FEDERATED NONPARAMETRIC HYPOTHESIS TESTING WITH DIFFERENTIAL PRIVACY CONSTRAINTS: OPTIMAL RATES AND ADAPTIVE TESTS

By T. Tony Cai[1,a], Abhinav Chakraborty[1,b] and Lasse Vuursteen[1,c]

[1]*Department of Statistics and Data Science, University of Pennsylvania,* [a]*tcai@wharton.upenn.edu;* [b]*abch@wharton.upenn.edu;* [c]*lassev@wharton.upenn.edu*

Federated learning has attracted significant recent attention due to its applicability across a wide range of settings where data is collected and analyzed across disparate locations. In this paper, we study federated nonparametric goodness-of-fit testing in the white-noise-with-drift model under distributed differential privacy (DP) constraints.

We first establish matching lower and upper bounds, up to a logarithmic factor, on the minimax separation rate. This optimal rate serves as a benchmark for the difficulty of the testing problem, factoring in model characteristics such as the number of observations, noise level, and regularity of the signal class, along with the strictness of the (ϵ,δ)-DP requirement. The results demonstrate interesting and novel phase transition phenomena. Furthermore, the results reveal an interesting phenomenon that distributed one-shot protocols with access to shared randomness outperform those without access to shared randomness. We also construct a data-driven testing procedure that possesses the ability to adapt to an unknown regularity parameter over a large collection of function classes with minimal additional cost, all while maintaining adherence to the same set of DP constraints.

Keywords: Distributed computation; Differential privacy; Federated learning; Nonparametric goodness-of-fit testing,

1. Introduction.

Federated learning is a collaborative distributed machine learning technique designed to address data governance and privacy concerns. It facilitates organizations or groups to collectively train a shared global model without the need to expose raw data externally. Federated learning has garnered increasing attention due to its applicability across a wide range of settings where data is collected and analyzed across disparate locations. This includes scenarios like medical data dispersed among different hospitals, financial customer data stored across various branches or databases, and the utilization of federated learning within user networks in modern technologies such as smartphones or self-driving cars. See, for example, [59, 52, 44, 13, 68]. In such contexts, privacy concerns often impede direct data pooling, making the development of efficient statistical inference methods that preserve privacy and harness the collective power of distributed data essential.

In this paper, we investigate federated nonparametric goodness-of-fit testing under distributed differential privacy (DP) constraints. Nonparametric hypothesis testing, a fundamental statistical problem, has been extensively studied in conventional settings, with a rich body of classical literature examining theoretically optimal performance. DP, introduced by Dwork et al. (2006), serves as a mathematical guarantee determining whether results or datasets can be deemed “privacy-preserving" and thus openly published. Many differentially private statistical methods have since been developed. See, for example, [11, 37, 38]. While several

MSC2020 subject classifications: Primary 62G10, 62C20; secondary 68P27.

Keywords and phrases: Federated learning, Differential privacy, Nonparametric goodness-of-fit testing, Distributed inference.

other privacy frameworks exist, DP holds a prominent position both theoretically and practically, finding application within industry giants like Google [40], Microsoft [32], Apple [78], as well as governmental entities such as the US Census Bureau [70].

DP may significantly impact the quality of statistical inference, particularly in testing problems where it may diminish statistical power that could be obtained with complete data availability. In the present paper, we quantify the cost of privacy in the canonical nonparametric goodness-of-fit testing setting, delineating the theoretical limits of performance achievable under DP constraints and developing methods that attain optimal performance. Contrasting our results with those derived for federated nonparametric estimation under DP [20], we observe that the differences persist and are exacerbated under privacy constraints, highlighting the unique challenges posed by privacy preservation in both testing and estimation contexts.

We first establish the minimax separation rate for the nonparametric goodness-of-fit testing problem and construct optimal tests in the oracle setting where the regularity parameters are known. The minimax separation rate serves as a benchmark for the difficulty of the testing problem. However, the regularity parameters are typically in applications. A natural question is: Without the knowledge of the regularity parameters, is it possible to construct a test that is as good as when the parameters are known? This is a question about *adaptation*, which has been a major goal in nonparametric statistics. We construct a data-driven test and show that the proposed method can adapt to unknown regularity parameters with minimal additional cost while adhering the same DP restrictions.

1.1. *Federated privacy-constrained testing.* We begin by formally introducing the general framework of federated inference under distributed DP constraints. Consider a family of probability measures $\{P_f\}_{f\in\mathcal{F}}$ on the measurable space $(\mathcal{X},\mathscr{X})$, parameterized by $f\in\mathcal{F}$. We consider a setting where $N=mn$ i.i.d. observations are drawn from a distribution P_f and distributed across m servers. Each server $j=1,\dots,m$ holding an equal amount (n many) observations.

Let us denote by $X^{(j)}=(X_i^{(j)})_{i=1}^n$ the n realizations from P_f on the j-th server. Based on $X^{(j)}$, each server outputs a (randomized) transcript $Y^{(j)}$ to the central server that satisfies the privacy constraint. The central server, utilizing all transcripts $Y:=(Y^{(1)},\dots,Y^{(m)})$, decides between a null hypothesis and an alternative hypothesis, through means of a test $T=T(Y)$. Since we are concerned with testing between a null and alternative hypothesis, we shall consider the decision space $\{0,1\}$, where 0 corresponds to DO NOT REJECT and 1 with REJECT. A *test* is then simply to be understood as a statistic taking values in $\{0,1\}$. Figure 1 gives an illustration of a federated (ϵ,δ)-DP-constrained testing procedure.

