshney. *Distributed detection and data fusion*. Springer Science & Business Media, 2012.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018. ISBN 1108415199.

R. Viswanathan and P. Varshney. Distributed detection with multiple sensors: Part I – Fundamentals. *Proceedings of IEEE*, 85(1):54–63, January 1997.

Michele Wigger and Roy Timo. Testing against independence with multiple decision centers. *IEEE International Conference on Signal Processing and Communications, IISc, Bangalore*, June 2016.

Aolin Xu and Maxim Raginsky. Information-theoretic lower bounds on Bayes risk in decentralized estimation. *IEEE Transactions on Information Theory*, 63(3):1580–1600, 2017.

Yuchen Zhang, John Duchi, Michael I. Jordan, and Martin J. Wainwright. Information-theoretic lower bounds for distributed statistical estimation with communication constraints. In *Advances in Neural Information Processing Systems 26*, pages 2328–2336, 2013.

Appendix A. Omitted Proofs: Upper Bounds

In this appendix, we provide the missing proofs from Section 3.

A.1. Proof of Theorem 16

Theorem 23 (Theorem 16, restated) *Given access only to $\mathbf{Z}_1, \dots \mathbf{Z}_n$, the Gaussian mean testing problem cannot be solved unless $n = \Omega(d/\varepsilon^4)$, regardless of the value of ℓ.*

Proof Let p be the distribution $\mathcal{G}(0, \mathbb{I}_d)$, and $q = \mathcal{G}(\mu, \mathbb{I}_d)$, where $\mu = (\varepsilon/\sqrt{d}, \varepsilon/\sqrt{d}, \dots, \varepsilon/\sqrt{d})$. When $\mathbf{X}_j \sim p$, then the $\mathbf{Z}_{j,i}$'s are independent standard chi-square random variables with density (for $z > 0$)

$$P(z) = \frac{1}{\sqrt{2}\Gamma(1/2)} e^{-z/2} \frac{1}{\sqrt{z}},$$

and when $\mathbf{X}_j \sim q$, then $\mathbf{Z}_{j,i}$'s are independent non-central chi-squared distributions with non-centrality $\lambda := \varepsilon^2/d$, and therefore have a pdf given by

$$Q(z) = \frac{1}{2} e^{-(z+\lambda)/2} \left(\frac{z}{\lambda}\right)^{-1/4} I_{-1/2}(\sqrt{\lambda z}),$$

where I_ν is the modified Bessel function of the first kind given by $I_\nu(y) = (y/2)^\nu \sum_{j=0}^\infty \frac{(y^2/4)^j}{j!\Gamma(\nu+j+1)}$. Let $g(y) := \sum_{j=0}^\infty \frac{(y^2/4)^j}{j!\Gamma(j+1/2)}$, then $I_\nu(y) = (y/2)^\nu g(y)$. Substituting $y = \sqrt{\lambda z}$, we obtain

$$\begin{aligned} Q(z) &= \frac{1}{2} e^{-(z+\lambda)/2} \left(\frac{z}{\lambda}\right)^{-1/4} \left(\frac{2}{\sqrt{\lambda z}}\right)^{1/2} g(\sqrt{\lambda z}) \\ &= \frac{1}{\sqrt{2}} e^{-z/2} e^{-\lambda/2} \frac{1}{\sqrt{z}} g(\sqrt{\lambda z}) \\ &= P(z) \cdot e^{-\lambda/2} g(\sqrt{\lambda z}) \Gamma(1/2). \end{aligned}$$

By the properties of Gamma function, note that $\Gamma(j+1/2) \le \Gamma(1/2) \cdot j!/2$, and $\Gamma(j+1/2) > \Gamma(1/2)(j-1)!/2$, and therefore, we can upper and lower bound the probabilities as follows:

$$P(z) \cdot e^{-\lambda/2} \left(1 + \frac{\lambda z}{2} + \sum_{j=2}^\infty \frac{2(\lambda z/4)^j}{j!j!}\right) \le Q(z) \le P(z) \cdot e^{-\lambda/2} \left(1 + \frac{\lambda z}{2} + \sum_{j=2}^\infty \frac{2(\lambda z/4)^j}{j!(j-1)!}\right).$$

Therefore, the Bhattacharyya parameter $B(P, Q)$ between Q and P can be bounded as

$$\int_0^\infty dz \sqrt{P(z)Q(z)} \ge \int_0^\infty dz\, P(z) \left(e^{-\lambda/4} \left(1 + \frac{\lambda z}{2} + \frac{\lambda^2 z^2}{32}\right)\right)^{1/2}. \tag{10}$$

