Optimal high-dimensional and nonparametric distributed testing under communication constraints

Szabó, Botond; Vuursteen, Lasse; Zanten, Harry Van

DOI
10.1214/23-AOS2269

Publication date
2023

Document Version
Final published version

Published in
Annals of Statistics

The Annals of Statistics
2023, Vol. 51, No. 3, 909–934
https://doi.org/10.1214/23-AOS2269

OPTIMAL HIGH-DIMENSIONAL AND NONPARAMETRIC DISTRIBUTED TESTING UNDER COMMUNICATION CONSTRAINTS

BY BOTOND SZABÓ[1,a], LASSE VUURSTEEN[2,b] AND HARRY VAN ZANTEN[3,c]

[1]*Department of Decision Sciences, Bocconi Institute for Data Science and Analytics (BIDSA), Bocconi University,* [a]*botond.szabo@unibocconi.it*

[2]*Delft Institute of Applied Mathematics (DIAM), Delft University of Technology,* [b]*l.vuursteen@tudelft.nl*

[3]*Department of Mathematics, Vrije Universiteit Amsterdam,* [c]*j.h.van.zanten@vu.nl*

We derive minimax testing errors in a distributed framework where the data is split over multiple machines and their communication to a central machine is limited to b bits. We investigate both the d- and infinite-dimensional signal detection problem under Gaussian white noise. We also derive distributed testing algorithms reaching the theoretical lower bounds.

Our results show that distributed testing is subject to fundamentally different phenomena that are not observed in distributed estimation. Among our findings we show that testing protocols that have access to shared randomness can perform strictly better in some regimes than those that do not. We also observe that consistent nonparametric distributed testing is always possible, even with as little as *one* bit of communication, and the corresponding test outperforms the best local test using only the information available at a single local machine. Furthermore, we also derive adaptive nonparametric distributed testing strategies and the corresponding theoretical lower bounds.

1. Introduction. Distributed methods are concerned with inference in a framework where the data resides at multiple machines. Such settings occur naturally, when data is observed and processed locally, at multiple locations before sent to a central location where they are aggregated to obtain a final result. By working with smaller sample sizes locally, distributed methods can substantially speed up the computation compared to centralized, classical methods. Furthermore, they reduce memory requirements and help protecting privacy by not storing all the information at a single location. For these reasons the study of distributed methods has attracted significant attention in recent years.

In our analysis we first consider the many normal means model, which is often used as a platform to investigate more complex statistical problems. In the classical version of the model, one obtains an observation X subject to the dynamics $X = f + n^{-1/2}Z$, where $f \in \mathbb{R}^d$ is an unknown signal, and Z an unobserved, d-dimensional standard normal noise vector. This is equivalent to observing n independent copies of a $N_d(f, I_d)$ vector. Our focus is on testing the absence or presence of the signal component f in the model. Rejecting the null hypothesis $H_0 : f = 0$ means declaring that there is a nonzero signal underlying the observation X. The difficulty of distinguishing between the two hypotheses depends on signal strength, the noise ratio n and dimension d. It is well known that the signal strength in terms of the Euclidean norm of f needs to be at least of the order $d^{1/4}/\sqrt{n}$ for the hypotheses to be distinguishable; see, for example, [6].

We study this signal detection problem in a distributed setting. In the distributed version of the above normal-means model, the n observations are divided over m machines (assuming without loss of generality that n is a multiple of m). Equivalently, each local machine $j \in$

Received August 2022; revised December 2022.
MSC2020 subject classifications. Primary 62G10, 62F30; secondary 62F03.
Key words and phrases. Distributed methods, nonparametric, hypothesis testing, minimax optimal.

$\{1, \dots, m\}$ observes

$$X^j = f + \sqrt{\frac{m}{n}} Z^j, \tag{1}$$

where $f \in \mathbb{R}^d$ and the noise vectors Z^j are independent d-dimensional standard normal random vectors. Each machine j transmits a b-bit transcript Y^j to a central machine. By aggregating these m local transcripts, the central machine computes a test for the hypothesis $H_0 : f = 0$. We derive, for this distributed setting, the order of the minimal signal strength ρ for which the null hypothesis can be distinguished from the alternative $H_1 : \|f\|_2 \geq \rho$. In the distributed setting, ρ is considered as a function of the number of machines m and the communication budget b, in addition to the dimension d and noise level n. We allow all the parameters b, m and d to depend on n.

The transcripts generated by the machines may be either deterministic or randomized. When randomizing the transcript, we consider two different possibilities for the source of randomness. In the *private coin* setup, the machines may only use their own local (independent) source of randomness. In the *public coin* setup, the machines have access to a shared source of randomness in addition to their own independent source. This is akin to a situation in which the machines have access to the same random seed. We show that, depending on the size of the communication budget, having access to a public coin strictly improves the distinguishability of the null and alternative hypotheses.

Our results indicate that, in the case where b and m are small relative to the dimension d in an appropriate sense, the one-bit protocols have similar properties, in terms of separation rate, as multibits protocols; that is, one can achieve the minimax optimal b-bit testing rates with taking the majority vote of appropriately chosen local (one-bit) test outcomes. This is a striking difference with estimation, where for small values of b, increases in the communication budget result in (sometimes exponential) improvements in convergence rate. We find that, as m increases, the local testing problems become more difficult as the local sample size deceases, but at a certain threshold, this effect is compensated for by the increase in total communication budget bm. This threshold occurs when bm exceeds the dimension. At this point we find that public coin protocols start to strictly outperform private coin protocols, in the sense that smaller signals can be detected with the same amount of transmitted bits b. This is also a dissimilarity with estimation, where having access to public randomness offers no benefit, as we show it in our paper. When the communication budget b per machine exceeds that of the dimension d of the problem, the minimax rates of the classical, nondistributed setting can be attained.

We then extend our results for the d-dimensional Gaussian model to the nonparametric signal in white noise setting. This latter model is of interest, as it serves as benchmark and starting point to investigate more complicated nonparametric models. Here the local observations for $j = 1, \dots, m$ constitute $\int_0^{\cdot} f(s)\, ds + \sqrt{\frac{m}{n}} W^j_{\cdot}$, where the W^j's are independent Brownian motions and $f \in L_2[0, 1]$ the unknown functional parameter of interest. Our results for the infinite dimensional model comes in the form of minimax rates for distributed protocols in terms of the strength of the signal in L_2-norm, the smoothness s of the signal, the amount of bits b allowed to be communicated by each machine, the signal to noise ratio n and the number of machines m. In contrast to nonparametric distributed estimation, we show that consistent distributed testing is always possible, even when m and b are small. Having a shared source of randomness results in better rates in certain regimes in the nonparametric setting, while we show that this is never the case for distributed estimation. Finally, we consider the more realistic, adaptive setting where the regularity s is considered to be unknown. We show that, in contrast to the nondistributed setting where the cost for adaptation is a multiplicative $\log\log n$ factor, in the distributed case a more severe $\log n$ penalty is necessary. We also propose a nonparametric distributed testing procedure, based on Bonferroni's correction, reaching the theoretical limits (up to a $\log\log n$ factor) and observe additional, unexpected phase transitions compared to the nonadaptive setting.

1.1. *Related literature.* Starting a few decades ago, earlier investigations into similar topics originate in the electrical engineering community under the names "decentralized decision theory/the CEO problem," for example, [4, 9, 20, 31–33] or "inference under multiterminal compression" (see [17] for an overview). Motivated by applications, such as surveillance systems and wireless communication, the inference problems are approached from a "rate-distortion" angle in this body of literature. However, these results typically consider fixed, finite sample spaces and a fixed number of machines m and investigate asymptotics only in the sample size n.

Understanding the fundamental statistical performance of distributed methods, in context of nondiscrete, higher-dimensional sample spaces, has been considered only recently. Most of the literature focused on estimating the parameter/signal of the model in a distributed framework. Minimax lower and (up to a possible logarithmic factor) matching upper bounds were derived for the minimax risk in terms of communication constraints and in context of the many normal means and simple parametric problems; see [11–13, 16, 18, 24, 35, 37]. These results were extended to nonparametric models, including Gaussian white noise [38], nonparametric regression [26], density estimation [7] and general, abstract settings [30]. Distributed techniques for adapting to the unknown regularity of the functional parameter of interest were derived in [14, 26, 27].

For distributed testing much less is known. In [1] the authors consider a setting in which each machine obtains a single observation from a distribution on a finite sample space and derive lower bounds for testing uniformity of this distribution. Similar distributed uniformity testing is considered in [2], where matching upper bounds are exhibited for this setting. In [28] the authors derive matching upper and lower bounds for the distributed version of the classical many normal means model (see (1) above) for the case that only the outcome of local tests can be communicated (e.g., *one* bit of communication). In [3] less stringent communication requirements are considered in the special case of the model in (1) above with $m = n$. Questions regarding nonparametric models and adaptation in the setting of distributed testing have remained completely open thus far.

To summarize the state of the art, the lower bounds derived in the literature so far are only optimal in case of constant communication budget in the public coin setting, that is, $b = O(1)$. So far no lower bound results are available in the public coin setting if b can tend to infinity as n increases. Furthermore, there is a lack of any lower bound result in the private coin setup. The traditional methods, based on mutual information and Taylor expansion as considered in [28] and [3], respectively, do not extend to the setting of multiple bits or private coin protocols. In this article we fill this gap and derive the first rigorous minimax lower bounds for distributed testing procedures in the normal means model for arbitrary communication budget b both for private and public coin settings. In order to prove the lower bounds, we provide a novel Bayesian testing argument based on a Brascamp–Lieb-type inequality with distributed version of testing lower bounding techniques.

The upper bounds derived in [3] are more complete for both the private and public coin settings and go beyond the above described restrictive setting in which the lower bounds were derived but do not cover all possible cases. For instance, in [3] it is assumed that the separation distance between the null and alternative hypotheses is bounded from above by one, which does not cover the case $\sqrt{dm} \gg n$. Also, only the $m = n$ case was considered in the preceding paper. Therefore, in certain regimes new testing procedures and proof techniques had to be derived for full treatment of the problem (e.g., our novel test T_{III} in the high-budget private coin case, see Section 4.3).

The literature on distributed testing has so far solely focused on finite dimensional models. We provide the first results for distributed testing in nonparametric models. Besides deriving lower and matching upper bounds, we also derive an adaptive testing procedure, not depending on the typically unknown regularity of the underlying functional parameter of interest.

1.2. *Overview of our results and organization.* For a quick overview, the main contributions of this article are:

- Sharp minimax upper and lower bounds for all values of n, m, d, b for the d-dimensional distributed-signal-in-white-noise model, for both private and public coin settings (Section 3), with accompanying methods achieving these rates (Theorem 3.1 and Theorem 3.2).
- We extend the d-dimensional distributed-signal-in-white-noise model to the nonparametric setting where the signal is a Sobolev regular functional parameter of known regularity and establish the minimax rates within this setting for all values of n, m, b for both the private and public coin settings (Theorem 6.1).
- We consider the nonparametric setting in which the regularity of underlying signal is unknown and derive adaptive private and public coin procedures. Furthermore, we establish private and public coin lower bounds for the adaptive setting that are tight up to a $\log\log n$ factor for all values of n, m, b (Theorem 7.1 and Theorem 7.2).

The remainder of the paper is organized as follows. In Section 2 we describe the distributed-signal-in-white-noise model with d-dimensional signal $f \in \mathbb{R}^d$ and formalize the distributed testing problem both for private and public coin protocols. In Section 3 we provide the minimax lower and matching upper bounds for both testing protocols. We exhibit constructive algorithms that achieve matching upper bounds in Section 4. Section 5 gives a sketch of the proof of the lower bound. We extend our results to the nonparametric distributed-signal-in-white-noise model with Sobolev regular functional parameter in Section 6. Here we also compare distributed testing and estimation rates and highlight the similarities and differences between them both in the private and public coin settings. In Section 7 we consider adaptation to the unknown regularity level in the nonparametric setting and present theoretical lower and matching upper bounds. In Section 8 we derive constructive algorithms achieving these upper bounds. The detailed proof of the lower bound for the d-dimensional signal is deferred to Section 9, and a key technical lemma is described in Section 10. Detailed proofs for this lemma as well as some of the technical details of the other main results and various auxilliary results have been deferred to the Supplementary Material [29] to this manuscript. Results, equations and sections in the Supplementary Material are indexed by capital letters, as opposed to numerals that are used in the article.

1.3. *Notation.* We write $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$. For two positive sequences a_n, b_n, we use the notation $a_n \lesssim b_n$ if there exists a universal positive constant C such that $a_n \leq C b_n$. We write $a_n \asymp b_n$, which holds if $a_n \lesssim b_n$ and $b_n \lesssim a_n$ are satisfied simultaneously. We shall use $a_n \gg b_n$ to denote $b_n/a_n \to 0$. The Euclidean norm of a vector $v \in \mathbb{R}^d$ is denoted by $\|\cdot\|_2$. For absolutely continuous probability measures $P \ll Q$, we denote by $D_{\mathrm{KL}}(P\|Q) = \int \log \frac{dP}{dQ}\, dP$ and $D_{\chi^2}(P\|Q) = \int (\frac{dP}{dQ} - 1)^2\, dP$ their Kullback–Leibler and Chi-square divergences, respectively. Throughout the whole text, for convenience we use the abbreviations RHS and LHS for right-hand side and left-hand side, respectively, and CDF for the cumulative distribution function.