The transcript $Y^{(j)}$ satisfies an (ϵ,δ)-DP constraint, which, loosely speaking, means that the transcript $Y^{(j)}$ cannot differ too much depending on whether a specific individual is in the data set or not. This is achieved through randomization, which is independent of the data. We will consider two types of sources for randomization; independently among the servers or through a shared source of randomness U (e.g., the same random seed). Formally, shared source of randomness means that the law of the transcript is given by a distribution conditionally on $X^{(j)}$ and U, $A\mapsto\mathbb{P}(Y^{(j)}\in A|X^{(j)},U)$, defined on a measurable space $(\mathcal{Y},\mathscr{Y})$. The presence of shared randomness is a slight, but important extension of the distributed protocols where $Y^{(j)}$ is allowed to be random only through locally generated randomness. See further discussion below and in Section 2. To assure that the source of shared randomness does not erode the notion of privacy, only the local source of randomness is used in the privacy mechanism, i.e. to guarantee privacy. We formalize this as follows.

We shall call two data sets $x,x'\in\mathcal{X}^n$ *neighboring* if they differ in one individual datum. That is, they are at most one apart in *Hamming distance* (see Section 1.6). A DP constraint is then defined as follows.

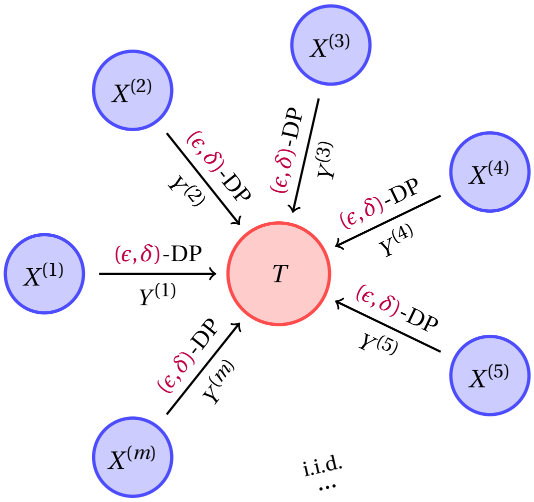

FIGURE 1. *Illustration of federated (ϵ,δ)-DP-constrained testing.*

DEFINITION 1. *The transcript $Y^{(j)}$ is (ϵ,δ)-differentially private ((ϵ,δ)-DP) if for all $A\in\mathscr{Y}^{(j)}$, $u\in\mathcal{U}$ and neighboring data sets $x,x'\in\mathcal{X}^n$ differing in one individual datum it holds that*

$$\mathbb{P}\left(Y^{(j)}\in A|X^{(j)}=x,U=u\right)\leqslant e^{\epsilon}\mathbb{P}\left(Y^{(j)}\in A|X^{(j)}=x',U=u\right)+\delta. \tag{1}$$

The above setting is concerned with distributed protocols for scenarios where multiple parties hold sensitive data and each publishes a differentially private summary, without sharing raw data between them. This approach is common in cases like separate studies by different hospitals on the same population, where privacy concerns prevent direct data pooling. We note that, in the above definition, the outcome of the shared source of randomness U does not diminish the privacy guarantee, even when it is publicly available. This is because the shared randomness is not used in the privacy mechanism, but as a means to enhance coordination between the transcripts, which allows the transcripts to be more informative about the underlying signal whilst each transcript is effectively sharing less information about their underlying individual data.

Formally, having access to shared randomness means that U can be defined on some probability space $(\mathcal{U},\mathscr{U},\mathbb{P}^U)$, such that U is independent of the data (i.e. by taking the appropriate product space for $\mathbb{P}^{X,U}$). Having no access to shared randomness effectively corresponds to considering U to be a degenerate random variable, or $\mathscr{U}=\{\varnothing,\mathcal{U}\}$. In order to stream line the notation between these two setups, we shall refer to the triplet $(T,\{\left(\mathbb{P}^{Y^{(j)}|X^{(j)}=x,U=u}\right)_{x\in\mathcal{X}^n,u\in\mathcal{U}}\}_{j=1}^m,(\mathcal{U},\mathscr{U},\mathbb{P}^U))$ as a (ϵ,δ)*-DP shared randomness distributed testing protocol* with $\{\left(\mathbb{P}^{Y^{(j)}|X^{(j)}=x,U=u}\right)_{x\in\mathcal{X}^n,u\in\mathcal{U}}\}_{j=1}^m$ satisfying Definition 1 for general U (i.e. general $(\mathcal{U},\mathscr{U},\mathbb{P}^U)$). The class of such triplets, but with $\mathscr{U}=\{\varnothing,\mathcal{U}\}$, shall be referred to as (ϵ,δ)*-DP local randomness distributed testing protocols*. We shall denote these classes as $\mathscr{T}^{(\epsilon,\delta)}_{\text{SHR}}$ and $\mathscr{T}^{(\epsilon,\delta)}_{\text{LR}}$, respectively, and note that the former is a superset of the latter.

1.2. *Problem formulation.* The white-noise-with-drift model serves as a benchmark model for nonparametric testing and has been extensively studied outside of the DP setting, see [41, 46, 56, 75, 45]. Furthermore, the problem bares a close relationship with "classical" nonparametric goodness-of-fit testing in the sense of [10, 73, 30, 84] and other nonparametric testing problems through asymptotic equivalence, see Section 1.4 in [47] and references therein.