Recalling that $\mathbb{E}_p[\mathbf{Z}_{j,i}] = 1$, we obtain $\int_0^\infty dz \sqrt{P(z)Q(z)} = (1 - \Theta(\lambda^2))$. By the multiplicativity of Bhattacharyya parameter, over the n samples, the distribution of $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ under p and q is then

$$B(p(\mathbf{Z}_1, \dots, \mathbf{Z}_n), q(\mathbf{Z}_1, \dots, \mathbf{Z}_n)) = (1 - \Theta(\lambda^2))^{nd}. \tag{11}$$

For the two distributions to be distinguishable, we want this parameter to be bounded away from 1 (i.e., the Hellinger distance bounded away from 0), which to happen requires $nd = \Omega(\lambda^2)$; recalling that $\lambda = \varepsilon^2/d$ yields the result. ■

A.2. Proof of Lemma 11

Lemma 24 (Lemma 11, restated) *Fix $\mu \in \mathbb{R}^d$. Let $X \sim \mathcal{G}(\mu, \mathbb{I}_d)$, and $Y \in \{-1,1\}^d$ be defined by $Y_i = \operatorname{sgn}(X_i)$ for all $i \in [d]$. Then Y follows a product distribution with mean vector $\nu \in \mathbb{R}^d$, such that*

- *If $\mu = \mathbf{0}_d$, then $\nu = \mathbf{0}_d$;*
- *further, $\|\nu\|_2 \geq \min\{\|\mu\|_2, \sqrt{d/2}\}/\sqrt{2}$.*

Proof The first part of the statement is immediate, so it suffices to prove the second. Suppose $\|\mu\|_2 > 0$; then, for every $i \in [d]$,

$$\mathbb{E}[Y_i] = 2\Pr[\,Y_i = 1\,] - 1 = 2\Pr[\,X_i > 0\,] - 1 = 2\operatorname{Erfc}\left(-\frac{1}{\sqrt{2}}\mu_i\right) - 1$$

and standard results on Erfc ensure that, for every $i \in [\ell]$, $\mathbb{E}[Y_i]^2 \geq \frac{1}{2}\min(\mu_i^2, \frac{1}{2})$. From this, we get

$$\|\nu\|_2^2 = \sum_{i=1}^{d} \mathbb{E}[Y_i]^2 \geq \frac{1}{2}\min\left(\frac{d}{2}, \|\mu\|_2^2\right). \tag{12}$$

concluding the proof. ■

Appendix B. Omitted Proofs: Lower Bounds

In this appendix, we provide the missing proofs from Section 4.

B.1. Proof of Lemma 20

We first restate the lemma, before providing its proof.

Lemma 25 (Lemma 20, restated) *For any channel $W \in \mathcal{W}_\ell$, and $\ell \leq \sqrt{d}/\varepsilon^2$, we have that*

$$\chi^2\left(W \mid \mathcal{P}_\varepsilon^{\mathrm{local}}\right) = O\left(\max\left(\frac{\varepsilon^2\ell}{d}, \frac{\varepsilon^4\ell^2}{d}\right)\right).$$

Proof Before delving into the proof, we can assume without loss of generality $\mathbb{E}_X[W(y \mid X)] \leq 1/4$ for all $y \in \mathcal{Y} = \{-1,1\}^\ell$. This will be useful at the end of the proof; to see why we can make this assumption, note that by allowing $\ell' = \ell + 2$ bits of communication instead of ℓ (which does not change the asymptotics of our bounds), we can convert any protocol into one satisfies this assumption. (Indeed, it suffices for each player to rerandomize uniformly any given message y of the original protocol among four copies in the new protocol. Each message in the new protocol has therefore probability at most $1/4$ to be sent.)

In view of bounding $\chi^2\left(W \mid \mathcal{P}_\varepsilon^{\mathrm{local}}\right)$ in this setting, we observe that, for the perturbed family $\mathcal{P}_\varepsilon^{\mathrm{local}}$ defined in (7), and since the reference distribution $\mathbf{p}$ is the standard Gaussian $\mathcal{G}(\mathbf{0}, \mathbb{I}_d)$ we have for every $z \in \mathbb{R}^d$

$$\forall x \in \mathbb{R}^d, \qquad \delta_z(x) = \frac{e^{-\|x-\mu_z\|_2^2/2}}{e^{-\|x\|_2^2/2}} - 1 = e^{-\frac{\varepsilon^2\|z\|_2^2}{2d}} e^{\frac{\varepsilon}{\sqrt{d}}\langle x, z\rangle} - 1\,.$$