2. Problem formulation and setting. We consider testing in the distributed-signal-in-white-noise model. In this section we provide the formulation of the distributed setup for data coming from the finite dimensional model. Except for obvious modifications to the sample space, the same setup is considered when the local data is from the infinite dimensional distributed-signal-in-white-noise model, which is formulated in Section 6. For $j = 1, \dots, m$ machines, the local observations constitute X^j taking values in $\mathcal{X} \subset \mathbb{R}^d$, subject to dynamics (1) under P_f. Each machine j communicates a b-bit transcript Y^j to a central machine. That is, the transcript Y^j takes values in some space $\mathcal{Y}^j$ with $|\mathcal{Y}^j| \leq 2^b$ for $b \in \mathbb{N}$. Let $Y = (Y^1, \dots, Y^m)$ denote the aggregated data in the central machine. The central machine computes a test $T(Y)$, where T is a map from $\mathcal{Y} := \bigotimes_{j=1}^m \mathcal{Y}^j$ to $\{0, 1\}$ that has to distinguish between the null hypothesis $f = 0$ and the alternative hypothesis. As an alternative hypothesis, we consider whether

$$f \in H_\rho := \{ f \in \mathbb{R}^d : \|f\|_2 \geq \rho \}$$

for some appropriately chosen $\rho = \rho_{m,n,d,b}$.

We distinguish two mechanisms through which the local machines $j = 1, \dots, m$ can generate their transcripts Y^j. In the first setup, machines can use only their local observation X^j when generating Y^j, possibly in combination with a local source of randomness. In the second setup, we allow the machines to access a common source of randomness U, which is independent of the data $X := (X^1, \dots, X^m)$. In the latter setup, which we call the public coin setting, the machines may use both local randomness, the observed draw of U and their local observation X^j when generating their transcript Y^j. The setup where only local randomness is available shall be referred to as the private coin setting. A formal definition of these two setups is as follows:

- A *private coin distributed testing protocol* consists of a map $T : \mathcal{Y} \to \{0, 1\}$ and a collection of Markov kernels $K^j : 2^{\mathcal{Y}^j} \times \mathcal{X}^j \to [0, 1]$, $j = 1, \dots, m$, and the transcript satisfies $Y^j | X^j \sim K^j(\cdot | X^j)$.
- A *public coin distributed testing protocol* consists of a map $T : \mathcal{Y} \to \{0, 1\}$, a random variable U taking values in a probability space $(\mathcal{U}, \mathscr{U}, \mathbb{P}^U)$ and a collection of Markov kernels $K^j : 2^{\mathcal{Y}^j} \times \mathcal{X}^j \times \mathcal{U} \to [0, 1]$, $j = 1, \dots, m$ such that $Y^j | (X^j, U) \sim K^j(\cdot | X^j, U)$.

The choices for the kernels induce the conditional distribution of $Y = (Y^1, \dots, Y^m)$, which we will denote $K := \bigotimes_{j=1}^m K^j$. For the joint distribution of X, Y and U, we shall write $\mathbb{P}_{f,K} \equiv \mathbb{P}_f$, where the f subscript indicates the dynamics underlying X and the subscript K is used to stress that the conditional distribution of Y induced by the choice of kernels. Furthermore, we denote by $\mathbb{P}_f^X$ the corresponding marginal distribution of X, that is, $\mathbb{P}_f^X = P_f$. Our distributed architecture in the public coin case then follows the following Markov chain structure at each local machine $j = 1, \dots, m$:

$$\begin{array}{ccccc} & & U & \searrow & \\ & & & & Y^j. \\ f & \longrightarrow & X^j & \nearrow & \end{array} \tag{2}$$

Note that any private coin testing protocol can effectively be considered a public coin testing protocol for which U has degenerate distribution, that is, $U = u \in \mathcal{U}$ almost surely. In our proofs below and for the sake of compactness, we consider without loss of generality that the private coin setting implies U has a degenerate distribution. When no confusion can arise, we will refer to a distributed testing protocol as "distributed test," and we will refer to the tuple $(T, \{K^1, \dots, K^m\}, \mathbb{P}^U)$ by T for ease of notation. We use $\mathcal{T}_{\text{priv}}(b)$ and $\mathcal{T}_{\text{pub}}(b)$ to denote the classes of all private and public coin distributed tests, respectively, each with communication budget b per machine.

We define the testing risk of a distributed test $T \equiv (T, K, \mathbb{P}^U)$ for the alternative hypothesis H_ρ as the sum of the Type I and Type II errors, that is,

$$\mathcal{R}(H_\rho, T) := \mathbb{P}_0(T(Y) = 1) + \sup_{f \in H_\rho} \mathbb{P}_f(T(Y) = 0). \tag{3}$$

3. Minimax upper and lower bounds in the normal means model. Our main results come in the form of two theorems. The first establishes the lower bounds for the detection threshold for both the public- and private coin distributed tests. We provide the proof of this theorem in Section 9. The second theorem establishes the optimality of the lower bound, posed in the first theorem, by providing distributed tests in both the public and private coin cases which attain the respective rates posed by the lower bounds. These optimal distributed testing procedures are described in Section 4. We note that our results are not asymptotic in nature, as they hold for every combination of b, n, m and d, hence going beyond the classical parametric framework.

THEOREM 3.1. [*Distributed testing lower bound*] *For each* $\alpha \in (0, 1)$, *there exists a constant* $c_\alpha > 0$ (*depending only on* α) *such that if*

$$\rho^2 < c_\alpha \frac{\sqrt{d}}{n}\left(\sqrt{\frac{d}{b \wedge d}} \wedge \sqrt{m}\right), \tag{4}$$

then in the public coin protocol case

$$\inf_{T \in \mathcal{T}_{\text{pub}}(b)} \mathcal{R}(H_\rho, T) > \alpha \quad \textit{for all } n, m, d, b \in \mathbb{N}.$$

Similarly, for

$$\rho^2 < c_\alpha \frac{\sqrt{d}}{n}\left(\frac{d}{b \wedge d} \wedge \sqrt{m}\right), \tag{5}$$

we have under the private coin protocol that

$$\inf_{T \in \mathcal{T}_{\text{priv}}(b)} \mathcal{R}(H_\rho, T) > \alpha \quad \textit{for all } n, m, d, b \in \mathbb{N}.$$

The approach to proving the lower bound theorem can be summarized as follows. We start out by lower bounding the testing risk by a type of Bayes risk, where the parameter f is drawn from an adversarial prior distribution π. By taking π to be Gaussian, we can exploit the conjugacy of the model in order to show that optimal transcripts are either invariant to the prior or "Gaussian" in an appropriate sense. After this the results follow by data-processing arguments that are geometric in nature. We defer a more elaborate sketch of the proof to Section 5 and the detailed proof to Section 9. The techniques used in this work are novel and drastically different than those used in [3, 28], which provide tight bounds only in the *one*-bit case.

The above theorem implies that if (4) holds, no consistent public coin distributed testing protocol with communication budget b bits per machine exists for the hypotheses $H_0 : f = 0$ vs. the alternative $H_1 : \|f\|_2 \geq \rho$. In other words, no public coin distributed test manages to consistently distinguish all signals from 0 if the signals are smaller than the RHS of (4). When considering only private coin distributed testing protocols, the detection threshold (5) is more stringent than the public coin threshold (4) for certain values of d, m and b. Theorem 3.2 below affirms that, in these cases, the best private coin protocol have a strictly worse performance compared to the best public coin protocol.

THEOREM 3.2. *For each* $\alpha \in (0, 1)$, *there exists a constant* $C_\alpha > 0$ (*depending only on* α) *such that if*

$$\rho^2 \geq C_\alpha \frac{\sqrt{d}}{n}\left(\sqrt{\frac{d}{b \wedge d}} \wedge \sqrt{m}\right),$$

there exists $T \in \mathcal{T}_{\text{pub}}(b)$ *such that*

$$\mathcal{R}(H_\rho, T) \le \alpha \quad \textit{for all } n, m, d, b \in \mathbb{N}.$$

Similarly, for

$$\rho^2 \ge C_\alpha \frac{\sqrt{d}}{n}\left(\frac{d}{b \wedge d} \wedge \sqrt{m}\right)$$

there exists $T \in \mathcal{T}_{\text{priv}}(b)$ *such that*

$$\mathcal{R}(H_\rho, T) \le \alpha \quad \textit{for all } n, m, d, b \in \mathbb{N}.$$

The achievability of arbitrarily small testing risk is shown using a constructive proof; see Section 4. That is, we derive distributed testing protocols that distinguish the null hypothesis from any $f \in \mathbb{R}^d$ in the alternative class.

Theorem 3.1 together with Theorem 3.2 establish the minimax distributed testing rate for public and private coin protocols. As a sanity check, note that when $m = 1$, we obtain the nondistributed minimax testing rate $\rho^2 = \sqrt{d}/n$. Furthermore, when $b \gtrsim d$, enough information about the coefficients can be communicated to obtain the nondistributed minimax rate also, for both the public coin and private coin distributed protocols. When the communication budget is smaller than the dimension ($b = o(d)$), the class of public coin protocols starts to exhibit strictly better performance than the private coin ones in scenarios as long as $d = o(mb)$. That is, as long as the total communication budget mb of the system exceeds the dimension d of the parameter, public coin protocols achieve a strictly better rate than private coin ones. This remarkable phenomenon disappears when the dimension is larger than the total communication budget (i.e., $mb = o(d)$), at which point there exists a one-bit private coin protocol achieving the optimal rate of $\rho^2 \asymp \frac{\sqrt{md}}{n}$ in both cases. Consistent distributed testing turns out to be possible, even for small values of b and m, as long as n is large enough compared to d. This stands in contrast to estimation in the d-dimensional Gaussian mean model, where consistent estimation is not possible when $mb = o(d)$, regardless of sample size n (see, e.g., [12]). Furthermore, as long as $mb = o(d)$ in the public coin case or $mb^2 = o(d^2)$ in the private coin case, an increase in communication budget does not lead to a better rate. This stands in stark contrast to estimation, where for small budgets an increase can lead to an exponential improvement in convergence rate.

4. Distributed testing protocols achieving the lower bound in the many normal means model. In this section we exhibit three distributed testing procedures achieving the rates posed by the lower bound. The first distributed testing procedure T_{I} communicates only a single bit per machine and can detect signals with a squared Euclidean norm of larger or equal order than $\frac{\sqrt{dm}}{n}$ and does not need a public coin. As a second procedure, we consider a test satisfying the public coin protocol T_{II} that achieves the rate $\frac{d}{n\sqrt{b \wedge d}}$. The third procedure satisfies the private coin protocol and achieves the corresponding slower rate $\frac{d}{n(b \wedge d)}$. Note that, depending on the values of n, m, d and b, the existence of such distributed testing protocols proves Theorem 3.2 and implies that the lower bounds in Theorem 3.1 are, in fact, tight.

A common denominator in the construction of the three protocols is that the transcripts Y^j are generated as vector of p_f^j-Bernoulli random variables taking values in $\{0, 1\}^b$, where $p_f^j \in [0, 1]^b$ depends on the underlying signal f, with $p_f^j = (1/2, \dots, 1/2)$ under the null hypothesis ($f = 0$). The concentration inequality for groups of Bernoulli random variables, given in Lemma 4.1, provides a recipe for the construction of a central test for each of the three regimes. The Type I error can be controlled since the distribution under the null hypothesis is known. The Type II error is small whenever the vectors of probabilities $p_f^1, \dots, p_f^m$ are sufficiently separated from $(1/2, \dots, 1/2)$ in Euclidean norm.

LEMMA 4.1. *Consider for $k, l \in \mathbb{N}$, $l \geq 2$, independent random variables $\{B_i^j : i = 1, \dots, k, j = 1, \dots, l\}$ with $B_i^j \sim \mathrm{Ber}(p_i)$. If $p_i = 1/2$ for $i = 1, \dots, k$, it holds that, for all $\alpha \in (0, 1)$, there exists $\kappa_\alpha > 0$ such that*

$$T := \mathbb{1}\left\{\left|\frac{1}{\sqrt{k}l}\sum_{i=1}^{k}\left(\sum_{j=1}^{l}\left(B_i^j - \frac{1}{2}\right)\right)^2 - \sqrt{k}/4\right| \geq \kappa_\alpha\right\}$$

satisfies $\mathbb{E}T \leq \alpha/2$. On the other hand, if

$$\eta_{p,l,k} := \frac{l-1}{2\sqrt{k}}\sum_{i=1}^{k}\left(p_i - \frac{1}{2}\right)^2 \geq \kappa_\alpha, \tag{6}$$

it holds that

$$\mathbb{E}(1 - T) \leq \frac{1/2 + 16\eta_{p,l,k}/\sqrt{k}}{\eta_{p,l,k}^2}. \tag{7}$$

The proof of the lemma can be found in Section A.2 of the Supplementary Material where it is restated as Lemma A.4. We also provide a version of this lemma (Lemma A.5 in the Supplementary Material) used in the high-budget private coin protocol case.

4.1. *Low communication budget*: *Construction of T_{I}*. The protocol presented here is similar to the one given in [28], with some adjustment allowing the application of Lemma 4.1 for a simpler proof.