In the distributed setting, the $j = 1, \dots, m$ machines each observe $i = 1, \dots, n$ i.i.d, $X_i^{(j)}$ taking values in $\mathcal{X} \subset L_2[0,1]$ and subject to the stochastic differential equation

$$dX_{t;i}^{(j)} = f(t)dt + \sigma dW_{t;i}^{(j)} \tag{2}$$

under P_f, with $t \mapsto W_{t;i}^{(1)}, \dots, t \mapsto W_{t;i}^{(m)}$ i.i.d. Brownian motions and $f \in L_2[0,1]$ for $i = 1, \dots, n$, with $\sigma > 0$ the known noise level for each observation. For notational convenience, we shall use $N = mn$ throughout the paper and will consider asymptotic regimes where $N \to \infty$. We note that, when $m = 1$, we recover the classical white-noise-with-drift model.

We consider the canonical signal detection problem, where the goal is to test for the presence or absence of the "signal component" f. More formally, we consider testing the null hypothesis $H_0 : f \equiv 0$ against the alternative hypotheses that

$$f \in H_\rho^{s,R} \equiv H_\rho^{s,R,p,q} := \left\{ f \in \mathcal{B}_{p,q}^{s,R} : \|f\|_{L_2} \geqslant \rho \text{ and } \|f\|_{\mathcal{B}_{p,q}^s} \leqslant R \right\}. \tag{3}$$

Here, the alternative hypothesis consists of s-smooth functions in a Besov space, with $\| \cdot \|_{\mathcal{B}_{p,q}^s}$ denoting the Besov-(s,p,q)-norm and $\mathcal{B}_{p,q}^{s,R} \subset L_2[0,1]$ corresponds to the Besov ball of radius R, see Section E in the Supplementary Material [21] for the definitions. Besov spaces are a very rich class of function spaces. They include many traditional smoothness spaces such as Hölder and Sobolev spaces as special cases. We refer the reader to [80] for a detailed discussion on Besov spaces.

Using a wavelet transform, the above testing problem is equivalent the observations under the Gaussian sequence model, where each of the $j = 1, \dots, m$ machines observes $i = 1, \dots, n$ observations $X_i^{(j)} := (X_{lk;i}^{(j)})_{l \geqslant 1, k=1,\dots,2^l}$

$$X_{lk;i}^{(j)} = f_{lk} + \sigma Z_{lk;i}^{(j)}, \tag{4}$$

where the $Z_{lk;i}^{(j)}$'s are i.i.d. standard Gaussian. The equivalent hypotheses (3) in the sequence model simply follows by replacing the $L_2[0,1]$-norm with the $\ell_2(\mathbb{N})$-norm and the Besov space $\mathcal{B}_{p,q}^{s,R}$ set to $\{f \in \ell_2(\mathbb{N}) : \|f\|_{\mathcal{B}_{p,q}^s} < \infty\}$, where the Besov norm on the sequence space $\ell_2(\mathbb{N})$ is defined as

$$\|f\|_{\mathcal{B}_{p,q}^s} := \begin{cases} \left(\sum_{l=1}^{\infty} \left(2^{l(s+1/2-1/p)} \left\| (f_{lk})_{k=1}^{2^l} \right\|_p \right)^q \right)^{1/q} & \text{for } 1 \leqslant q < \infty, \\ \sup_{l \geqslant 1} 2^{l(s+1/2-1/p)} \left\| (f_{lk})_{k=1}^{2^l} \right\|_p & \text{for } q = \infty. \end{cases} \tag{5}$$

In other words, the results for testing under DP derived for the sequence model of (4) with hypothesis (3) apply to the model described by (2) also, with the same corresponding hypothesis.

Given a $\{0,1\}$ valued test T, where $T(Y) = 1$ corresponds to rejecting the null hypothesis, we define the testing risk sum of the type I and worst case type II error over the alternative class;

$$\mathcal{R}(H_\rho^{s,R}, T) = \mathbb{P}_0 T(Y) + \sup_{f \in H_\rho^{s,R}} \mathbb{P}_f T(Y).$$

For the range of values $2 \leqslant p \leqslant \infty$, $1 \leqslant q \leqslant \infty$, the *minimax separation rate* in the unconstrained case is known to be $\rho \asymp (\sigma^2/N)^{\frac{s}{2s+1/2}}$ (see e.g. [46]). This means that, for $\rho \gg$

$(\sigma^2/N)^{\frac{s}{2s+1/2}}$, there exists a sequence of consistent tests $T \equiv T_N$ such that $\mathcal{R}(H^{s,R}_{\rho}, T) \to 0$, whilst no such sequence of tests exists whenever $\rho \ll (\sigma^2/N)^{\frac{s}{2s+1/2}}$.

The minimax separation rate captures how the testing problem becomes easier, or more difficult, for different model characteristics. For distributed (ϵ, δ)-DP testing protocols, the minimax separation rate depends on the stringency of the privacy requirement, given by $\epsilon, \delta > 0$, as well as the model characteristics m, n, s and σ. That is, we aim to find ρ as a function of $m, n, s, \sigma, \epsilon, \delta$, such that $\inf_{T \in \mathscr{T}^{(\epsilon,\delta)}} \mathcal{R}(H^{s,p,q}_{\rho',R}, T)$ converges to either 0 or 1 depending on whether $\rho' \ll \rho$ or $\rho' \gg \rho$. The class of alternatives under consideration are subsets of the Besov ball $\mathcal{B}^{s,R}_{p,q}$, where $2 \leqslant p < \infty$, $1 \leqslant q \leqslant \infty$, which offers a framework for functions in (2) with specific smoothness characteristics. Our results extend easily to the case where $p = \infty$ at the cost of an additional logarithmic factor in the rate and a few additional technicalities in the proofs.