For convenience of some of the computations, we will deviate a little from the definition of $\mathcal{P}_\varepsilon^{\mathrm{local}}$, and actually choose the distribution of Z, ζ, to be $\mathcal{G}(\mathbf{0}, \mathbb{I}_d)$ instead of uniform on $\{-1,1\}^d$. As

mentioned earlier, this does not change the result, but will simplify some of the (already cumbersome) arguments.[6] Further, we hereafter will write $\gamma := \varepsilon/\sqrt{d}$ to alleviate notation. Now, we get that, for all $x, x' \in \mathbb{R}^d$,

$$\mathbb{E}_Z\big[\delta_Z(x)\delta_Z(x')\big] = e^{-\varepsilon^2}\mathbb{E}_Z\Big[e^{\gamma\langle x+x', Z\rangle}\Big] - e^{-\frac{\varepsilon^2}{2}}\mathbb{E}_Z\Big[e^{\gamma\langle x, Z\rangle} + e^{\gamma\langle x', Z\rangle}\Big] + 1\,. \tag{13}$$

Now, since ζ is a standard Gaussian, we can use the fact that for any $v \in \mathbb{R}^d$,

$$\mathbb{E}_{U\sim\mathcal{G}(0,1)}\Big[e^{\alpha x U - \beta\alpha^2 U^2}\Big] = \frac{e^{\frac{\alpha^2 x^2}{2(1+2\beta\alpha^2)}}}{\sqrt{1+2\beta\alpha^2}} \tag{14}$$

in order to rewrite this as

$$\begin{aligned}
1+\mathbb{E}_Z\big[\delta_Z(x)\delta_Z(x')\big] &= \frac{e^{\frac{\gamma^2}{2(1+2\gamma^2)}\|x+x'\|_2^2}}{(1+2\gamma^2)^{d/2}} - \left(\frac{e^{\frac{\gamma^2}{2(1+\gamma^2)}\|x\|_2^2}}{(1+\gamma^2)^{d/2}} + \frac{e^{\frac{\gamma^2}{2(1+\gamma^2)}\|x'\|_2^2}}{(1+\gamma^2)^{d/2}}\right) \\
&= \sum_{k=0}^{\infty}\frac{\varepsilon^{2k}}{2^k k!}\left(\frac{\|x+x'\|_2^2}{d(1+2\gamma^2)} - \frac{\ln(1+2\gamma^2)}{\gamma^2}\right)^k - \sum_{k=0}^{\infty}\frac{\varepsilon^{2k}}{2^k k!}\left(\left(\frac{\|x\|_2^2}{d(1+\gamma^2)} - \frac{\ln(1+\gamma^2)}{\gamma^2}\right)^k + \left(\frac{\|x'\|_2^2}{d(1+\gamma^2)} - \frac{\ln(1+\gamma^2)}{\gamma^2}\right)^k\right).
\end{aligned}$$

From here, it follows that, for any fixed $y \in \mathcal{Y}$, and writing $a(x) := W(y \mid X)$ for convenience,

$$\begin{aligned}
\mathbb{E}_{XX'}\big[\mathbb{E}_Z\big[\delta_Z(X)\delta_Z(X')\big]a(X)a(X')\big] = \sum_{k=1}^{\infty}\frac{\varepsilon^{2k}}{2^k k!}\mathbb{E}_{XX'}\left[\left(\frac{\|X+X'\|_2^2}{d(1+2\gamma^2)} - \frac{\ln(1+2\gamma^2)}{\gamma^2}\right)^k a(X)a(X')\right] \\
- 2\mathbb{E}_X[a(X)]\sum_{k=1}^{\infty}\frac{\varepsilon^{2k}}{2^k k!}\mathbb{E}_X\left[\left(\frac{\|X\|_2^2}{d(1+\gamma^2)} - \frac{\ln(1+\gamma^2)}{\gamma^2}\right)^k a(X)\right]
\end{aligned}$$

In order to handle this expression, we consider separately the first (linear) term, before bounding the remaining non-linear ones.