We first compute the local test statistic $S_{\mathrm{I}}^j = (n/m)\|X^j\|_2^2$ at every machine $j = 1, \dots, m$. Under the null hypothesis, S_{I}^j follows a chi-square distribution with d degrees of freedom, that is, $S_{\mathrm{I}}^j \sim \chi_d^2$. Letting $F_{\chi_d^2}$ denote χ_d^2-CDF, the quantity $F_{\chi_d^2}(S_{\mathrm{I}}^j)$ can be seen as the p-value for the local test statistic S_{I}^j. Based on these "local p-values," we then generate the randomized transcript Y_{I}^j for every j using Bernoulli random variables,

$$Y_{\mathrm{I}}^j | S_{\mathrm{I}}^j \sim \mathrm{Ber}\big(F_{\chi_d^2}(S_{\mathrm{I}}^j)\big).$$

For a given $\alpha \in (0, 1)$, we can construct the test

$$T_{\mathrm{I}} = \mathbb{1}\left\{\left|\frac{1}{m}\left(\sum_{j=1}^{m}(Y_{\mathrm{I}}^j - 1/2)\right)^2 - 1/4\right| \geq \kappa_\alpha\right\} \tag{8}$$

at the central machine. In applications one could set, for instance, κ_α such that $\mathbb{P}_0 T_{\mathrm{I}} \approx \alpha$ by considering that $\sum_{j=1}^m Y_{\mathrm{I}}^j$ is $(m, 1/2)$-binomially distributed under the null. Lemma A.6 in the Supplementary Material yields that, for each $\alpha \in (0, 1)$, there exist constants $\kappa_\alpha, C_\alpha, M_\alpha, D_0 > 0$ such that, for $m \geq M_\alpha$ and $d \geq D_0$, it holds that $\mathcal{R}(H_\rho, T_{\mathrm{I}}) \leq \alpha$, whenever $\rho^2 \geq C_\alpha \frac{\sqrt{md}}{n}$.

The case $m \leq M_\alpha$ corresponds essentially to the nondistributed setting and is treated separately for technical reasons. In practice, one would simply use the test, given in (8), also for $m \leq M_\alpha$. Furthermore, if one allows for a slightly larger amount of bits (e.g., $\log_2(n)$ bits), one could opt to transmit an (approximation of) the test statistics S_{I}^j themselves;

see, for example, Lemma 2.3 in [27] for which it is easy to prove that the rate of $\frac{\sqrt{md}}{n}$ is achieved without requiring any assumptions on m. For the sake of completeness, by considering $\rho^2 \geq C_\alpha \sqrt{M_\alpha} \frac{\sqrt{d}}{n}$, we see that the optimal rate of $\frac{\sqrt{md}}{n}$ can be achieved in the $m \leq M_\alpha$ case by simply taking

$$T_{\mathrm{I}}' := Y_{\mathrm{I}}^1 := \mathbb{1}\left\{ \frac{1}{\sqrt{d}} (S_{\mathrm{I}}^1 - d) \geq \kappa_\alpha \right\} \tag{9}$$

for an appropriately large choice of the constant κ_α. Similarly, the requirement that d is larger than some constant D_0 (which is independent of α) appears for technical reasons. The case where $d \leq D_0$ is covered by the private coin protocol T_{III} in Section 4.3.

4.2. *Public coin*, *high communication budget*: *Construction of* T_{II}. We now switch our attention to exhibiting a testing procedure that is optimal when $bm \gtrsim d$ in the public coin case. That a shared source of randomness in distributed settings can be strictly better than private ones, in terms of *communication complexity*, is an idea that goes back to [36]. Essentially, the use of shared randomness allows for the machines coordinate their efforts in "covering" each of the d dimensions of the data, even though all communication happens in just one round; see also, for example, Chapter 3 in [23] for an extensive treatment of this phenomenon. We adopt ideas proposed by [3], who consider the setting where $m = n$ with asymptotics in m. We exhibit this testing protocol below and provide a full proof covering also the case where $m \neq n$. To that extend, let U be a random rotation, that is, U is drawn from the Haar measure (see, e.g., Theorem F.13 in [5]) on the set of orthonormal matrices in $\mathbb{R}^{d\times d}$. At each machine and for $b \leq d$, we can compute the b-bit transcript $Y_{\mathrm{II}}^j \in \{0, 1\}^b$ conditionally on the shared public coin draw U, where each of the $1 \leq i \leq b$ components is defined through

$$(Y_{\mathrm{II}}^j)_i | U, X^j = \mathbb{1}\{(\sqrt{n/m} U X^j)_i > 0\},$$

where $(v)_i$ denotes the projection onto the ith coordinate of the vector $v \in \mathbb{R}^d$. The random rotation fulfills a similar purpose, as the random reweighting algorithm proposed in [28], but leads to an easier proof in the d-dimensional case because of rotational invariance of the Gaussian distribution.

Centrally, after transmitting $(Y^1, \ldots, Y^m)$, we compute the aggregated test statistics $S_{\mathrm{II}} = \sum_{j=1}^m Y_{\mathrm{II}}^j$ and define the corresponding test as

$$T_{\mathrm{II}} = \mathbb{1}\left\{ \left| \frac{1}{\sqrt{bm}} \sum_{i=1}^b \left((S_{\mathrm{II}})_i - \frac{m}{2} \right)^2 - \sqrt{b}/4 \right| > \kappa_\alpha \right\}. \tag{10}$$

Lemma A.7 in the Supplementary Material shows that this test achieves the public coin lower bound when $mb \gtrsim d$ and $m \geq M_\alpha$.

4.3. *Private coin*, *high total communication budget*: *Constructing* T_{III}. Finally, we consider the case of not having access to a public coin but having a relatively large communication budget ($b^2 m \gtrsim d^2$). Note that we can assume without loss of generality that $m \geq M_\alpha d^2/b^2$ for a constant $M_\alpha > 0$, as otherwise the optimal rate is $\sqrt{md}/n$, obtained by the *one*-bit private coin test described by (8) (see Section 4.1). This case is the most involved one, and we construct a test consisting two subtests optimal in different subregimes.

The most obvious approach in this case is to divide the communication budget of each machine over the d coordinates as uniformly as possible. That is to say, to partition the coordinates $\{1, \ldots, d\}$ into approximately d/b sets of size b (we assume without loss of generality that $b \leq d$, as we can always throw away excess budget and $b = d$ bits suffice for achieving the minimax rate). The machines are then equally divided over each of these partitions and communicate the coefficients corresponding to their partition. More formally, such a strategy entails taking sets $\mathcal{I}_i \subset \{1, \dots, m\}$ such that $|\mathcal{I}_i| = \lfloor \frac{mb}{d} \rfloor$ and each $j \in \{1, \dots, m\}$ is in $\mathcal{I}_i$ for b different indexes $i \in \{1, \dots, d\}$. For $i = 1, \dots, d$ and $j \in \mathcal{I}_i$, generate the transcripts according to

$$Y_i^j | X_i^j = \mathbb{1}\{X_i^j > 0\}. \tag{11}$$

Centrally, a natural test based on these transcripts is

$$T_{\mathrm{III}}^1 := \mathbb{1}\left\{\left|\frac{1}{|\mathcal{I}_1|\sqrt{d}} \sum_{i=1}^{d} \left(\sum_{j \in \mathcal{I}_i} (Y_i^j - 1/2)\right)^2 - \sqrt{d}/4\right| > \kappa_\alpha\right\}. \tag{12}$$

It turns out that such a test does not cover all regimes, where $m \gtrsim d^2/b^2$, because there is a certain amount of information loss due to the nonlinearity of the quantization step (11); that is, the test induces soft thresholding for the signal components, which is suboptimal for (relatively) large signal components. For the exact statement on the testing error of this test, see Lemma A.9 below.

For detecting signals, including large coordinates, we propose an adaptation of test T_{III}^1. We start by assuming that $b \geq 2\log(d+1)$; otherwise, we do not construct the test. Then for $i = 1, \dots, d$ and $j = 1, \dots, m$, let us generate

$$B_{li}^j \overset{\text{iid}}{\sim} \mathrm{Ber}\big(F_{\chi_1^2}((\sqrt{n/m} X_i^j)^2)\big), \quad l \in \{1, \dots, C_{b,d} = \lfloor 2^b/(d+1) \rfloor\}.$$

Note that $C_{b,d} \geq 1$ by assumption. Then machine j communicate the transcripts

$$N^j = \sum_{l=1}^{C_{b,d}} \sum_{i=1}^{d} B_{li}^j \in \{0, 1, \dots, C_{b,d} d\}, \tag{13}$$

which can be done using $\log_2(C_{b,d} d + 1) \leq b$ bits in total. Based on these transcripts, we compute the test

$$T_{\mathrm{III}}^2 = \mathbb{1}\left\{\left|\frac{1}{dmC_{b,d}} \left(\sum_{j=1}^{m} (N^j - Ld/2)\right)^2 - \frac{1}{4}\right| \geq \kappa_\alpha\right\}$$

centrally. The testing risk bound for the above test is given in Lemma A.10 below.

Finally, we construct our test by combining the above ones. We construct both partial tests T_{III}^1 and T_{III}^2 if $b \geq 2\log(d+1)$ by transmitting $b' = \lfloor b/2 \rfloor$ bits per machine for each; otherwise, we just construct T_{III}^1. Then we merge them by taking

$$T_{\mathrm{III}} = T_{\mathrm{III}}^1 \vee T_{\mathrm{III}}^2 \mathbb{1}_{\{b \geq 2\log(d+1)\}}, \tag{14}$$

where the indicator should be understood to rule out cases in which the transcripts for T_{III}^2 cannot necessarily be communicated. This case, as shown below, is covered by the first test T_{III}^1. Lemma A.8 in the Supplementary Material shows that T_{III} has sufficiently small testing risk in all cases where $m \geq M_\alpha d^2/b^2$.

5. A sketch of proof for the testing lower bound (Theorem 3.1). In this section we provide a sketch of proof of Theorem 3.1 of which the full details are given in Section 9. The proof starts out the same way for both the private and public coin cases but bifurcates later on. We consider for the time being a generic collection of b-bit distributed testing protocols $\mathcal{T}(b)$.

As a first step, we introduce a prior distribution π on $\mathbb{R}^d$ and lower bound the testing risk by a type of Bayes risk and the mass of π that resides outside of the alternative hypothesis

H_ρ, akin to, for example, [19]. Recall that $\mathbb{P}_f$ denotes the joint distribution of Y, U and X where X^j follows (1) and $Y \sim \mathbb{E}_f^{X,U} K(\cdot|X,U) =: \mathbb{P}_{f,K}^Y = \mathbb{P}_f^Y$. For π a given distribution on $\mathbb{R}^d$, define the mixture distribution $\mathbb{P}_\pi^X = P_\pi$ on $\mathbb{R}^{md}$ by $P_\pi(A) = \int P_f(A)\, d\pi(f)$, where we recall the notational convention $\mathbb{P}_f^X = P_f$ from Section 2.

Through the Markov chain relation $f \to X \to Y$, this defines a distribution $\mathbb{P}_\pi^Y = \mathbb{P}_{\pi,K}^Y$ on $\mathcal{Y}$ and lets us denote by $\mathbb{E}_\pi^Y$ the corresponding expectation. Lemma A.1 in the Supplementary Material lower bounds the infimum testing risk $\inf_{T\in\mathcal{T}} \mathcal{R}(H_\rho, T)$ using a version of Le Cam's lemma adapted to the distributed setting. The lemma yields that, for any distribution on U,

$$\inf_{T\in\mathcal{T}}\Big(\mathbb{E}_0^Y T(Y) + \sup_{f\in H_\rho} \mathbb{E}_f^Y\big(1 - T(Y)\big)\Big) \geq \inf_K \sup_\pi \big(1 - \|\mathbb{P}_{0,K}^Y - \mathbb{P}_{\pi,K}^Y\|_{\mathrm{TV}} - \pi(H_\rho^c)\big),$$

where the infimum on the RHS is over all kernels on $\mathcal{Y}$.

Using that the measure $d\mathbb{P}_f^Y$ disintegrates as $d\mathbb{P}_f^{Y|U=u}\, d\mathbb{P}_f^U(u)$ and the fact that U is independent of the prior π, we find by Jensen's inequality that

$$\|\mathbb{P}_{0,K}^Y - \mathbb{P}_{\pi,K}^Y\|_{\mathrm{TV}} \leq \int \|\mathbb{P}_{0,K}^{Y|U=u} - \mathbb{P}_{\pi,K}^{Y|U=u}\|_{\mathrm{TV}}\, d\mathbb{P}^U(u).$$

By Pinsker's second inequality and the fact that $\log(x) \leq x - 1$, we obtain that

$$\inf_{T\in\mathcal{T}(b)} \mathcal{R}(H_\rho, T) \geq 1 - \sup_K \inf_\pi \Big(\int \sqrt{2D_{\chi^2}\big(\mathbb{P}_{0,K}^{Y|U=u}; \mathbb{P}_{\pi,K}^{Y|U=u}\big)}\, d\mathbb{P}^U(u) + \pi(H_\rho^c)\Big), \tag{15}$$

where

$$D_{\chi^2}\big(\mathbb{P}_{0,K}^{Y|U=u}; \mathbb{P}_{\pi,K}^{Y|U=u}\big) = \mathbb{E}_{0,K}^{Y|U=u}\bigg(\frac{\mathbb{P}_{\pi,K}^{Y|U=u}}{\mathbb{P}_{0,K}^{Y|U=u}}(Y)\bigg)^2 - 1. \tag{16}$$

From hereon, the proof can be broken down into two steps. We provide the skeleton of the proof here and defer the full details to Section 9:

1. The first term on the RHS of (16) can be expressed in terms of a conditional expectation of the likelihood of X,

$$\mathbb{E}_0^{Y|U=u}\Bigg(\mathbb{E}_0\bigg[\int \prod_{j=1}^m \frac{d\mathbb{P}_f^{X^j}}{d\mathbb{P}_0^{X^j}}(X^j)\, d\pi(f) \Big| Y, U=u\bigg]^2\Bigg), \tag{17}$$

which we compare to the quantity

$$\prod_{j=1}^m \mathbb{E}_0^{Y^j|U=u}\bigg(\mathbb{E}_0\Big[\int \frac{d\mathbb{P}_f^{X^j}}{d\mathbb{P}_0^{X^j}}(X^j)\, d\pi(f) \Big| Y^j, U=u\Big]^2\bigg), \tag{18}$$

which corresponds to the product of the first terms of the local chi-square divergences. In particular, we compare the ratio of the expressions in the above two displays and show that when π is taken to be centered Gaussian, this ratio is maximized when the protocol's kernel $K : L_2(\mathcal{Y}) \to L_2(\mathcal{X})$ with Hilbert space adjoint K^* satisfies that $K^*K : L_2(\mathcal{X}) \to L_2(\mathcal{X})$ is Gaussian in an appropriate sense. This is the content of Lemma 10.1, which forms the crux of our proof. This lemma, on which we expound in Section 10, exploits the conjugacy between the prior and the model, which enables the use of techniques applied in [21]. Consequently, we obtain that the first term on the RHS of (16) is bounded from above by a multiple of

$$\prod_{j=1}^m \mathbb{E}_0\big[\mathscr{L}_\pi(X^j)^2\big] \int \exp(f^\top \Xi_u g)\, d(\pi\times\pi)(f,g)\, d\mathbb{P}^U(u), \tag{19}$$

where

$$\Xi_u := \sum_{j=1}^{m} \mathbb{E}_0^{Y^j|U=u} \mathbb{E}_0[X^j|Y^j, U=u]\mathbb{E}_0[X^j|Y^j, U=u]^\top. \tag{20}$$

2. The final step combines data-processing techniques with what is essentially a geometric argument. The first term in (19) is handled using classical, nondistributed techniques, that is, decoupling argument of the measure and the moment generating function of the Gaussian chaos; see, for example, [34]. In the second term in (19), the $d \times d$ matrix Ξ_u geometrically captures how well Y allows to "reconstruct" the compressed sample X. The information lost by compressing a d dimensional observation X^j into a b-bit transcript Y^j is captured in a data-processing inequality for the matrix Ξ_u and its trace, which comes in the form of Lemma A.2 and Lemma A.3. From hereon out, the proof of the private and the public coin cases separate. Recalling the order of the supremum, infimum and expectation with respect to the public coin in (15), we see that, in the private coin case, π can be chosen with knowledge of Ξ_u, as U is degenerate in this case. To obtain the stricter lower bound in the private coin case, we choose π's covariance in order to exploit the "weakest directions" of the protocol Y, and the proof is finished by matrix algebra arguments.

6. Nonparametric testing with known regularity. A natural extension of the above finite dimensional signal in Gaussian noise setting is the infinite dimensional signal in white noise model. Here the $j = 1, \dots, m$ machines observe i.i.d. X^j, taking values in $\mathcal{X} \subset L_2[0,1]$ and subject to the stochastic differential equation

$$dX_t^j = f(t)\,dt + \sqrt{\frac{m}{n}}\,dW_t^j \tag{21}$$

under P_f, with $W^1, \dots, W^m$ i.i.d. Brownian motions and $f \in L_2[0,1]$. Besides the difference in the local observations, the distributed setup considered for this model remains exactly the same. The results derived for the alternatives H_ρ in the finite dimensional model translate to testing in the infinite dimensional model against the alternative hypotheses

$$f \in H_\rho^{s,R} := \{ f \in \mathcal{H}^{s,R}[0,1] : \|f\|_{L_2} \geq \rho \text{ and } \|f\|_{\mathcal{H}^s} \leq R \}.$$

Here $\mathcal{H}^{s,R} = \mathcal{H}^{s,R}([0,1])$ denotes the Sobolev ball of radius R in the space of s-smooth Sobolev functions and $\|\cdot\|_{\mathcal{H}^s}$ the Sobolev norm; see Section G for recalling the definitions. The smoothness parameter $s > 0$ determines the difficulty of the classical (nondistributed, $m = 1$) nonparametric testing problem, as considered in, for example, [19]. The asymptotic minimax rate for the nondistributed case is $\rho^2 \asymp n^{-\frac{2s}{2s+1/2}}$ for the s-smooth Sobolev alternative class.

We allow for asymptotics in b and m, in the sense that they can depend on n. Consequently, we consider the separation rate ρ in the nonparametric problem to be a sequence of positive numbers in both n, m and the budget b. A distributed test T in the nonparametric setting is called α-consistent for $\alpha \in (0,1)$ if $\mathcal{R}(H_\rho^{s,R}, T) \leq \alpha$ for all n large enough.

The distributed setting for the nonparametric model remains unchanged in comparison with the finite dimensional model introduced in Section 2, except, of course, for the sample space in which the observations X^j take values. This becomes $L_2[0,1]$, instead of $\mathbb{R}^d$. The following theorem describes the minimax rate for the nonparametric distributed problem.

THEOREM 6.1 (Nonparametric signal in white noise minimax rate). *Take $f \in H^{s,R}$ for some $s, R > 0$, let $b \equiv b_n$ and $m \equiv m_n$ be sequences of natural numbers and let $\rho \equiv \rho_{n,b,m,s}$*

be a sequence of positive numbers satisfying

$$\rho^2 \asymp \begin{cases} n^{-\frac{2s}{2s+1/2}} & \text{if } b \geq n^{\frac{1}{2s+1/2}}, \\ (\sqrt{b}n)^{-\frac{2s}{2s+1}} & \text{if } n^{\frac{1}{2s+1/2}}/m^{\frac{2s+1}{2s+1/2}} \leq b < n^{\frac{1}{2s+1/2}}, \\ (n/\sqrt{m})^{-\frac{2s}{2s+1/2}} & \text{if } b < n^{\frac{1}{2s+1/2}}/m^{\frac{2s+1}{2s+1/2}}. \end{cases} \tag{22}$$

In the public coin protocol case, the minimax testing rate is ρ^2 *given in* (22); *that is, for all* $\alpha \in (0, 1)$, *there exist constants* $C_\alpha, c_\alpha > 0$, *depending only on* α, s *and* R, *such that for all* n *large enough,*

$$\inf_{T \in \mathcal{T}_{\text{pub}}(b)} \mathcal{R}\big(H^{s,R}_{c_\alpha \rho}, T\big) > 1 - \alpha \quad \textit{and} \quad \inf_{T \in \mathcal{T}_{\text{pub}}(b)} \mathcal{R}\big(H^{s,R}_{C_\alpha \rho}, T\big) \leq \alpha.$$

Similarly, in the private coin protocol case $\rho \equiv \rho_{n,b,m}$ *given below,*

$$\rho^2 \asymp \begin{cases} n^{-\frac{2s}{2s+1/2}} & \text{if } b \geq n^{\frac{1}{2s+1/2}}, \\ (bn)^{-\frac{2s}{2s+3/2}} & \text{if } n^{\frac{1}{2s+1/2}}/m^{\frac{s+3/4}{2s+1/2}} \leq b < n^{\frac{1}{2s+1/2}}, \\ (n/\sqrt{m})^{-\frac{2s}{2s+1/2}} & \text{if } b < n^{\frac{1}{2s+1/2}}/m^{\frac{s+3/4}{2s+1/2}}, \end{cases} \tag{23}$$

provides the minimal testing rate; that is, for all $\alpha \in (0, 1)$, *there exist constants* $C_\alpha, c_\alpha > 0$, *depending only on* α *and* R, *such that for all* n *large enough,*

$$\inf_{T \in \mathcal{T}_{\text{priv}}(b)} \mathcal{R}\big(H^{s,R}_{c_\alpha \rho}, T\big) > 1 - \alpha \quad \textit{and} \quad \inf_{T \in \mathcal{T}_{\text{priv}}(b)} \mathcal{R}\big(H^{s,R}_{C_\alpha \rho}, T\big) \leq \alpha.$$

The proof of the theorem is given in Section B. The theorem reveals the relationship between the signal-to-noise-ratio n, communication budget per machine b, the number of machines m and the smoothness of the signal s. Before providing the proof, we briefly discuss the connection with distributed minimax estimation rates.

The distributed minimax estimation rates under private coin protocol were established in Corollary 2.2 of [26] or Theorem 3.1 in [38]. A slight reformulation of the latter yields that

$$\begin{aligned} &\inf_{(\hat{f}, \mathcal{L}(Y)) \in \mathcal{E}_{\text{priv}}(b)} \sup_{f \in \mathcal{H}^{s,R}} \mathbb{E}_f^Y \|\hat{f}(Y) - f\|_{L_2}^2 \\ &\asymp \begin{cases} n^{-\frac{2s}{2s+1}} & \text{if } b \geq n^{\frac{1}{2s+1}}, \\ (bn)^{-\frac{2s}{2s+2}} & \text{if } \big(n/m^{2+2s}\big)^{\frac{1}{2s+1}} \leq b \leq n^{\frac{1}{2s+1}}, \\ (bm)^{-2s} & \text{if } b \leq \big(n/m^{2+2s}\big)^{\frac{1}{2s+1}}, \end{cases} \end{aligned} \tag{24}$$

where $\mathcal{E}_{\text{priv}}(b)$ is the class of all distributed estimators based on b-bit transcripts $Y = (Y^1, \dots, Y^m)$.

A first observation is that consistent testing is possible in any regime of $b \geq 1$ and m, whereas this is not the case in estimation. Consider, for instance, the regime where m and b are fixed. In nonparametric distributed estimation, the L_2-risk does not improve once the sample size is large enough. In fact, even when allowing for asymptotics in b and m (but assuming that $(n/m^{2+2s})^{\frac{1}{2s+1}} \geq b$), one is better off performing the estimation locally using just one of the machines with local signal-to-noise-ratio n/m, attaining the locally optimal rate $(n/m)^{-\frac{2s}{2s+1}}$.

In the case of nonparametric testing, not only can we consistently test for any fixed m and b, the distributed testing rate is bounded from above by $(n/\sqrt{m})^{-2s/(2s+1/2)}$ (regardless of the communication budget b), which is significantly smaller (for large m) than the minimax testing rate, based on the local signal-to-noise-ratio $(n/m)^{-2s/(2s+1/2)}$, which can be achieved by using only a single local machine. One possible explanation for this discrepancy is that, in nonparametric estimation, the output of the inference is a high-dimensional object, which requires a large total communication budget to be reconstructed with sufficient granularity. In testing, the output of our inference is binary.

A perhaps less surprising difference is that a larger budget is needed for testing at the nondistributed minimax testing rate compared to estimation. That is, in order to obtain the nondistributed minimax rate of $\rho^2 \asymp n^{-\frac{2s}{2s+1/2}}$, the communication budget needs to satisfy $b \gtrsim n^{\frac{1}{2s+1/2}}$. On the other hand, the nondistributed minimax estimation rate $n^{-\frac{2s}{2s+1}}$ requires only $b \gtrsim n^{\frac{1}{2s+1}}$. This follows from the fact that the L_2 testing rate is faster than the estimation rate, and hence to achieve this faster rate, one has to collect information about the signal at higher frequency level as well (up to the $O(n^{\frac{1}{2s+1/2}})$ coefficients in the spectral decomposition).

Increasing m decreases the local signal-to-noise-ratio. When the total budget bm grows at a similar or faster rate than the "effective dimension" of the model, the rate that can be achieved no longer depends on m in both estimation and testing settings. In this regime this effect is offset by the total number of bits being received by the central machine. What is different in testing problem, however, is that having access to shared randomness strictly improves the performance (until the local communication budget b reaches the effective dimension $n^{\frac{1}{2s+1/2}}$ as after that both method reaches the minimax nondistributed testing rate $n^{-\frac{2s}{2s+1/2}}$). One might wonder whether having access to a public coin improves the rate in the estimation setting also. It turns out that this is not the case. We show in Theorem C.1 in the Supplementary Material that, under the public coin protocol, the distributed minimax estimation rate does not improve compared to the private coin protocol.

7. Adaptation in nonparametrics. In the previous section we have derived minimax lower and matching upper bounds for the nonparametric distributed testing problem in context of the Gaussian white noise model. The proposed tests, however, depend on the regularity hyperparameter s of the functional parameter of interest f. Typically, the regularity of the function is not known in practice, and one has to use data driven methods to find the best testing strategies. In this section we derive distributed tests adapting to this unknown regularity. We derive both lower and upper bounds and observe surprising, additional phase transition in the small budget regime, which was not present in the nonadaptive setting.

First, we note that, even in the nondistributed setting, we have to pay an additional $\log\log n$ factor as a price for adaptation (see, e.g., Theorem 2.3 in [25] or Section 7 in [19]). More concretely, if $\rho_s \asymp n^{-s/(2s+1/2)}$, it holds that, for any $s_{\min} < s_{\max}$,

$$\sup_{s\in[s_{\min},s_{\max}]} \mathcal{R}\big(H^{s,R}_{c_n M_{n,s}\rho_s}, T\big) \to 1$$

for all tests T, $M_{n,s} = (\log\log n)^{\frac{s/4}{2s+1/2}}$ and any $c_n = o(1)$ while there exists a test T, satisfying

$$\sup_{s\in[s_{\min},s_{\max}]} \mathcal{R}\big(H^{s,R}_{C M_{n,s}\rho_s}, T\big) \to 0,$$

for large enough constant $C > 0$.

The distributed testing problem is more complicated, as we have to consider different regimes based on the number of transmitted bits; see Theorem 6.1. These regimes, however, depend on the unknown regularity hyperparameter and require different testing procedures to achieve consistent testing. The transcripts transmitted require a larger communication budget to attain the same performance as in Theorem 6.1. Theorems 7.1 and 7.2 below capture this increased difficulty in terms of lower and upper bounds on the detection rate (tight up to a log-log factor). In the proof of the theorem, we derive such an adaptive distributed testing method which adapts to the smoothness. These methods are, in principle, based on taking a $1/\log n$ grid of the regularity interval $[s_{\min}, s_{\max}]$, constructing optimal tests for each of the grid points and combining them using Bonferroni's correction. This results in loosing a logarithmic factor in the intermediate case, as the budget has to be divided over $O(\log n)$ tests, each capturing a different possible level of smoothness.