1.3. *Main results and our contribution.* We quantify the difficulty of the federated testing problem outlined in the previous section in terms of the minimax separation rate. To achieve this, we present constructive distributed (ϵ, δ)-DP testing protocols that achieve consistent testing for the problem described above for certain values of ρ (Theorem 4 in Section 3). Additionally, we establish matching minimax lower bounds, up to logarithmic factor, for the testing risk (Theorem 5 in Section 5), providing a lower bound on the performance of distributed (ϵ, δ)-DP testing protocols.

Our analysis uncovers several novel and intriguing findings, which we briefly highlight here. The performance guarantees for the methods demonstrated in Section 3, along with the lower bounds established in Section 5, indicate that the distributed (ϵ, δ)-DP testing problem for the hypotheses given in (3) is governed by the minimax separation rate (up to logarithmic factors)

(6)
$$\rho^2 \asymp \left(\frac{\sigma^2}{mn}\right)^{\frac{2s}{2s+1/2}} + \left(\frac{\sigma^2}{mn^{3/2}\epsilon\sqrt{1 \wedge n\epsilon^2}}\right)^{\frac{2s}{2s+1}} \wedge \left(\left(\frac{\sigma^2}{\sqrt{m}n\sqrt{1 \wedge n\epsilon^2}}\right)^{\frac{2s}{2s+1/2}} + \frac{\sigma^2}{mn^2\epsilon^2}\right).$$

The precise statement is deferred to Theorem 1.

The derived rate indicates that the distributed testing problem under privacy constraints undergoes multiple phase transitions, resulting in different regimes where ϵ affects the detection boundary differently. Specifically, a smaller ϵ, which implies a stronger privacy guarantee, leads to an increased detection threshold. When δ decreases polynomially with N, its impact on the detection boundary is limited to a logarithmic factor, making its effect on the error rate minor compared to that of ϵ.

For $m = 1$, our theorems establish the optimal separation rate for nonparametric goodness-of-fit testing in the central DP setting, where all data is available on a single machine. When $\epsilon \lesssim 1/\sqrt{N}$, the privacy constraint affects the rate polynomially. In contrast, for $\epsilon \gtrsim 1/\sqrt{N}$, the rate approximates the classical minimax rate, up to logarithmic factors. Thus, the privacy constraint significantly impacts the rate only when ϵ is relatively small compared to the total number of observations N.

When $n = 1$, we establish the optimal separation rate for the testing problem in the local DP setting. Here, ϵ can be seen to have a pronounced effect on the rate whenever $\epsilon \lesssim 1$.

In the general federated setting, with $m \gg 1$, we see that m and n come into play with different powers in the minimax rate whenever $\epsilon^2 \lesssim \sigma^{\frac{1}{2s+1}} m^{\frac{1}{4s+1}} n^{\frac{1/2-2s}{4s+1}}$. This means that if one distributes $N = mn$ observations across m machines, the task becomes more challenging as the N observations are spread over a greater number of machines, rather than having many observations on a smaller number of machines. This phenomenon is also observed in the comparable estimation setting of nonparametric regression, as recently investigated in [20]. The phase transitions, however, are not observed in its estimation counterpart. We provide a detailed interpretation of the phase transitions in Section 2.

Our analysis also reveals that the minimax rate obtained in Theorem 1 becomes worse without access to shared randomness. This is revealed by Theorem 1 in Section 2. For certain values of ϵ, we show that the performance is strictly worse for methods that use only local randomness, and we exhibit optimal local and shared randomness methods for these regimes, respectively, in Sections 3.2 and 3.3. The shared source of randomness does not reveal any information on the identity of the individuals comprising the sample, even when it is publicly available. This means that the shared randomness does not violate the privacy constraints, and the improvement in the rate is a direct result of the shared randomness. We comment on this further in Section 2.

In many practical applications, the regularity parameter s is unknown. In Section 4, we extend the methods of Section 3 that attain the minimax rates, up to additional logarithmic factors, when the regularity is unknown. That is, establish that adaptive testing is possible under DP constraints with minimal additional cost in terms of the separation rate.

The lower bound relies on several technical innovations, which are summarized in Section 5. There, we provide a sketch of the proof and highlight the key innovations. The optimal methods proposed in this paper rely on carefully tailored private test statistics that extend to the adaptive setting with minimal cost due to exponential concentration bounds. We describe the construction of these tests in Sections 3 and 4.

1.4. *Related Work.* The literature on the theoretical properties of DP can be mostly divided into those studying *local* DP or *central* DP. In local DP, the privacy protection is applied at the level of individual data entries or observations, which corresponds to $n = 1$ in our setting. This is a stringent form of DP because each item of data is independently given privacy protection. In the other extreme, central DP, only the inference output needs to satisfy the DP constraint (i.e. $m = 1$ in our setting), meaning that if the output is a test, only the final decision needs to satisfy a DP constraint.