- Upon expanding $\|X+X'\|_2^2 = \|X\|_2^2 + \|X'\|_2^2 + 2\langle X, X'\rangle$, the first term ($k=1$) is equal to

$$\begin{aligned}
\Delta_1 &= \frac{\varepsilon^2}{2}\cdot\mathbb{E}_{XX'}\left[\left(\frac{2\langle X,X'\rangle}{d(1+2\gamma^2)} + \frac{\|X\|_2^2+\|X'\|_2^2}{d}\left(\frac{1}{1+2\gamma^2} - \frac{1}{1+\gamma^2}\right) + \frac{2\ln(1+\gamma^2)-\ln(1+2\gamma^2)}{\gamma^2}\right)a(X)a(X')\right] \\
&= \frac{\varepsilon^2}{d(1+2\gamma^2)}\mathbb{E}_{XX'}\big[\langle X, X'\rangle a(X)a(X')\big] + \frac{\varepsilon^2}{2}\mathbb{E}_X[a(X)]^2\frac{2\ln(1+\gamma^2)-\ln(1+2\gamma^2)}{\gamma^2} + \frac{\varepsilon^2}{2}\Delta(X,X') \\
&\le \frac{\varepsilon^2}{d}\|\mathbb{E}_X[Xa(X)]\|_2^2 + \frac{\varepsilon^4}{2d}\mathbb{E}_X[a(X)]^2 + \frac{\varepsilon^2}{2}\Delta(X,X') \\
&\le \frac{\varepsilon^2}{d}\|\mathbb{E}_X[Xa(X)]\|_2^2 + \frac{\varepsilon^4}{2d}\mathbb{E}_X[a(X)]^2
\end{aligned} \tag{15}$$

  where $\Delta(X,X') := \frac{\|X\|_2^2+\|X'\|_2^2}{d}\left(\frac{1}{1+2\gamma^2} - \frac{1}{1+\gamma^2}\right) \le 0$, and the inequality stems from the fact that $2\ln(1+\gamma^2) - \ln(1+2\gamma^2) \le \gamma^4$. This leaves us with two terms, the second of which will be easy to take care of, and the first ($\|\mathbb{E}_X[Xa(X)]\|_2^2$) which can handle with Lemma 5.

- On the other hand, we can deal with the terms of degree $k \ge 2$ by considering the function $f_k\colon \mathbb{R}^{2d} \to \mathbb{R}$ given by

$$f_k(x,x') := \left(\frac{\|x+x'\|_2^2}{d(1+2\gamma^2)} - \frac{\ln(1+2\gamma^2)}{\gamma^2}\right)^k - \left(\frac{\|x\|_2^2}{d(1+\gamma^2)} - \frac{\ln(1+\gamma^2)}{\gamma^2}\right)^k - \left(\frac{\|x'\|_2^2}{d(1+\gamma^2)} - \frac{\ln(1+\gamma^2)}{\gamma^2}\right)^k$$

6. Specifically, we lose the convenient fact that $\|Z\|_2^2 = d$ a.s. for Z uniform on $\{-1,1\}^d$; however, we now have a closed-form expression for some of the expectations, instead of an unwieldy product of $\cosh$.

which has degree at most $2k$, and $b\colon \mathbb{R}^{2d} \to [0,1]$ given by $b(x,x') := a(x)a(x')$, thus getting, for some absolute constant $\beta > 0$,

$$\begin{aligned}\mathbb{E}_{XX'}\left[\left|f_k(X,X')\right| a(X)a(X')\right] &\le \beta^k \|f_k\|_2 \cdot \mathbb{E}_{XX'}\left[b(X,X')\right] \ln^k \frac{e}{\mathbb{E}_{XX'}[b(X,X')]^{1/2k}} \\ &= \beta^k \|f_k\|_2 \cdot \mathbb{E}_X[a(X)]^2 \ln^k \frac{e}{\mathbb{E}_X[a(X)]^{1/k}}\end{aligned}$$

from an application of a straightforward generalization of level-k-type inequalities (Lee, 2019, Lemma 10). Putting it together, this leads to

$$\begin{aligned}&\mathbb{E}_{XX'}\left[\mathbb{E}_Z\left[\delta_Z(X)\delta_Z(X')\right]a(X)a(X')\right] \\ &\le \tfrac{\varepsilon^2}{d}\|\mathbb{E}_X[Xa(X)]\|_2^2 + \tfrac{\varepsilon^2}{d}\mathbb{E}_X[a(X)]^2 + \mathbb{E}_X[a(X)]^2 \sum_{k=2}^{\infty} \tfrac{\beta^k \varepsilon^{2k}}{2^k k!}\|f_k\|_2 \ln^k \tfrac{e}{\mathbb{E}_X[a(X)]^{1/k}} \quad (16)\end{aligned}$$

To proceed, of course, it remains to give a bound on $\|f_k\|_2$, which the following lemma (whose proof is deferred to the end of this appendix) does.