This additional incurred cost in the distributed setting, due to additional communication budget required, is fundamental, as our accompanying lower bound shows. This additional difficulty translates to a $\sqrt{\log(n)}$ and $\log(n)$ factor more observations required in the intermediate budget regimes for the public and private coin settings, respectively. In the small budget regime, such a loss is incurred when the local communication budget b is of smaller order than $\log(n)$. When $b \gtrsim \log(n)$ in the small budget regime, the same rate as in Theorem 6.1 can be obtained, up to the $\log\log(n)$ factor incurred by the Bonferroni correction.

The above described results are split over two theorems. The first, Theorem 7.1, concerns the case where $b \gtrsim \log(n)$. In the second, Theorem 7.2, the case where $b \lesssim \log(n)$ (both theorems coincide when $b \asymp \log(n)$). The case where $b = O(1)$ is of special interest, as $b = 1$ means each machine's local transcript forms a test itself and the global test can be seen as a "meta-analysis" on the basis of these m tests. The proofs of the upper bounds in both theorems are given in Section 8, while the proofs of the lower bound are deferred to Section D in the Supplementary Material.

THEOREM 7.1. *Let us consider some* $0 < s_{\min} < s_{\max} < \infty$, $R > 0$; *let* $b \equiv b_n$ *such that* $b \gg \log n$ *and* $m \equiv m_n$ *be sequences of natural numbers, and take a sequence of positive numbers* $\rho_s \equiv \rho_{n,b,m,s}$ *satisfying*

$$(25)\qquad \rho_s^2 \asymp \begin{cases} n^{-\frac{2s}{2s+1/2}} & \text{if } b \geq \log(n) n^{\frac{1}{2s+1/2}}, \\ \left(\dfrac{\sqrt{bn}}{\sqrt{\log(n)}}\right)^{-\frac{2s}{2s+1}} & \text{if } \log(n)\left(\dfrac{n^{\frac{1}{2s+1/2}}}{m^{\frac{2s+1}{2s+1/2}}} \vee 1\right) \leq b < \log(n) n^{\frac{1}{2s+1/2}}, \\ \left(\dfrac{n}{\sqrt{m}}\right)^{-\frac{2s}{2s+1/2}} & \text{if } \log(n) \leq b < \log(n)\left(\dfrac{n^{\frac{1}{2s+1/2}}}{m^{\frac{2s+1}{2s+1/2}}} \vee 1\right) \end{cases}$$

in the public coin case, and

$$(26)\qquad \rho_s^2 \asymp \begin{cases} n^{-\frac{2s}{2s+1/2}} & \text{if } b \geq \log(n) n^{\frac{1}{2s+1/2}}, \\ \left(\dfrac{bn}{\log(n)}\right)^{-\frac{2s}{2s+3/2}} & \text{if } \log(n)\left(\dfrac{n^{\frac{1}{2s+1/2}}}{m^{\frac{s+3/4}{2s+1/2}}} \vee 1\right) \leq b < \log(n) n^{\frac{1}{2s+1/2}}, \\ \left(\dfrac{n}{\sqrt{m}}\right)^{-\frac{2s}{2s+1/2}} & \text{if } \log(n) \leq b < \log(n)\left(\dfrac{n^{\frac{1}{2s+1/2}}}{m^{\frac{s+3/4}{2s+1/2}}} \vee 1\right) \end{cases}$$

in the case of a private coin. Then there exits a sequence of distributed testing procedures in the respective setups such that

$$\sup_{s \in [s_{\min}, s_{\max}]} \mathcal{R}(H^{s,R}_{M_n \rho_s}, T) \to 0,$$

for arbitrary $M_n \gg (\log\log(n))^{1/4}$. *Similarly, for all distributed testing procedures in the respective setups, we have that, for all* $\alpha \in (0, 1)$, *there exists* $c_\alpha > 0$ *such that*

$$\sup_{s \in [s_{\min}, s_{\max}]} \mathcal{R}(H^{s,R}_{c_\alpha \rho_s}, T) > \alpha.$$

The above theorem recovers (up to log-factors) the three rates corresponding to the three regimes also found in Theorem 6.1, the different regimes corresponding to different testing strategies. Since the true smoothness is unknown, these different distributed testing strategies are to be conducted simultaneously.

We note that, for $m \geq n^{\frac{1}{2s_{\min}+1}}$ or $m \geq n^{\frac{1}{s_{\min}+3/4}}$ in the public and private coin cases, respectively, the small budget regime no longer occurs. The reason for this is that, even though b could be relatively small, the total communication budget bm is large enough to warrant the strategy for the intermediate and high budget regimes. Furthermore, whenever $b > \log(n) n^{\frac{1}{2s+1/2}}$, the budget is large enough to recover the nondistributed regime rate.

For $b \lesssim \log(n)$, the separation rate is different from the nonadaptive low budget regime. Depending on the interplay between n and m, either the minimax rate corresponding to the intermediate case applies or an additional $(\log(n)/b)^{\delta}$ factor is present compared to the nonadaptive low budget regime, both in the private and public coin settings. This results in an additional phase transition at $b = \log n$. The reason for this is that, in order to cover approximately $\log(n)$ different levels of smoothness using less than $\log(n)$ bits, each of the machines can no longer send an adequate amount of information on all of the relevant smoothness levels. Instead, an optimal strategy is to divide the different machines over each of the smoothness levels, where each machines foregoes sending information regarding certain smoothness levels all together.

THEOREM 7.2. *Assume the conditions of Theorem 7.1 with* $b \lesssim \log(n)$, *and assume* $bm \gg \log(n)$. *Let us consider*

$$\rho_s^2 \asymp \begin{cases} \left(\frac{\sqrt{b}n}{\sqrt{\log(n)}}\right)^{-\frac{2s}{2s+1}} & \text{if } m \geq n^{\frac{1}{2s+1}}, \\ \left(\frac{\sqrt{b}n}{\sqrt{m\log(n)}}\right)^{-\frac{2s}{2s+1/2}} & \text{if } m < n^{\frac{1}{2s+1}} \end{cases} \tag{27}$$

in the public coin case and

$$\rho_s^2 \asymp \begin{cases} \left(\frac{bn}{\log(n)}\right)^{-\frac{2s}{2s+3/2}} & \text{if } m \geq n^{\frac{2}{2s+3/2}}\left(\frac{b}{\log(n)}\right)^{\frac{s-1/4}{2s+3/2}}, \\ \left(\frac{n\sqrt{b}}{\sqrt{m\log(n)}}\right)^{-\frac{2s}{2s+1/2}} & \text{if } m < n^{\frac{2}{2s+3/2}}\left(\frac{b}{\log(n)}\right)^{\frac{s-1/4}{2s+3/2}} \end{cases} \tag{28}$$

in the private coin case. Then there exits a sequence of distributed testing procedures in the respective setups such that

$$\sup_{s\in[s_{\min},s_{\max}]} \mathcal{R}\big(H^{s,R}_{M_n\rho_s}, T\big) \to 0$$

for arbitrary $M_n \gg (\log\log(n))^{1/4}$. *Similarly, for all distributed testing procedures in the respective setups, we have that, for all* $\alpha \in (0,1)$, *there exists* $c_\alpha > 0$ *such that*

$$\sup_{s\in[s_{\min},s_{\max}]} \mathcal{R}\big(H^{s,R}_{c_\alpha\rho_s}, T\big) > \alpha.$$

REMARK 7.3. Both theorems together cover all cases where $mb \gg \log(n)$. The cases where $mb \lesssim \log(n)$ are excluded for technical reasons as well as the fact that, when $mb \lesssim \log(n)$, the optimal rate in (27)–(28) (up to at most a $\sqrt{\log\log(n)}$ factor) is attained by using a standard nondistributed method using just the data of one machine (see, e.g., [25]). Similarly, in order to contain the level of technicality, we have foregone the $(\log\log(n))^{1/4}$ additional factor in the lower bound, which we esteem also to be present in the distributed setting. We refer the reader to the argument of Theorem 2.3 in [25] for how to obtain the $(\log\log(n))^{1/4}$ factor in the lower bound in addition to the $\sqrt{\log(n)}$ and $\log(n)$ factors in the public and private coin cases, respectively.

8. Adaptive tests attaining the adaptation bounds in Theorems 7.1 and 7.2. Let us consider the smooth orthonormal wavelet basis $\{\psi_{li} : l \in \mathbb{N}_0, i = 0, 1, \dots, 2^l - 1\}$; see Section G for a brief introduction of wavelets and collection of properties used in this proof. For $L = L \in \mathbb{N}$, let $V_L = \{\psi_{li} : l \leq L, i = 0, 1, \dots, 2^l - 1\}$. For $f \in L_2[0, 1]$, let f^L denote the projection of f onto V_L, that is,

$$f^L = \sum_{l=0}^{L} \sum_{i=0}^{2^l-1} \tilde{f}_{li} \psi_{li} \tag{29}$$

with $\tilde{f}_{li} := \int f \psi_{li}$. We denote the wavelet coefficients of X^j by $\tilde{X}^j_{li} := \int_0^1 \psi_{li}\, dX^j_t$. For the coefficients at resolution level L, write $\tilde{X}^j_L = (\tilde{X}^j_{L0}, \dots, \tilde{X}^j_{L(2^L-1)}) \in \mathbb{R}^{2^L}$, and let $\tilde{X}^j_{L':L}$ denote the concatenated coefficients from resolution level $L' < L$ up to resolution level L, that is, $\tilde{X}^j_{L':L} = (\tilde{X}^j_{L'}, \dots, \tilde{X}^j_L) \in \mathbb{R}^{2^{L+1}-2^{L'+1}}$. The vector $\tilde{X}^j_{0:L} := (\tilde{X}^j_0, \tilde{X}^j_1, \dots, \tilde{X}^j_L)$ follows the dynamics

$$\tilde{X}^j_{0:L} = \tilde{f}^L + \sqrt{\frac{m}{n}} Z^j, \tag{30}$$

where $Z^j \sim^{\text{iid}} N(0, I_{2^{L+1}-1})$, $j = 1, \dots, m$, and $\tilde{f}^L := (\tilde{f}_{li})_{l=0,\dots,L; i=0,\dots,2^l-1}$.

Let $\nu_L = 2^{L+1} - 1$, and let us introduce the notations $L_s = \lfloor s^{-1} \log(1/\rho_s) \rfloor \vee 1$; for shorthand, write $L_{\min} = L_{s_{\max}}$ and $L_{\max} = L_{s_{\min}}$, and note that $L_s \in \mathcal{C} := \{L_{\min}, \dots, L_{\max}\}$ for all $s \in [s_{\min}, s_{\max}]$. Note that $|\mathcal{C}| \leq \log n$.

For each regularity hyperparameter s, we distinguish low-budget ($2^{L_s} \gtrsim mb$ in the public coin, and $2^{\frac{3}{2}L_s} \gtrsim mb$ in the private coin setting) and high-budget (corresponding to $2^{L_s} \lesssim mb$ in the public coin, and $2^{\frac{3}{2}L_s} \lesssim mb$ in the private coin setting) cases. Since m and b are known for any given regularity s, we know which regime it falls and is sufficient to construct that test. For notational convenience and without loss of generality, for each s we construct both the high-budget and the low-budget optimal tests using all the m machines (and do not split them between these two cases).

8.1. *Proof of the upper bound in the low-budget regime.* First, we deal with the low-budget case (where the total budget is small compared to the effective dimension), which coincides in both setups. For each $L \in \mathcal{C}$, we take a subset of machines $M_L \subset \{1, \dots, m\}$ such that $|M_L| = m' := \frac{m(\log(n) \wedge b)}{\log(n)}$, and each machine appears in at most b such subsets. We note that this is possible since $m'|\mathcal{C}| \leq mb$. Then for each $j \in M_L$, $L \in \mathcal{C}$, we communicate

$$Y^j_{\mathrm{I}}(L) | X^j \sim \mathrm{Ber}\big(\chi^2_{\nu_L}\big(\sqrt{n/m} \|\tilde{X}^j_{0:L}\|_2^2\big)\big), \tag{31}$$

and at the central machine, we can compute

$$S_{\mathrm{I}}(L) = \frac{1}{\sqrt{m'}} \sum_{j \in M_L} \big(2Y^j_{\mathrm{I}}(L) - 1\big).$$

Then we consider the following adaptive test based on Bonferroni's correction:

$$T^{\text{adapt}}_{\mathrm{I}} = \mathbb{1}\Big\{\max_{L \in \mathcal{C}} S_{\mathrm{I}}(L) \geq 2\sqrt{\log\log n}\Big\}.$$

Since for $L \in \mathcal{C}$, it holds that $L \asymp \log(n)$, the above $\sqrt{\log\log n}$ blow up suffices to guarantee that the test has asymptotically vanishing Type I error control, that is, $\mathbb{E}_0 T_{\mathrm{I}}^{\mathrm{adapt}} = o(1)$ by Lemma E.1 in the Supplementary Material (as the random variables $2Y_{\mathrm{I}}^j(L) - 1$ are i.i.d. Rademacher under $\mathbb{P}_0$).