Locally differential private estimation has been studied in the context of the many-normal-means model, discrete distributions and parametric models in [34, 35, 1, 85]. The problem of density estimation under local DP constraints has been considered by [35, 71, 53, 18], of which the latter three works consider adaptation. In the context of hypothesis testing, [42, 72, 7, 2, 14, 4, 3] study testing under local DP for discrete distributions. Nonparametric goodness-of-fit testing under local DP is considered in [33, 55], where in [55], the authors consider adaptation as well. In [19], the authors consider estimation of a quadratic functional under local differential privacy constraints, which has connections to goodness-of-fit testing.

Settings in which the full data is assumed to be on a single server (i.e. $m = 1$), where a single privacy constraint applies to all the observations, have also been studied for various parametric high-dimensional problems [74, 39, 12, 49, 50, 22, 65, 17, 24]. In [54], nonparametric density estimation with known smoothness is considered. When it comes to hypothesis testing under central DP, [25] study simple hypothesis testing. [8] considers uniformity and independence testing in the multinomial model and [26, 66] study signal detection in the many-normal-means model. In [9], hypothesis testing in a linear regression setting is considered.

Investigations into the more general federated setting have been much more limited, with estimation being considered in [63, 6], which study estimation for discrete distributions and [57, 67] which study mean estimation, [20] which study nonparametric regression, and [58] which consider sparse linear regression. In the paper [27], the authors consider discrete distribution testing in a two server setting ($m = 2$) with differing DP constraints.

1.5. *Organization of the paper.* The rest of the paper is organized as follows. In Section 2, we present the main results of the paper, for the known smoothness case and the adaptive setting. Next, Section 3 presents the methods that achieve the optimal rates derived in Section 2. In Section 4, we extend these methods to be adaptive in the case that the smoothness is unknown. In Section 5, we present the lower bound theorems for the testing problem and give a sketch of its proof. Further proofs are deferred to the supplemental material to the article, [21].

1.6. *Notation, definitions and assumptions.* Throughout the paper, we shall write $N := mn$. For two positive sequences a_k, b_k we write $a_k \lesssim b_k$ if the inequality $a_k \leqslant C b_k$ holds for some universal positive constant C. Similarly, we write $a_k \asymp b_k$ if $a_k \lesssim b_k$ and $b_k \lesssim a_k$ hold simultaneously and let $a_k \ll b_k$ denote that $a_k / b_k = o(1)$.

We use the notations $a \vee b$ and $a \wedge b$ for the maximum and minimum, respectively, between a and b. For $k \in \mathbb{N}$, $[k]$ shall denote the set $\{1, \ldots, k\}$. Throughout the paper c and C denote universal constants whose value can differ from line to line. The Euclidean norm of a vector $v \in \mathbb{R}^d$ is denoted by $\|v\|_2$. For a matrix $M \in \mathbb{R}^{d \times d}$, the norm $M \mapsto \|M\|$ is the spectral norm and $\text{Tr}(M)$ is its trace. Furthermore, we let I_d denote the $d \times d$ identity matrix.

Throughout the paper, d_H is the Hamming distance on $\mathcal{X}^n$ is defined as $\mathrm{d}_H(x, \breve{x}) := \sum_{i=1}^n \mathbb{1}\{x_i \neq \breve{x}_i\}$ for $x = (x_i)_{i=1}^n, \breve{x} = (\breve{x}_i)_{i=1}^n \in \mathcal{X}^n$. Furthermore, for a vector space $\mathcal{X}$ and $x = (x_i)_{i \in [n]} \in \mathcal{X}^n$, we shall write $\overline{x}$ for the average $n^{-1} \sum_{i=1}^n x_i$.

2. Minimax optimal testing rates under privacy constraints.

In this section, we discuss the main results in detail. We start the discussion with results for the oracle case where the regularity parameter is known in Section 2.1. Section 2.2 describes the main results for when the regularity is not known.

2.1. *Description of the minimax separation rate.* We first give a precise statement concerning the minimax separation rate shown in (6).

THEOREM 1. *Let $s, R > 0$ be given and consider any sequences of natural numbers $m \equiv m_N$ and $n := N/m$ such that $N = mn \to \infty$, $1/N \ll \sigma \equiv \sigma_N = O(1)$, $\epsilon \equiv \epsilon_N$ in $(N^{-1}, 1]$ and $\delta \equiv \delta_N \lesssim N^{-(1+\omega)}$ for any constant $\omega > 0$. Let ρ a sequence of positive numbers satisfying (6).*

Then,

$$\inf_{T \in \mathscr{T}_{\text{SHR}}^{(\epsilon,\delta)}} \mathcal{R}(H^{s,R}_{\rho M_N}, T) \to \begin{cases} 0 & \textit{for any } M_N^2 \gg \log\log(N) \log^{3/2}(N) \log(1/\delta), \\ 1 & \textit{for any } M_N \to 0. \end{cases}$$

The proof of the theorem is given in Section B.4 of the Supplementary Material [21]. It is based on a combination of upper and lower bounds, where the lower bound is established in Section 5. The upper bound is given in Section 3, where we present an (ϵ, δ)-DP distributed testing protocol that attains the rate in Theorem 1. These upper and lower bounds are in fact non-asymptotic, meaning that they do not require the assumption that $N \to \infty$.

Theorem 1 shows multiple regime changes, where the distributed testing problem under privacy constraints undergoes a change in the minimax separation rate. Later on in this section, we highlight the different regimes and give an interpretation to each of them.