Lemma 26 *For $k \ge 2$ and f_k defined as above, we have $\|f_k\|_2^2 \lesssim \frac{(2C\cdot k)^{2k}}{d^k}$, where $C > 0$ is an absolute constant independent of k.*

Therefore, absorbing said absolute constant C in the other absolute constant β, (16) then yields

$$\begin{aligned}&\mathbb{E}_{XX'}\left[\mathbb{E}_Z\left[\delta_Z(X)\delta_Z(X')\right]a(X)a(X')\right] \\ &\quad\le \frac{\varepsilon^2}{d}\|\mathbb{E}_X[Xa(X)]\|_2^2 + \mathbb{E}_X[a(X)]^2 \sum_{k=2}^{\infty} \frac{2^k\beta^k\varepsilon^{2k}}{d^{k/2}k!}\left(k^k + \ln^k \frac{1}{\mathbb{E}_X[a(X)]}\right)\end{aligned}$$

Note that the series $\sum_{k=2}^{\infty} \frac{2^k\beta^k\varepsilon^{2k}k^k}{d^{k/2}k!}$ converges (for d sufficiently large with respect to some constant related to β), with

$$\sum_{k=2}^{\infty} \frac{2^k\beta^k\varepsilon^{2k}k^k}{d^{k/2}k!} = O\left(\frac{\varepsilon^4}{d}\right).$$

In addition, we have

$$\sum_{k=2}^{\infty} \frac{2^k\beta^k\varepsilon^{2k}}{d^{k/2}k!}\left(\ln\frac{1}{\mathbb{E}_X[a(X)]}\right)^k = e^{2\beta\frac{\varepsilon^2}{\sqrt{d}}\ln(1/\mathbb{E}[a(X)])} - 2\beta\frac{\varepsilon^2}{\sqrt{d}}\ln\frac{1}{\mathbb{E}[a(X)]} - 1$$

which "should" behave as $\frac{\varepsilon^4}{d}\ln^2\frac{1}{\mathbb{E}[a(X)]}$ (although we cannot immediately conclude this). For convenience, let $\phi(x) = e^x - x - 1$ for $x \in \mathbb{R}$, so that the above is equal to $\phi(2\beta\frac{\varepsilon^2}{\sqrt{d}}\ln(1/\mathbb{E}[a(X)]))$.

To summarize, so far we have arrived to

$$\begin{aligned}&\chi^2\left(W \mid \mathcal{P}_\varepsilon^{\mathrm{local}}\right) \\ &\le \tfrac{\varepsilon^2}{d}\sum_{y\in\mathcal{Y}} \tfrac{\|\mathbb{E}_X[XW(y|X)]\|_2^2}{\mathbb{E}_X[W(y|X)]} + \sum_{y\in\mathcal{Y}} \mathbb{E}_X[W(y\mid X)]\cdot O\left(\tfrac{\varepsilon^4}{d}\right) + \sum_{y\in\mathcal{Y}} \mathbb{E}_X[W(y\mid X)]\cdot\phi\left(2\beta\tfrac{\varepsilon^2}{\sqrt{d}}\ln\tfrac{1}{\mathbb{E}_X[W(y|X)]}\right) \\ &\lesssim \frac{\varepsilon^2\ell}{d} + \frac{\varepsilon^4}{d} + \sum_{y\in\mathcal{Y}} \mathbb{E}_X[W(y\mid X)]\cdot\phi\left(2\beta\frac{\varepsilon^2}{\sqrt{d}}\ln\frac{1}{\mathbb{E}_X[W(y\mid X)]}\right) \quad (17)\end{aligned}$$

where the first term follows from Lemma 5 and $\sum_{y\in\mathcal{Y}} \mathbb{E}_X[W(y \mid X)] \log \frac{1}{\mathbb{E}_X[W(y|X)]} \leq \ell$. To conclude, we turn to the last term; and let $\tau := 2\beta\frac{\varepsilon^2}{\sqrt{d}} \ll 1$, so that

$$\begin{aligned}
&\mathbb{E}_X[W(y \mid X)]\phi\Big(2\beta\tfrac{\varepsilon^2}{\sqrt{d}} \ln \tfrac{1}{\mathbb{E}_X[W(y|X)]}\Big) \\
&\quad = \mathbb{E}_X[W(y \mid X)]\phi(\tau \ln \tfrac{1}{\mathbb{E}_X[W(y|X)]}) \\
&\quad = \mathbb{E}_X[W(y \mid X)]^{1-\tau} - \tau\mathbb{E}_X[W(y \mid X)] \ln \frac{1}{\mathbb{E}_X[W(y \mid X)]} - \mathbb{E}_X[W(y \mid X)]
\end{aligned}$$