For the Type II error, note that

$$\mathbb{E}_f\big(1 - T_{\mathrm{I}}^{\mathrm{adapt}}\big) \le \mathbb{P}_f\big(S_{\mathrm{I}}(L_s) < 2\sqrt{\log\log n}\big),$$

and aim to apply Lemma A.4. In view of Lemma A.6, (with $\|f\|_2$ replaced by $\|\tilde{f}^{L_s}\|_2$ and $d = \nu_{L_s}$), noting that by triangle inequality $\|\tilde{f}^{L_s}\|_2^2 \ge \|f\|_2^2/2 - 2^{-2L_s s}R^2$ (see also Section B in the Supplementary Material), we get, for $\|f\|_2^2 \ge C_0^2\sqrt{\log\log(n)}\rho_s^2 \ge C_0^2\sqrt{\log\log(n)}\frac{\sqrt{2^{L_s} m \log(n)}}{n\sqrt{b \wedge \log(n)}}$, that for m large enough

$$\eta_{p,m',1} \gtrsim (m'-1)\bigg(\frac{n\|\tilde{f}^{L_s}\|_2^2}{m 2^{L_s/2}} \wedge \frac{1}{2}\bigg)^2 \gtrsim m'\bigg(\bigg(\tilde{C}\frac{\log\log n}{m'}\bigg) \wedge (1/4)\bigg)$$

with $\tilde{C} = C_0^2/2 - R^2$. By the assumption that $bm \gg \log(n)$, m' can be taken larger than arbitrary constant $M_0 > 0$. This means that, in view of Lemma A.4 with $c_{\alpha,n} = 4\log\log n$ and large enough constant C_0 (depending on R), the Type II error is bounded by α.

8.2. *Proof of the upper bound in the public coin, high budget regime.* We use similar arguments as before, applying a Bonferroni-type of correction. First, let us consider the public coin setting, and take a one-to-one mapping ξ_L from $\{1, \dots, \nu_L\}$ to $\{(l, i) : l = 0, \dots, L, i = 0, 1, \dots, 2^l - 1\}$. Let us define the test

$$(32) \qquad \big(Y_{\mathrm{II}}^j(L)\big)_i | U_L = \mathbb{1}\big\{\big(\sqrt{n/m}\,U_L \tilde{X}_{\xi_L(i)}^j\big)_i > 0\big\},$$

where the random variable $U_L \in \mathbb{R}^{\nu_L \times \nu_L}$ is drawn from the Haar measure on the rotation group on $\mathbb{R}^{\nu_L}$. Similarly to before for each L, we take a subset of machines $M_L \subseteq \{1, \dots, m\}$ such that $|M_L| = m' := \frac{m(b \wedge \log(n))}{\log(n)}$, and each machine appears at most in b such sets.

Then machine $j \in M_L$, $L \in \mathcal{C}$, transmits the bits $(Y_{\mathrm{II}}^j(L))_i$, $i = 1, \dots, b' := \frac{mb}{m'|\mathcal{C}|} \wedge \nu_L$ to the central machine, where these local test statistics are aggregated, similarly to (10), as

$$(33) \qquad S_{\mathrm{II}}(L) = \frac{1}{\sqrt{b'm'}}\sum_{i=1}^{b'}\bigg[\bigg(\sum_{j \in M_L}\big[\big(Y_{\mathrm{II}}^j(L)\big)_i - 1/2\big]\bigg)^2 - \frac{m'}{4}\bigg].$$

In view of Lemma E.1, the Type I error of the test

$$T_{\mathrm{II}}^{\mathrm{pub,adapt}} := \mathbb{1}\Big\{\max_{L \in \mathcal{C}} S_{\mathrm{II}}(L) \ge 2\sqrt{\log\log n}\Big\}$$

is $o(1)$. For the Type II error, note that

$$\mathbb{E}_f\big(1 - T_{\mathrm{II}}^{\mathrm{pub,adapt}}\big) \le \mathbb{E}_f\mathbb{1}\big\{S_{\mathrm{II}}(L_s) < 2\sqrt{\log\log n}\big\}.$$

By Lemma E.2, the above display is $o(1)$ whenever $\rho^2 \gtrsim M_n \frac{2^{L_s}}{n\sqrt{\frac{b}{\log(n)} \wedge 2^{L_s}}}$, which, for the choice of $L_s = \lfloor s^{-1}\log(1/\rho_s)\rfloor \vee 1$, yields the rates of Theorem 7.1 and 7.2.

8.3. *Proof of the upper bound in the private coin, high-budget regime.* We proceed by adapting the test T_{III}, provided in Section 4.3, to the nonparametric setting with unknown regularity using again a Bonferroni type correction to achieve adaptation. For simplicity we again apply the map ξ_L, introduced previously, to move between the single and double index notations of the sequence model.

For all $L \in \mathcal{C}$, similarly to the previous cases, we consider a collection of machines M_L with $|M_L| = m' = \frac{m(b\wedge\log(n))}{\log(n)}$, and similarly to Section 4.3, let us use the notation $\mathcal{I}_i(L) \subset M_L$ for the collection of machines corresponding the ith coordinate. We note that, without loss of generality, we can assume that $m' \geq M_\alpha\sqrt{\log\log n}2^{2L_s}/(b')^2$ for some large enough constant M_α; otherwise, the test $T_{\text{I}}^{\text{adapt}}$ above covers the corresponding range. Then we modify the test, given in (12), by increasing the threshold with the Bonferroni correction, that is,

$$T_{\text{III}}^{\text{priv,adapt},1} = \mathbb{1}\Big\{\max_{L\in\mathcal{C}} S^{\text{III},1}(L) \geq 2\sqrt{\log\log n}\Big\}, \quad \text{where}$$

$$S^{\text{III},1}(L) = \left|\frac{1}{|\mathcal{I}_1(L)|2^{L/2}}\sum_{i=1}^{\nu_L}\left(\sum_{j\in\mathcal{I}_i(L)}(Y_i^j - 1/2)\right)^2 - 2^{L/2}/4\right|, Y_i^j|\tilde{X}^j_{\xi_L(i)} = \mathbb{1}_{\tilde{X}^j_{\xi_L(i)}>0}.$$

To deal with large signal components, similarly to (12) (with $d = \nu_L$ and including the Bonferroni correction in the threshold), we propose the test

$$T_{\text{III}}^{\text{priv,adapt},2} = \mathbb{1}\Big\{\max_{L\in\mathcal{C},2\log(L)\leq b} S^{\text{III},2}(L) \geq \kappa_\alpha\sqrt{\log\log n}\Big\}, \quad \text{where}$$

$$S^{\text{III},2}(L) = \left|\frac{1}{dm'C_{b,L}}\left(\sum_{j=1}^{m'}(N^j - C_{b,L}2^{L-1})\right)^2 - \frac{1}{4}\right|$$

with $C_{b,L} = 2^{b-L}$ and N^j given in (13). Finally, we aggregate these tests by taking

$$T_{\text{III}}^{\text{priv,adapt}} = T_{\text{III}}^{\text{priv,adapt},1} \vee T_{\text{III}}^{\text{priv,adapt},2}.$$

In view of the law of Lemma E.1, the Type I error tends to zero for both tests. Therefore, it remained to show that the Type II error is bounded by α. Similarly to the previous cases, note that

$$E_f(1 - T_{\text{III}}^{\text{priv,adapt}}) \leq \mathbb{E}_f\big(\mathbb{1}\{S^{\text{III},1}(L_s) < 2\sqrt{\log\log n}\} \wedge \mathbb{1}\{S^{\text{III},2}(L_s) < 2\sqrt{\log\log n}\}\big).$$

Following the proofs of Lemmas A.8, A.9 and A.10 (with $d = \nu_{L_s}$, f taken to be the ν_{L_s} dimensional vector $\tilde{f}^{L_s}$, b replaced by b' and M_α replaced by $M_0\sqrt{\log\log n}$, for some large enough $M_0 > 0$) and noting that, for $C_0^2 > 4R^2$,

$$\begin{aligned}\|\tilde{f}^{L_s}\|_2^2 &\geq \|f\|_2^2/2 - R^2 2^{-2L_s s} \gtrsim C_0\sqrt{\log\log(n)}\rho_s^2\\ &= \frac{C_0 2^{3L_s/2}\sqrt{\log\log n}}{2n(\frac{b}{\log(n)}\wedge 2^{L_s})} \gtrsim \frac{C_0 2^{L_s}\sqrt{\log\log n}}{nb'\frac{m'}{m}},\end{aligned}$$

and applying Lemmas A.11 and A.4 with $c_{n,\alpha} = 2\sqrt{\log\log n}$, we get that the Type II error of $T_{\text{III}}^{\text{priv,adapt}}$ is bounded from above by $\alpha/2$.

Finally, we combine the above tests by taking

$$T^{\text{priv,adapt}} = T_{\text{III}}^{\text{priv,adapt}} \vee T_{\text{I}}^{\text{priv,adapt}} \quad \text{and} \quad T^{\text{pub,adapt}} = T_{\text{II}}^{\text{pub,adapt}} \vee T_{\text{I}}^{\text{pub,adapt}}.$$

Note that both of the above tests still have vanishing Type I error, while the Type II errors are bounded by the prescribed level α in view of taking the union of the above optimal tests.

9. Proof of the testing lower bound. We provide the details for Steps 1 and 2, as outlined in Section 5. We shall write $\mathscr{L}_\pi(x) = \int \mathscr{L}_f(x)\, d\pi(f)$ with $\mathscr{L}_f(x) := \frac{dP_f}{dP_0}(x)$ and $P_f = \mathbb{P}_f^X$:

Step 1. In view of the Markov chain structure given in (2), the probability measure $d\mathbb{P}_\pi(x,u,y)$ disintegrates as $d\mathbb{P}_K^{Y|(X,U)=(x,u)}\, d\mathbb{P}_f^X(x)\, d\mathbb{P}^U(u)\, d\pi(f)$. Using the Markov chain structure, the first term on the RHS of (16) can be seen to equal

$$\sum_{y \in \mathcal{Y}} \mathbb{P}_0^{Y|U=u}(y) \left(\int \mathscr{L}_\pi(x) \frac{K(y|x,u)}{\mathbb{P}_0^{Y|U=u}(y)}\, dP_0(x) \right)^2 = \mathbb{E}_0^{Y|U=u} \mathbb{E}_0\big[\mathscr{L}_\pi(x)|Y, U=u\big]^2. \tag{34}$$

Decoupling the square in X and using Fubini's theorem, we can write the above display as

$$\int \mathscr{L}_\pi(x_1)\mathscr{L}_\pi(x_2) q_u(x_1,x_2)\, d(P_0 \times P_0)(x_1,x_2), \tag{35}$$

where by independence between the transcripts,

$$q_u(x_1,x_2) := \sum_{y \in \mathcal{Y}} \frac{K(y|x_1,u)K(y|x_2,u)}{\mathbb{P}_0^{Y|U=u}(y)} = \prod_{j=1}^m \left(\sum_{y^j \in \mathcal{Y}^j} \frac{K^j(y^j|x_1^j,u)K^j(y^j|x_2^j,u)}{\mathbb{P}_0^{Y^j|U=u}(y^j)} \right).$$

Note that in the above display, x_i^j and y^j denote the projection of x_i and y on the coordinates indexed by $\{(j-1)d+1, \dots, jd\}$, respectively. In addition, let us denote by $\prod_{j=1}^m q_u^j(x_1^j, x_2^j)$ the RHS of the preceding display. Since K is a Markov kernel, the function $q_u \in L_2(\mathbb{R}^{2dm}, P_0 \times P_0)$ is bounded and nonnegative. Furthermore,

$$\int q_u(x_1,x_2)\, dP_0(x_1) = \sum_{y \in \mathcal{Y}} \frac{K(y|x_2,u)}{\mathbb{P}_0^{Y|U=u}(y)} \int K(y|x_1,u)\, dP_0(x_1) = \sum_{y \in \mathcal{Y}} K(y|x_2,u) = 1,$$

similarly $\int q_u(x_1,x_2)\, dP_0(x_2) = 1$,

$$\int x_i q_u(x_1,x_2)\, d(P_0 \times P_0)(x_1,x_2) = \int x_i\, dP_0(x_i) = 0 \in \mathbb{R}^{md} \tag{36}$$

for $i = 1, 2$, and

$$\int \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \begin{pmatrix} x_1^\top & x_2^\top \end{pmatrix} q_u(x_1,x_2)\, d(P_0 \times P_0)(x_1,x_2) =: \Sigma \in \mathbb{R}^{2md \times 2md}, \tag{37}$$

where the former display can be seen to follow by the law of total expectation, $\Sigma = \operatorname{Diag}(\Sigma^1, \dots, \Sigma^m) \in \mathbb{R}^{2md}$ for

$$\Sigma^j := \begin{pmatrix} \frac{m}{n} I_d & \Xi_u^j \\ \Xi_u^j & \frac{m}{n} I_d, \end{pmatrix}$$

with

$$\Xi_u^j := \mathbb{E}_0^{Y^j|U=u} \mathbb{E}_0\big[X^j|Y, U=u\big] \mathbb{E}_0\big[X^j|Y^j, U=u\big]^\top.$$

Writing $\mathscr{L}_\pi^j := \int \frac{d\mathbb{P}_f^{X^j}}{d\mathbb{P}_0^{X^j}}\, d\pi(f)$, (18) can be seen to equal

$$\prod_{j=1}^m \int \mathscr{L}_\pi^j(x_1^j)\mathscr{L}_\pi^j(x_2^j) q_u^j(x_1^j, x_2^j)\, d(P_0 \times P_0)(x_1^j, x_2^j). \tag{38}$$

Lemma 10.1 below applies to the ratio between (35) and (38) whenever π is chosen to be centered Gaussian. The lemma yields that the aforementioned ratio is maximized when