Theorem 1 considers the minimax rate for the class of distributed protocols with access to shared randomness, $\mathscr{T}_{\text{SHR}}^{(\epsilon,\delta)}$. Theorem 2 below considers the minimax rate for the (strictly smaller) class of distributed protocols without access to shared randomness, $\mathscr{T}_{\text{LR}}^{(\epsilon,\delta)}$. Here, transcripts depend *only* on their local data and possibly a local source of randomness.

THEOREM 2. *Let $s, R > 0$ be given and consider any sequences of natural numbers $m \equiv m_N$ and $n := N/m$ such that $N = mn \to \infty$, $1/N \ll \sigma \equiv \sigma_N = O(1)$ and $\epsilon \equiv \epsilon_N$ in $(N^{-1}, 1]$ and $\delta \equiv \delta_N \lesssim N^{-(1+\omega)}$ for any constant $\omega > 0$. Let $\rho \equiv \rho_N$ a sequence of positive numbers satisfying*

$$\rho^2 \asymp \left(\frac{\sigma^2}{mn}\right)^{\frac{2s}{2s+1/2}} + \left(\frac{\sigma^2}{mn^2\epsilon^2}\right)^{\frac{2s}{2s+3/2}} \wedge \left(\left(\frac{\sigma^2}{\sqrt{mn}\sqrt{1 \wedge n\epsilon^2}}\right)^{\frac{2s}{2s+1/2}} + \left(\frac{\sigma^2}{mn^2\epsilon^2}\right)\right). \tag{7}$$

Then,

$$\inf_{T \in \mathscr{T}_{LR}^{(\epsilon,\delta)}} \mathcal{R}(H^{s,R}_{\rho M_N}, T) \to \begin{cases} 0 \text{ for any } M_N^2 \gg \log\log(N) \log^{3/2}(N) \log(1/\delta), \\ 1 \text{ for any } M_N \to 0. \end{cases}$$

The proof of Theorem 1 is given in Section B.4 of the supplemental material [21]. The theorem shows that, depending on the value of ϵ, the minimax rate for protocols that do not have access to shared randomness is strictly worse than those for protocols that do have access to shared randomness.

To more easily compare the two theorems, we provide a table in Table 1 below, where we separate six "regimes" to aid interpretability below. Each of the regimes correspond to the dominating term in the minimax separation rates of Theorems 1 and 2. Which term dominates depends on the value of ϵ, in comparison to n, m, σ, s and the availability of shared randomness.

	Regime 1	**Regime 2**	**Regime 3**	**Regime 4**	**Regime 5**	**Regime 6**
Shared U	$\left(\frac{\sigma^2}{mn}\right)^{\frac{2s}{2s+1/2}}$	$\left(\frac{\sigma^2}{mn^{3/2}\epsilon}\right)^{\frac{2s}{2s+1}}$	$\left(\frac{\sigma^2}{mn^2\epsilon^2}\right)^{\frac{2s}{2s+1}}$	$\left(\frac{\sigma^2}{\sqrt{mn}}\right)^{\frac{2s}{2s+1/2}}$	$\left(\frac{\sigma^2}{\sqrt{m}n^{3/2}\epsilon}\right)^{\frac{2s}{2s+1/2}}$	$\frac{\sigma^2}{mn^2\epsilon^2}$
Local only	$\left(\frac{\sigma^2}{mn}\right)^{\frac{2s}{2s+1/2}}$	$\left(\frac{\sigma^2}{mn^2\epsilon^2}\right)^{\frac{2s}{2s+3/2}}$	$\left(\frac{\sigma^2}{mn^2\epsilon^2}\right)^{\frac{2s}{2s+3/2}}$	$\left(\frac{\sigma^2}{\sqrt{mn}}\right)^{\frac{2s}{2s+1/2}}$	$\left(\frac{\sigma^2}{\sqrt{m}n^{3/2}\epsilon}\right)^{\frac{2s}{2s+1/2}}$	$\frac{\sigma^2}{mn^2\epsilon^2}$

TABLE 1

The minimax separation rates for the testing problem under privacy constraints, for both the local randomness and shared randomness settings. The rates are given up to logarithmic factors. The regimes are defined by the values of ϵ and the model characteristics m, n, σ, s.

The rates in Regimes 4, 5 and 6 are the same for both types of protocols, and these rates are attained by the same testing protocol for each of the classes, which does not require shared randomness. We shall refer to Regime 4, 5 and 6 as the "low privacy-budget" regimes, as these rates occur for relatively small values of ϵ.

We shall refer to Regimes 1, 2 and 3 as the "high privacy-budget" regimes, as these rates occur for relatively large values of ϵ. These rates are achieved by a different protocol, for the classes of protocols with and without shared randomness, respectively. These protocols are given in Sections 3.2 and 3.3. Proving the tighter lower bound in case of the class of protocols with access to local randomness only, requires a different technique to that of the class of shared randomness protocols, which we outline in Section 5.

The improvement in the rate for the shared randomness protocols, compared to the local randomness protocols, are visible for Regime 2 and Regime 3, but also the values of ϵ for which the different regimes occur are different for the two types of protocols. The improvement in the rate is loosely speaking a consequence of the improved coordination between the servers made possible by shared randomization. We exhibit a distributed (ϵ, δ)-DP shared randomness protocol attaining the above rate in Section 3.3.