To conclude, we will finally use our assumption that $\mathbb{E}_X[W(y \mid X)] \leq 1/4$ for all $y \in \mathcal{Y}$. Then, since for all $\tau \leq \tau_0$, where $\tau_0 > 0$ is absolute constant, the function $x \mapsto x^{1-\tau} - \tau x \ln \frac{1}{x} - x$ is concave on $(0, 1/4)$, we can bound

$$\sum_{y\in\mathcal{Y}} \mathbb{E}_X[W(y \mid X)] \cdot \phi\left(\tau \ln \frac{1}{\mathbb{E}_X[W(y \mid X)]}\right) \leq |\mathcal{Y}|^\tau - \tau \ln |\mathcal{Y}| - 1 = O\left(\frac{\varepsilon^4\ell^2}{d}\right)$$

the last equality as $\ell \lesssim \sqrt{d}/\varepsilon^2$. Plugging this in Eq. (17), we finally obtain

$$\chi^2\left(W \mid \mathcal{P}_\varepsilon\right) = O\left(\frac{\varepsilon^2\ell}{d} + \frac{\varepsilon^4\ell^2}{d}\right)$$

as claimed. ■

We now provide the last piece, the proof of Lemma 26.

Proof [Proof of Lemma 26]

$$\begin{aligned}
\|f_k\|_2^2 &= \mathbb{E}_{XX'}\left[f_k(X, X')^2\right] \\
&\leq 5\mathbb{E}_{XX'}\left[\left(\frac{\|X + X'\|_2^2}{d(1+2\gamma^2)} - \frac{\ln(1+2\gamma^2)}{\gamma^2}\right)^{2k} + \left(\frac{\|X\|_2^2}{d(1+\gamma^2)} - \frac{\ln(1+\gamma^2)}{\gamma^2}\right)^{2k} + \left(\frac{\|X'\|_2^2}{d(1+\gamma^2)} - \frac{\ln(1+\gamma^2)}{\gamma^2}\right)^{2k}\right] \\
&= 5 \cdot 2^{2k}\mathbb{E}_{XX'}\left[\left(\frac{1}{d(1+2\gamma^2)}\left\|\frac{X + X'}{\sqrt{2}}\right\|_2^2 - \frac{\ln(1+2\gamma^2)}{2\gamma^2}\right)^{2k}\right] + 10\mathbb{E}_X\left[\left(\frac{\|X\|_2^2}{d(1+\gamma^2)} - \frac{\ln(1+\gamma^2)}{\gamma^2}\right)^{2k}\right] \\
&= \frac{5 \cdot 2^{2k}}{(1+2\gamma^2)^{2k}}\mathbb{E}_{XX'}\left[\left(\frac{1}{d}\|X\|_2^2 - \frac{(1+2\gamma^2)\ln(1+2\gamma^2)}{2\gamma^2}\right)^{2k}\right] + \frac{10}{(1+\gamma^2)^{2k}}\mathbb{E}_X\left[\left(\frac{1}{d}\|X\|_2^2 - \frac{(1+\gamma^2)\ln(1+\gamma^2)}{\gamma^2}\right)^{2k}\right] \\
&\leq 5 \cdot 2^{2k}\mathbb{E}_{XX'}\left[\left(\frac{1}{d}\|X\|_2^2 - \frac{(1+2\gamma^2)\ln(1+2\gamma^2)}{2\gamma^2}\right)^{2k}\right] + 10\mathbb{E}_X\left[\left(\frac{1}{d}\|X\|_2^2 - \frac{(1+\gamma^2)\ln(1+\gamma^2)}{\gamma^2}\right)^{2k}\right] \\
&\leq 10 \cdot 2^{4k}\mathbb{E}_X\left[\left(\frac{1}{d}\|X\|_2^2 - 1\right)^{2k}\right] + 5 \cdot 2^{4k}\left(\frac{(1+2\gamma^2)\ln(1+2\gamma^2)}{2\gamma^2} - 1\right)^{2k} + 10 \cdot 2^{2k}\left(\frac{(1+\gamma^2)\ln(1+\gamma^2)}{\gamma^2} - 1\right)^{2k} \\
&\leq 10 \cdot 2^{4k}\mathbb{E}_X\left[\left(\frac{1}{d}\|X\|_2^2 - 1\right)^{2k}\right] + 15 \cdot 2^{4k}\gamma^{4k} \\
&\leq \frac{10 \cdot 2^{2k}}{d^{2k}} \cdot (C \cdot k\sqrt{d})^{2k} + 15 \cdot (2\varepsilon^2/d)^{2k} \lesssim \frac{(2C \cdot k)^{2k}}{d^k}
\end{aligned}$$

for some absolute constant $C > 0$, using the fact that $\|X\|_2^2 \sim \chi_d^2$ is a subexponential random variable (see, e.g., (Vershynin, 2018, Section 2) (and in particular Proposition 2.7.1 and Exercise 2.7.10). ■