$q_u(x_1, x_2)\, d(P_0 \times P_0)(x_1, x_2)$ is a Gaussian distribution on $\mathbb{R}^{2md}$ with covariance Σ, where the maximization is among all choices of q_u such that q_u is nonnegative, bounded and satisfying (36)–(37). Deliberation and proof of the lemma is deferred to Section 10 and the Supplementary Material to the article. For π a centered Gaussian distribution on $\mathbb{R}^d$, the above lemma applies with $k = 2d$, $\sigma^2 = m/n$, we obtain that the ratio between (35) and (38) is bounded above by

$$\frac{\int \mathscr{L}_\pi(x_1)\mathscr{L}_\pi(x_2)\, dN(0, \Sigma)(x_1, x_2)}{\prod_{j=1}^m \int \mathscr{L}_\pi^j(x_1^j)\mathscr{L}_\pi^j(x_2^j)\, dN(0, \Sigma^j)(x_1^j, x_2^j)}. \tag{39}$$

Combining the result of the lemma with the bound

$$\prod_{j=1}^m \mathbb{E}_0^{Y^j|U=u} \mathbb{E}_0[\mathscr{L}_\pi(X^j)|Y^j, U = u]^2 \leq \prod_{j=1}^m \mathbb{E}_0^{X^j|U=u}[\mathscr{L}_\pi(X^j)^2] \tag{40}$$

following from Jensen's inequality, we obtain that

$$\begin{aligned} \mathbb{E}_0^{Y|U=u}\left(\frac{\mathbb{P}_\pi^{Y|U=u}}{\mathbb{P}_0^{Y|U=u}}(Y)\right)^2 \leq{}& \frac{\int \mathscr{L}_\pi(x_1)\mathscr{L}_\pi(x_2)\, dN(0, \Sigma)(x_1, x_2)}{\prod_{j=1}^m \int \mathscr{L}_\pi^j(x_1^j)\mathscr{L}_\pi^j(x_2^j)\, dN(0, \Sigma)(x_1, x_2)} \\ &\times \prod_{j=1}^m \mathbb{E}_0^{X^j|U=u}[\mathscr{L}_\pi(X^j)^2]. \end{aligned} \tag{41}$$

By the block diagonal matrix structure of Σ, the denominator in the first factor of the RHS of (41) satisfies

$$\begin{aligned} \prod_{j=1}^m \int \mathscr{L}_\pi^j(x_1^j)\mathscr{L}_\pi^j(x_2^j)\, dN(0, \Sigma)(x_1, x_2) &= \prod_{j=1}^m \int e^{\frac{n}{2m}(\frac{n}{m}\|\sqrt{\Sigma^j}(f,g)\|_2^2 - \|(f,g)\|_2^2)}\, d(\pi \times \pi)(f, g) \\ &= \prod_{j=1}^m \int e^{\frac{n^2}{m^2} f^\top \Xi_u^j g}\, d(\pi \times \pi)(f, g) \\ &\geq \prod_{j=1}^m e^{\frac{n^2}{m^2}\int f^\top \Xi_u^j g\, d(\pi\times\pi)(f,g)} = 1. \end{aligned}$$

Similarly, the numerator is equal to

$$\int \mathscr{L}_\pi(x_1)\mathscr{L}_\pi(x_2)\, dN(0, \Sigma)(x_1, x_2) = \int e^{\frac{n^2}{m^2} f^\top \sum_{j=1}^m \Xi_u^j g}\, d(\pi \times \pi)(f, g).$$

Combining the above displays (i.e., (16) and the last three displays), we obtain that

$$\begin{aligned} D_{\chi^2}(\mathbb{P}_{0,K}^{Y|U=u}; \mathbb{P}_{\pi,K}^{Y|U=u}) \leq{}& \prod_{j=1}^m \mathbb{E}_0^{X^j|U=u}[\mathscr{L}_\pi(X^j)^2] \\ &\cdot \int e^{\frac{n^2}{m^2} f^\top \sum_{j=1}^m \Xi_u^j g}\, d(\pi \times \pi)(f, g) - 1. \end{aligned} \tag{42}$$

Step 2. What is left to show in this step is that, for $\pi = N(0, \Gamma)$, $\Gamma \in \mathbb{R}^{d\times d}$ can be chosen such that the RHS of the previous display is small enough, while also ensuring that $\pi(H_\rho^c)$ is controlled whenever ρ^2 satisfies (4)–(5) for c_α depending only on $\alpha \in (0, 1)$.

For a given $c_\alpha > 0$, set $\epsilon := \frac{\rho}{c_\alpha^{1/4} d^{1/2}}$ and $\Gamma := \epsilon^2 \bar{\Gamma}$ for some $\bar{\Gamma} \in \mathbb{R}^{d\times d}$ to be specified later, separately for the private and public coin protocols. The remaining mass $\pi(H_\rho)$ can now be seen to equal

$$\pi(f : \|f\|_2^2 \leq \rho^2) = \Pr(Z^\top \bar{\Gamma} Z \leq \sqrt{c_\alpha} d),$$

where Z is a d-dimensional standard normal vector. If $\bar{\Gamma}$ is symmetric, idempotent and has rank (proportional to) d, the concentration inequality in Lemma A.13 yields that the probability on the RHS of the above display can be made arbitrarily small for small enough choice of $c_\alpha > 0$.

We now proceed to bound the first factor in the product on the RHS of (42), which for a positive semidefinite choice of $\bar{\Gamma}$ equals

$$\prod_{j=1}^{m} \int \mathbb{E}_0^{X^j|U=u} \exp\biggl(\frac{n}{m}\bigl(\sqrt{\bar{\Gamma}}(f+g)\bigr)^\top X^j - \frac{n}{2m}\|\sqrt{\bar{\Gamma}}f\|_2^2 - \frac{n}{2m}\|\sqrt{\bar{\Gamma}}g\|_2^2\biggr)\, dN\bigl(0, \epsilon^2 I_{2d}\bigr)(f,g).$$

By direct computation the latter display equals

$$\prod_{j=1}^{m} \int \exp\biggl(\frac{n\epsilon^2}{m} z^\top \bar{\Gamma} z'\biggr)\, dN(0, I_{2d})\bigl(z, z'\bigr).$$

By applying the moment generating function of the Gaussian chaos, for example, Lemma 6.2.2 in [34] to the above display and using that ρ^2 satisfies (4) or (5), we obtain that, for $\frac{n\epsilon^2}{m}\|\bar{\Gamma}\| \lesssim \frac{n\rho^2}{c_\alpha^{1/2} m\sqrt{d}} \le \sqrt{c_\alpha/m} \le \sqrt{c_\alpha}$ small enough, where $\|\cdot\|$ denotes the spectral norm of a matrix, there exists a constant $C \ge \|\bar{\Gamma}\|^2/d$ such that

$$\prod_{j=1}^{m} \mathbb{E}_0^{X^j|U=u}\bigl[\mathscr{L}_\pi\bigl(X^j\bigr)^2\bigr] \le \exp\biggl(C c_\alpha^{-1} \frac{n^2\rho^4}{md}\biggr) \le \exp(C c_\alpha). \tag{43}$$

The exponent can be made arbitrarily close to zero per choice of $c_\alpha > 0$.

We switch our attention now to the second factor in the product on the RHS of (42), which we bound by applying Lemma 6.2.2 in [34] once more,

$$\int e^{\frac{n^2}{m^2}(\sqrt{\Gamma}f)^\top \sum_{j=1}^m \Xi_u^j(\sqrt{\Gamma}g)}\, dN\bigl(0, \epsilon^2 I_{2d}\bigr)(f,g) \le e^{C\frac{n^4\epsilon^4}{m^4}\operatorname{Tr}((\sqrt{\bar{\Gamma}}^\top \Xi_u \sqrt{\bar{\Gamma}})^2)}, \tag{44}$$

whenever

$$\frac{n^2\epsilon^2}{m^2}\bigl\|\sqrt{\bar{\Gamma}}^\top \Xi_u \sqrt{\bar{\Gamma}}\bigr\| \tag{45}$$

is small enough.

It remains to choose a symmetric, idempotent positive semidefinite $\bar{\Gamma}$ that sufficiently bounds (45) and to combine the above displays providing the stated lower bound for the testing risk. For the exact choice of $\bar{\Gamma}$, we distinguish between the public coin and private coin cases. In both cases we employ the data-processing inequalities of Lemma A.3 and Lemma A.2, which yield that

$$\operatorname{Tr}(\Xi_u) = \sum_{j=1}^{m} \operatorname{Tr}\bigl(\Xi_u^j\bigr) \le \min\biggl\{2\log 2 \cdot \frac{b}{d}, 1\biggr\}\frac{m^2 d}{n}. \tag{46}$$

The public coin case: In this case it suffices to take $\bar{\Gamma} = I_d$, which is trivially symmetric, idempotent and positive semidefinite. By Lemma A.2, $\frac{n}{m}\Xi_u^j \le I_d$, so (45) holds as well for this choice of $\bar{\Gamma}$,

$$\frac{n^2\epsilon^2}{m^2}\|\Xi_u\| \le n\epsilon^2 \le \frac{n\rho^2}{\sqrt{c_\alpha} d} \le \sqrt{c_\alpha},$$

where the second to last inequality holds for ρ^2 satisfying (4).

It remains to combine our results and provide a lower bound for the testing risk. Note that

$$\mathrm{Tr}(\Xi_u^2) = \|\Xi_u\| \,\mathrm{Tr}(\Xi_u) \le \frac{m^2}{n}\,\mathrm{Tr}(\Xi_u) \lesssim \frac{(b\wedge d)m^4}{n^2},$$

where the last inequality follows from (46). Combining the above bound with assertions (44), (43), (42), (16) and (15), $\epsilon^4 = c_\alpha^{-1} d^{-2}\rho^4$ and the fact that $\pi(H_\rho) \le \alpha/2$, we obtain that

$$\inf_{T\in\mathcal{T}_{\text{pub}}(b)} \mathcal{R}(H_\rho, T) \ge 1 - \sqrt{2\left(e^{C(\frac{n^2\rho^4}{c_\alpha md} + \frac{n^2\rho^4(b\wedge d)}{c_\alpha d^2})} - 1\right)} - \pi\left(H_\rho^c\right)$$

$$\ge 1 - \sqrt{2\left(e^{2Cc_\alpha} - 1\right)} - \alpha/2 > 1-\alpha,$$

whenever ρ^2 satisfies (4) for $c_\alpha > 0$ small enough. This finishes the proof for the public coin case.

The private coin case: Since without loss of generality we can assume that U is degenerate in the private coin case, $\Xi_u = \Xi$ for $\mathbb{P}^U$-almost every u. The matrix Ξ is positive definite and symmetric; therefore, it possesses a spectral decomposition $V^\top \mathrm{Diag}(\xi_1, \dots, \xi_d)V$. Without loss of generality, assume that $\xi_1 \ge \xi_2 \ge \dots \ge \xi_d$ with corresponding eigenvectors $V = (v_1 \quad \dots \quad v_d)$. Let $\check{V}$ denote the $d \times \lceil d/2 \rceil$ matrix $(v_{\lfloor d/2\rfloor+1} \quad \dots \quad v_d)$. The choice of prior may depend on Ξ; to see this, note the order of the supremum and infimum in (15) and the fact that Ξ soley depends on the choice of kernel. To that extent, set $\bar{\Gamma} = \check{V}\check{V}^\top$. It holds that

$$\mathrm{Tr}(\check{V}\check{V}^\top) = \sum_{i=1}^{d} \sum_{k=\lfloor d/2\rfloor+1}^{d} (v_k)_i^2 = \lceil d/2 \rceil.$$

The choice $\Gamma = \epsilon^2\bar{\Gamma}$ is thus seen to satisfy the conditions of symmetry and positive definiteness and is idempotent with rank $\lceil d/2\rceil$.

Since the eigenvalues are decreasingly ordered,

$$\xi_{\lfloor d/2\rfloor} \le \frac{2}{d}\sum_{i=1}^{\lfloor d/2\rfloor} \xi_i \le \frac{2}{d}\,\mathrm{Tr}(\Xi).$$

By orthogonality of the columns of V, $\check{V}^\top \Xi \check{V} = \mathrm{Diag}(\xi_{\lfloor d/2\rfloor+1}, \dots, \xi_d)$. Combining this inequality with the last display and assertion (46), we get that, for ρ^2 satisfying (5), the term (45) can be made arbitrarily small for small enough choice of c_α, that is,

$$\frac{n^2\epsilon^2}{m^2}\left\|\sqrt{\bar{\Gamma}}^\top \Xi_u \sqrt{\bar{\Gamma}}\right\| \le \frac{n^2\epsilon^2}{m^2}\xi_{\lfloor d/2\rfloor} \le 2\frac{n^2\rho^2}{\sqrt{c_\alpha}d^2m^2}\,\mathrm{Tr}(\Xi)$$

$$\le (4\log 2)\frac{n\rho^2(b\wedge d)}{\sqrt{c_\alpha}d^2} \le (4\log 2)\sqrt{c_\alpha/d}.$$

Finally, a similar argument will be used to bound the RHS of (44) and finally to provide a lower bound for the testing risk. Note that

$$\mathrm{Tr}\big((\sqrt{\bar{\Gamma}}^\top \Xi_u \sqrt{\bar{\Gamma}})^2\big) = \mathrm{Tr}\big((\check{V}^\top \Xi \check{V})^2\big) = \sum_{i=\lfloor d/2\rfloor+1}^{d} \xi_i^2 \le d\xi_{\lfloor d/2\rfloor}^2 \le \frac{4}{d}\,\mathrm{Tr}(\Xi)^2,$$

which implies in turn that

$$\frac{n^4\epsilon^4}{m^4}\,\mathrm{Tr}\big((\check{V}^\top \Xi \check{V})^2\big) \le 4\frac{n^4\rho^4}{c_\alpha m^4 d^3}\,\mathrm{Tr}(\Xi)^2 \le 4\frac{n^2\rho^4(b\wedge d)^2}{c_\alpha d^3},$$

where the last inequality follows from (46). Consequently, we have obtained that

$$\inf_{T \in \mathcal{T}_{\text{priv}}(b)} \mathcal{R}(H_\rho, T) \geq 1 - \sqrt{2\big(e^{C(\frac{n^2\rho^4}{c_\alpha m d} + \frac{n^2\rho^4(b\wedge d)^2}{c_\alpha d^3})} - 1\big)} - \pi\big(H_\rho^c\big)$$
$$\geq 1 - \sqrt{2\big(e^{2Cc_\alpha} - 1\big)} - \alpha/2 > 1 - \alpha$$

for ρ^2 satisfying (5) and $c_\alpha > 0$ small enough.