In case of access to shared randomness, the high privacy-budget regime occurs whenever $\epsilon \gtrsim \sigma^{-\frac{2}{4s+1}} m^{-\frac{2s}{4s+1}} n^{\frac{1/2-2s}{4s+1}}$ and $\epsilon \geqslant n^{-1/2}$, or $\sigma^{-\frac{1}{2s}} m^{-\frac{1}{2}} n^{\frac{1-2s}{4s}} \leqslant \epsilon < n^{-1/2}$. In the case of local randomness only, the high privacy-budget regimes occur for larger values of ϵ, namely $\epsilon \gtrsim \sigma^{-\frac{2}{4s+1}} m^{\frac{1}{4s+1}} n^{\frac{1/2-2s}{4s+1}}$ if $\epsilon \geqslant n^{-1/2}$ or $\epsilon \gtrsim \sigma^{-\frac{4}{4s-1}} m^{-\frac{1}{2}} n^{\frac{5/2-2s}{4s-1}}$ for $\epsilon \lesssim n^{-1/2}$, whenever $s > 1/4$. When $s \leqslant 1/4$ and no shared randomness is available, we are always in the low privacy-budget regime for the range of $N^{-1} \lesssim \sigma \lesssim 1$ and $N^{-1} < \epsilon \lesssim 1$ considered.

The testing protocols that attains the rate in the high privacy-budget regimes are given in Section 3.2 and 3.3. These protocols bear some resemblance with the estimation strategy in [20], where transcripts constitute noisy, lower dimensional approximations of the original data. As ϵ increases, the dimensionality of these approximations increases and the rate improves. In Regime 2 and 3, the minimax rate is at least a polynomial factor of ϵ larger than the unconstrained, non-private minimax separation rate, which is attained in Regime 1 (up to a logarithmic factor).

The observation that shared randomness can improve performance in our problem has been noted in other contexts involving distributed privacy and communication constraints, see for example [4, 5, 3, 76, 77]. [33, 19] studies interactive versus non-interactive protocols and finds a difference in terms of minimax performance between the two in the local differential privacy setting. Interestingly, when $n = 1$ (i.e. in the local differential privacy setting), we find the similar minimax rates for nonparametric goodness-of-fit testing in the high privacy-budget regimes, for the shared randomness and local randomness protocols, as they do for interactive and non-interactive protocols, whenever ϵ is in the high-budget regime. Although they study a different model, observations from smooth densities; it is interesting to see that the same rates seem to be attainable without sequential interaction, by using shared randomness instead. We note here that, when sequential- or interactive protocols are allowed, shared randomness can be employed in particular. In real applications without interaction, one should always use shared randomness if at all possible.

In the low privacy-budget case, i.e. Regimes 4, 5 and 6, the minimax rate for both local and shared randomness protocols coincides. We note that the regimes occur at different values of ϵ for the two types of protocols, however. Within the low privacy-budget range, we find essentially three different regimes. When $n^{-1/2} \leqslant \epsilon < \sigma^{-\frac{2}{4s-1}} m^{\frac{1/4-s}{4s+1}} n^{\frac{1/2-2s}{4s+1}}$, the rate is given by $(\frac{\sigma^2}{\sqrt{mn}})^{\frac{2s}{2s+1/2}}$. What is remarkable here, is that whilst the rate is polynomially worse in m than the unconstraint rate, the rate is otherwise independent of ϵ. This regime essentially corresponds to a setting where, even though a high privacy-budget strategy is not feasible, the desired level of privacy is achieved “for free" with the locally optimal test statistic.

We exhibit an (ϵ, δ)-DP distributed testing protocol that attains this rate in Section 3.1. This strategy can roughly be described as first computing a locally optimal private test statistic – a test statistic that would result in the optimal private test using just the local data – and then averaging these private test statistics; essentially combining the power of the local tests. That this strategy performs well when the privacy constraint is sufficiently stringent can intuitively be explained as that it is easier to retain privacy when only (a private version of) a single real valued local test statistic is shared, rather than a (private approximation of) the original data.

When $\epsilon \lesssim n^{-1/2}$ within the low privacy-budget range, ϵ affects the rate polynomially. For the smallest values of ϵ, i.e. $\epsilon \lesssim \sigma^{\frac{1}{2s+1}} m^{-\frac{1}{2}} n^{-\frac{1+s}{2s+1}}$, the rate is given by $\frac{\sigma^2}{mn^2\epsilon^2}$. Strikingly, the regularity parameter does not appear in the rate in this regime. This phenomenon has the following explanation: for such small values of ϵ, signals of size $\frac{\sigma^2}{mn^2\epsilon^2}$ are of larger order than the local estimation rate of $(\frac{\sigma^2}{n})^{\frac{2s}{2s+1/2}}$. Consequently, signal can locally be estimated with high accuracy, and the bottleneck is purely the privacy constraint, not the high-dimensional nature of the problem.

In the case of central DP (i.e. $m = 1$), only the low privacy-budget regime is observed. In this case, our results show that the non-private rate is attainable (up to logarithmic factors) for $\epsilon \gtrsim 1/\sqrt{N}$. This is in contrast to the local DP setting, where both the high- and low privacy budget regimes are observed (depending on the values of σ and s). Whenever m is larger than say polynomial in N in the low privacy-budget range, (i.e. $m \asymp N^{\omega}$ for some $\omega > 0$) the unconstrained minimax rate cannot be reached.

To further illustrate the difference between these classes of protocols, we provide two plots in Figure 2. The plots show the relationship between the minimax testing rate ρ and ϵ for fixed values of m, n, σ and four different choices for the regularity s. The regimes correspond to the six regimes in Table 1. The plots show that the shared randomness setting strictly improves the rate for certain values of ϵ, and that the values of ϵ for which the different regimes occur are different for the two types of protocols. A full case-wise breakdown of when each of the regimes occur is given in Section C of the Supplementary Material [21]. Below, we give an interpretation for each of the regimes.