B.2. Proof of Lemma 22

We now prove Lemma 22, restated below.

Lemma 27 (Lemma 22, restated) *For the channel $W \in \mathcal{W}_\ell$ given in (9), we have that*

$$\chi^{(2)}\left(W^n \mid \mathcal{P}_\varepsilon^{\mathrm{local}}\right) = O\left(\max(n^2, n\ell) \cdot \frac{\varepsilon^4 \ell}{d^2}\right).$$

Proof Our objective is now to bound (4), for which we start by analyzing (part of) (5): since, for a Rademacher perturbation Z, $\|Z\|_2^2 = d$ a.s., we get

$$\begin{aligned}
\mathbb{E}_X[\delta_Z(X)W(y \mid X)] &= \mathbb{E}_X\left[(e^{-\frac{\varepsilon^2}{2}} e^{\gamma\langle X, Z\rangle} - 1)W(y \mid X)\right] \\
&= -\mathbb{E}_X[W(y \mid X)] + e^{-\frac{\varepsilon^2}{2}} \prod_{i=1}^{\ell} \mathbb{E}_X\left[e^{\gamma X_i Z_i} \mathbb{1}\{y_i X_i \geq 0\}\right] \cdot \prod_{i=\ell+1}^{d} \mathbb{E}_X\left[e^{\gamma X_i Z_i}\right] \\
&= -\mathbb{E}_X[W(y \mid X)] + e^{-\frac{\varepsilon^2}{2}} \cdot \prod_{i=1}^{\ell} \frac{e^{\frac{\gamma^2}{2}}}{2}\left(1 + \mathrm{Erf}\left(\frac{\gamma y_i Z_i}{\sqrt{2}}\right)\right) \cdot e^{(d-\ell)\frac{\gamma^2}{2}} \\
&= \mathbb{E}_X[W(y \mid X)]\left(\prod_{i=1}^{\ell}\left(1 + \mathrm{Erf}\left(\frac{\gamma y_i Z_i}{\sqrt{2}}\right)\right) - 1\right).
\end{aligned}$$

again, since $\mathbb{E}_X[W(y \mid X)] = 2^{-\ell}$ for all $y \in \mathcal{Y}$, and $d\gamma^2 = \varepsilon^2$. Using this expression in (4) gives us

$$\left\langle \delta_Z^W, \delta_{Z'}^W \right\rangle = \sum_{y \in \mathcal{Y}} \mathbb{E}_X[W(y \mid X)]\left(\prod_{i=1}^{\ell}\left(1 + \mathrm{Erf}\left(\frac{\gamma y_i Z_i}{\sqrt{2}}\right)\right) - 1\right)\left(\prod_{i=1}^{\ell}\left(1 + \mathrm{Erf}\left(\frac{\gamma y_i Z'_i}{\sqrt{2}}\right)\right) - 1\right)$$

This does look a tad unwieldy, but we can expand this and use our assumption that $\mathbb{E}_X[W(y \mid X)] = 1/2^\ell$ to get the following, where we take the expectation over a uniformly random $y \in \{-1, 1\}^\ell$ and use the fact that Erf is an odd function. In what follows, we set $\gamma' := \mathrm{Erf}(\gamma/\sqrt{2})$ to avoid cluttering the equations: since Erf is odd, $\mathrm{Erf}(\alpha y_i Z_i) = \mathrm{Erf}(\alpha) y_i Z_i$, so that