10. Lemma 10.1: Gaussian maximization. Before giving the detailed statement of the lemma below, we briefly contemplate on its aim and proof. The lemma bears a close connection to Brascamp–Lieb inequalities [8, 10, 21]. Brascamp–Lieb-type inequalities have appeared in context of information theory in the literature before; see, for example, [15, 22], where Gaussian extremality is established for certain information theoretic optimization problems. Instead of the information theoretic entropy based route, we rely on the technique of [21]. The resulting lemma allows us to bound the ratio between (35) and (38), that is,

$$\frac{\int \mathscr{L}_\pi(x_1)\mathscr{L}_\pi(x_2) q_u(x_1, x_2)\, d(P_0 \times P_0)(x_1, x_2)}{\prod_{j=1}^m \int \mathscr{L}_\pi^j(x_1^j)\mathscr{L}_\pi^j(x_2^j) q_u^j(x_1^j, x_2^j)\, d(P_0 \times P_0)(x_1^j, x_2^j)}, \tag{47}$$

by (39), that is, a Gaussian distribution with matching mean and covariance. Consequently, we obtain a quadratic form in the covariance that we would otherwise obtain via a Taylor expansion. That such a quadratic form does not follow through more standard means such as Taylor expansion is described in [3], Section 4.

The proof of the lemma exploits the conjugacy between likelihood of the observation X and the Gaussian prior on the parameter to obtain that a Gaussian distribution is, in fact, an extremal case. For reasons of space, we defer the proof to Section F of the Supplementary Material.

LEMMA 10.1. *For $x \in \mathbb{R}^{mk}$, let $x^j \in \mathbb{R}^k$, $j = 1, \dots, m$ denote the projection of x on the coordinates $\{(j-1)k + 1, \dots, jk\}$. Let $\Lambda \in \mathbb{R}^{k\times k}$ a positive definite symmetric matrix and $\Lambda^{\otimes m} = \mathrm{Diag}(\Lambda, \dots, \Lambda) \in \mathbb{R}^{mk\times mk}$. For $h \in \mathbb{R}^k$, let p_h denote the density of a $N(h, \Lambda)$ distribution with respect to the Lebesgue measure on $\mathbb{R}^k$, and let $p_h^m(x) := \prod_{j=1}^m p_h(x^j)$. Consider for some $M > 0$, $\mathcal{Q} \equiv \mathcal{Q}(M, \Sigma)$ the class of all nonnegative functions $q \in L_\infty(\mathbb{R}^{mk})$ satisfying $\frac{q(x)}{\int q(x) p_0^m(x)\,dx} \leq M$ P_0^m-a.e., $\int x q(x) p_0^m(x)\,dx = 0$ and $\int x x^\top q(x) p_0^m(x)\,dx = \Sigma$. Furthermore, let H a $N(0, \Upsilon)$-distributed random vector in $\mathbb{R}^k$. Then*

$$\sup_{q\in\mathcal{Q}} \frac{\int \mathbb{E}^H \prod_{j=1}^m \frac{p_H}{p_0}(x^j) q(x) p_0^m(x)\,dx}{\int \prod_{j=1}^m \mathbb{E}^H \frac{p_H}{p_0}(x^j) q(x) p_0^m(x)\,dx} \leq \frac{\int \mathbb{E}^H \prod_{j=1}^m \frac{p_H}{p_0}(x^j)\, dN(0,\Sigma)(x)}{\int \prod_{j=1}^m \mathbb{E}^H \frac{p_H}{p_0}(x^j)\, dN(0,\Sigma)(x)}.$$

Acknowledgments. We would like to thank Elliot H. Lieb for a helpful comment regarding the proof of Lemma 10.1.

Funding. Cofunded by the European Union (ERC, BigBayesUQ, project number: 101041064). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

SUPPLEMENTARY MATERIAL

Supplementary Material to Optimal high-dimensional and nonparametric distributed testing under communication constraints (DOI: 10.1214/23-AOS2269SUPP; .pdf). In the supplement to this paper [29], we present the detailed proofs for the main theorems in the paper "Optimal high-dimensional and nonparametric distributed testing under communication constraints".

REFERENCES

[1] ACHARYA, J., CANONNE, C. L. and TYAGI, H. (2020). Inference under information constraints I: Lower bounds from chi-square contraction. *IEEE Trans. Inf. Theory* **66** 7835–7855. MR4191030 https://doi.org/10.1109/TIT.2020.3028440

[2] ACHARYA, J., CANONNE, C. L. and TYAGI, H. (2020). Inference under information constraints II: Communication constraints and shared randomness. *IEEE Trans. Inf. Theory* **66** 7856–7877. MR4191031 https://doi.org/10.1109/TIT.2020.3028439

[3] ACHARYA, J., CANONNE, C. L. and TYAGI, H. (2020). Distributed signal detection under communication constraints. In *Proceedings of Thirty Third Conference on Learning Theory* (J. Abernethy and S. Agarwal, eds.). *Proceedings of Machine Learning Research* **125** 41–63. PMLR.

[4] AHLSWEDE, R. and CSISZÁR, I. (1986). Hypothesis testing with communication constraints. *IEEE Trans. Inf. Theory* **32** 533–542. MR0852123 https://doi.org/10.1109/TIT.1986.1057194

[5] ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices*. *Cambridge Studies in Advanced Mathematics* **118**. Cambridge Univ. Press, Cambridge. MR2760897

[6] BALAKRISHNAN, S. and WASSERMAN, L. (2019). Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *Ann. Statist.* **47** 1893–1927. MR3953439 https://doi.org/10.1214/18-AOS1729

[7] BARNES, L. P., HAN, Y. and ÖZGÜR, A. (2020). Lower bounds for learning distributions under communication constraints via Fisher information. *J. Mach. Learn. Res.* **21** Paper No. 236, 30. MR4209522

[8] BENNETT, J., CARBERY, A., CHRIST, M. and TAO, T. (2008). The Brascamp–Lieb inequalities: Finiteness, structure and extremals. *Geom. Funct. Anal.* **17** 1343–1415. MR2377493 https://doi.org/10.1007/s00039-007-0619-6

[9] BERGER, T. and ZHANG, Z. (1994). On the CEO problem. In *Proceedings of* 1994 *IEEE International Symposium on Information Theory* 201. https://doi.org/10.1109/ISIT.1994.394767

[10] BRASCAMP, H. J. and LIEB, E. H. (1976). Best constants in Young's inequality, its converse, and its generalization to more than three functions. *Adv. Math.* **20** 151–173. MR0412366 https://doi.org/10.1016/0001-8708(76)90184-5

[11] BRAVERMAN, M., GARG, A., MA, T., NGUYEN, H. L. and WOODRUFF, D. P. (2016). Communication lower bounds for statistical estimation problems via a distributed data processing inequality. In *STOC'*16—*Proceedings of the* 48*th Annual ACM SIGACT Symposium on Theory of Computing* 1011–1020. ACM, New York. MR3536632 https://doi.org/10.1145/2897518.2897582

[12] CAI, T. T. and WEI, H. (2020). Distributed Gaussian mean estimation under communication constraints: Optimal rates and communication-efficient algorithms. cs, math, stat, available at arXiv:2001.08877.

[13] CAI, T. T. and WEI, H. (2022). Distributed adaptive Gaussian mean estimation with unknown variance: interactive protocol helps adaptation. arXiv preprint. Available at arXiv:2001.08877.

[14] CAI, T. T. and WEI, H. (2022). Distributed nonparametric function estimation: Optimal rate of convergence and cost of adaptation. *Ann. Statist.* **50** 698–725. MR4404917 https://doi.org/10.1214/21-aos2124

[15] CARLEN, E. A. and CORDERO-ERAUSQUIN, D. (2009). Subadditivity of the entropy and its relation to Brascamp–Lieb type inequalities. *Geom. Funct. Anal.* **19** 373–405. MR2545242 https://doi.org/10.1007/s00039-009-0001-y

[16] DUCHI, J. C., JORDAN, M. I., WAINWRIGHT, M. J. and ZHANG, Y. (2014). Optimality guarantees for distributed statistical estimation. cs, math, stat, available at arXiv:1405.0782.

[17] HAN, T. S. and AMARI, S. (1998). Statistical inference under multiterminal data compression. *IEEE Trans. Inf. Theory* **44** 2300–2324. MR1658791 https://doi.org/10.1109/18.720540

[18] HAN, Y., ÖZGÜR, A. and WEISSMAN, T. (2018). Geometric lower bounds for distributed parameter estimation under communication constraints. In *Conference on Learning Theory* 3163–3188. PMLR.

[19] INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. *Lecture Notes in Statistics* **169**. Springer, New York. MR1991446 https://doi.org/10.1007/978-0-387-21580-8

[20] KREIDL, O. P., TSITSIKLIS, J. N. and ZOUMPOULIS, S. I. (2011). On decentralized detection with partial information sharing among sensors. *IEEE Trans. Signal Process.* **59** 1759–1765. https://doi.org/10.1109/TSP.2010.2099223

[21] LIEB, E. H. (1990). Gaussian kernels have only Gaussian maximizers. *Invent. Math.* **102** 179–208. MR1069246 https://doi.org/10.1007/BF01233426

[22] LIU, J., COURTADE, T. A., CUFF, P. W. and VERDÚ, S. (2016). Brascamp–Lieb inequality and its reverse: An information theoretic view. In 2016 *IEEE International Symposium on Information Theory* (*ISIT*) 1048–1052. IEEE, Barcelona, Spain. https://doi.org/10.1109/ISIT.2016.7541459

[23] RAO, A. and YEHUDAYOFF, A. (2020). *Communication Complexity: And Applications*. Cambridge Univ. Press, Cambridge.

[24] SHAMIR, O. (2014). Fundamental limits of online and distributed algorithms for statistical learning and estimation. *Adv. Neural Inf. Process. Syst.* **27** 163–171.

[25] SPOKOINY, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498. MR1425962 https://doi.org/10.1214/aos/1032181163

[26] SZABÓ, B. and VAN ZANTEN, H. (2020). Adaptive distributed methods under communication constraints. *Ann. Statist.* **48** 2347–2380. MR4134798 https://doi.org/10.1214/19-AOS1890

[27] SZABÓ, B. and VAN ZANTEN, H. (2022). Distributed function estimation: adaptation using minimal communication. arXiv preprint. Available at arXiv:2003.12838.

[28] SZABÓ, B., VUURSTEEN, L. and VAN ZANTEN, H. (2022). Optimal distributed composite testing in high-dimensional Gaussian models with 1-bit communication. *IEEE Trans. Inf. Theory* **68** 4070–4084. MR4433269

[29] SZABÓ, B., VUURSTEEN, L. and VAN ZANTEN, H. (2023). Supplement to "Optimal high-dimensional and nonparametric distributed testing under communication constraints." https://doi.org/10.1214/23-AOS2269SUPP

[30] SZABÓ, B. and ZAMAN, A. (2022). Distributed nonparametric estimation under communication constraints. Preprint.

[31] TARIGHATI, A., GROSS, J. and JALDÉN, J. (2017). Decentralized hypothesis testing in energy harvesting wireless sensor networks. *IEEE Trans. Signal Process.* **65** 4862–4873. MR3689096 https://doi.org/10.1109/TSP.2017.2716909

[32] TENNEY, R. R. and SANDELL, N. R. JR. (1981). Detection with distributed sensors. *IEEE Trans. Aerosp. Electron. Syst.* **17** 501–510. MR0624526 https://doi.org/10.1109/TAES.1981.309178

[33] TSITSIKLIS, J. N. (1988). Decentralized detection by a large number of sensors. *Math. Control Signals Systems* **1** 167–182. MR0936332 https://doi.org/10.1007/BF02551407

[34] VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. With a foreword by Sara van de Geer. MR3837109 https://doi.org/10.1017/9781108231596

[35] XU, A. and RAGINSKY, M. (2017). Information-theoretic lower bounds on Bayes risk in decentralized estimation. *IEEE Trans. Inf. Theory* **63** 1580–1600. MR3625982 https://doi.org/10.1109/TIT.2016.2646342

[36] YAO, A. C.-C. (1979). Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing. STOC'*79 209–213. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/800135.804414

[37] ZHANG, Y., DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Information-theoretic lower bounds for distributed statistical estimation with communication constraints. In *NIPS* 2328–2336. Citeseer.

[38] ZHU, Y. and LAFFERTY, J. (2018). Distributed nonparametric regression under communication constraints. In *Proceedings of the* 35*th International Conference on Machine Learning* (J. Dy and A. Krause, eds.). *Proceedings of Machine Learning Research* **80** 6009–6017. PMLR.