$$\begin{aligned}
\left\langle \delta_Z^W, \delta_{Z'}^W \right\rangle &= \mathbb{E}_y\left[\prod_{i=1}^{\ell}\left(1 + \gamma' y_i Z_i\right)\left(1 + \gamma' y_i Z'_i\right) - \prod_{i=1}^{\ell}\left(1 + \gamma' y_i Z_i\right) - \prod_{i=1}^{\ell}\left(1 + \gamma' y_i Z'_i\right) + 1\right] \\
&= \prod_{i=1}^{\ell} \mathbb{E}_{y_i}\left[\left(1 + \gamma' y_i Z_i\right)\left(1 + \gamma' y_i Z'_i\right)\right] - 1 = \prod_{i=1}^{\ell}\left(1 + \gamma'^2 Z_i Z'_i\right) - 1 \\
&\leq e^{\gamma'^2 \sum_{i=1}^{\ell} Z_i Z'_i} - 1 \;= \frac{2}{\pi}\gamma^2 \sum_{i=1}^{\ell} Z_i Z'_i + O(\gamma^4 \ell^2),
\end{aligned}$$

recalling that $\left|\sum_{i=1}^{\ell} Z_i Z'_i\right| \leq \ell$, and $\gamma^2 \ell \leq \varepsilon^2 \ll 1$. From the above, the n-fold induced decoupled chi-square fluctuation of $\mathcal{P}_\varepsilon^{\mathrm{local}}$ for our choice of $W^n \in \mathcal{W}^n$ is

$$\begin{aligned}
\chi^{(2)}\left(W^n \mid \mathcal{P}_\varepsilon^{\mathrm{local}}\right) &= \log \mathbb{E}_{ZZ'}\left[\exp\left(\frac{2}{\pi} n\gamma^2 \sum_{i=1}^{\ell} Z'_i Z_i\right)\right] + O\big(n\gamma^4 \ell^2\big) \\
&\leq \frac{2\ell n^2 \gamma^4}{\pi^2} + O\big(n\gamma^4 \ell^2\big) = O\big(\ell n^2 \gamma^4 + n\gamma^4 \ell^2\big) = O\big(\ell n^2 \gamma^4\big).
\end{aligned}$$

This concludes the proof. ∎

Appendix C. Omitted Proofs: Miscellaneous

We here provide a simple self-contained proof of our measure change bound, restated below:

Lemma 28 (Theorem 5, restated) *For $X \sim \mathcal{G}(\mathbf{0}, \mathbb{I}_d)$ and any function $a(X)$,*

$$\frac{\|\mathbb{E}[a(X)X]\|_2^2}{\mathbb{E}[a(X)]^2} \leq 2\frac{\mathbb{E}_P[a(X)\log a(X)]}{\mathbb{E}_P[a(X)]} + 2\log\frac{1}{\mathbb{E}_P[a(X)]}.$$

In particular, for $a(X) = \mathbb{1}\{X \in A\}$, $\frac{\|\mathbb{E}[a(X)X]\|_2^2}{\mathbb{E}[a(X)]^2} \leq 2\log\frac{1}{P(A)}$.

Proof By Gibb's variational principle (which can be proved simply by nonnegativity of KL divergence), for every random variable Z with distribution P and $Q \ll P$

$$\lambda\mathbb{E}_Q[Z] \leq \log\mathbb{E}_P\left[e^{\lambda Z}\right] + D(Q\|P).$$

Setting $Z = X_i$ and $Q \ll P$ as

$$\frac{dQ}{dP} = \frac{a(X)}{\mathbb{E}_P[a(X)]},$$

we get

$$\lambda\mathbb{E}_Q[X_i] \leq \log\mathbb{E}_P\left[e^{\lambda X_i}\right] + D(Q_i\|P_i) = \frac{\lambda^2}{2} + D(Q_i\|P_i),$$

where we denote by Q_i and P_i the marginals of the ith coordinates of X under Q and P, respectively. In particular, for $\lambda = \sqrt{2D(Q_i\|P_i)}$, we get

$$\mathbb{E}_Q[X_i] \leq \sqrt{2D(Q_i\|P_i)},$$

whereby

$$\mathbb{E}_Q[X_i]^2 \leq 2D(Q_i\|P_i).$$

and upon summing both sides over i,

$$\|\mathbb{E}_Q[X]\|_2^2 \leq 2\sum_{i=1}^d D(Q_i\|P_i)\,.$$

Note that the sum on the right-side above is less than

$$\sum_{i=1}^d \mathbb{E}_{X^{i-1}}\left[D\big(Q_{X_i|X^{i-1}}\|P_{X_i}\big)\right] = D(Q\|P) = \frac{\mathbb{E}_P[a(X)\log a(X)]}{\mathbb{E}_P[a(X)]} + \log\frac{1}{\mathbb{E}_P[a(X)]},$$

which completes the proof since $\mathbb{E}_Q[X] = \frac{\mathbb{E}_P[a(X)X]}{\mathbb{E}_P[a(X)]}$. ■