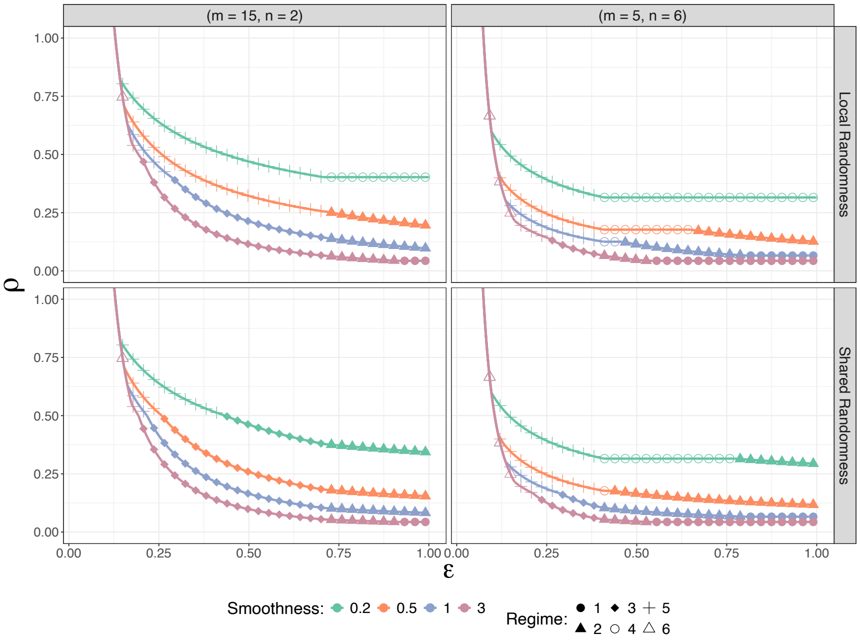

FIGURE 2. *The relationship of the minimax testing rate ρ and ϵ, given by (7) and (6), for $(n, m) = (5, 5)$ in the left column and $(n, m) = (2, 15)$ in the right column, $\sigma = 1$ and smoothness levels $s = 1/5$, $s = 1/2$, $s = 1$ and $s = 3$. The panels on the first row correspond to distributed (ϵ, δ)-DP (local randomness only) protocols (i.e. (7)), the bottom row corresponds to distributed (ϵ, δ)-DP protocols with shared randomness (i.e. (6)). The regimes correspond to the six regimes (e.g. different rates) in Table 1.*

What constitute "moderate" or "large" values, depends on the size of m relative to n, as can be seen when comparing Figure 2, which compares $N = 30$ observations distributed either between $m = 15$ servers with 2 observations each, and $m = 5$ servers with $n = 6$ observations each. It can be seen that, as the local sample n is larger compared to the number of times the total number of data points N is divided m, the cost of privacy is less. This underlines the idea that, in large samples, it is easier to retain privacy.

When ϵ becomes "very small" (smaller than a threshold depending on s, m and n), the smoothness starts to matter less and less, up to the point where the difficulty of the problem is no different for (very) different regularity levels. These scenarios correspond to settings where the privacy requirement underlying the problem is so stringent, that it effectively becomes the bottleneck of the testing problem.

2.2. *Adaptation.* In the previous section, we derived the minimax separation rate for the nonparametric distributed testing problem. However, the proposed tests constructed in Section 3 require knowledge of the regularity parameter s of the underlying f. Typically, the regularity of the function is unknown in practice, necessitating the use of data-driven methods to find the best adaptive testing strategies.

Given that the regularity of the underlying signal class is unknown, it makes sense to consider the minimax testing risk

$$\sup_{s\in[s_{\min},s_{\max}]} \mathcal{R}\left(H^{s,R}_{M_{N,s}\rho_s}, T\right),$$

for certain predetermined values $0 < s_{\min} < s_{\max} < \infty$. Here, we consider separation rates ρ_s depending on the underlying smoothness. In the case that the true underlying smoothness is $s = s_{\min}$, the separation rate is relatively larger than when (for example) $s = s_{\max}$. In the case that the true smoothness s is larger than $s_{\min}$, we would like to attain the smaller of the two rates ρ_s.

In the non-privacy constraint setting, adaptation for the above risk can be achieved with only a minor additional cost in the separation rate (a $\log\log N$ factor). See for example Theorem 2.3 in [75] or Section 7 in [47]. Theorem 3 below shows that also under privacy constraint, the optimal private rate can be attained by a protocol that is adaptive to the regularity parameter s, with minimal additional cost; at most a logarithmic factor.

THEOREM 3. *Let $0 < s_{\min} < s_{\max} < \infty$, $R > 0$ be given and consider any sequences of natural numbers $m \equiv m_N$ and $n := N/m$ such that $N = mn \to \infty$, $1/N \ll \sigma \equiv \sigma_N = O(1)$, $\epsilon \equiv \epsilon_N$ in $(N^{-1}, 1]$ and $\delta \equiv \delta_N \lesssim N^{-(1+\omega)}$ for any constant $\omega > 0$.*

If ρ a sequence of positive numbers satisfies (6), there exists a sequence of distributed (ϵ, δ)-DP testing protocols T_N such that

$$\sup_{s\in[s_{\min},s_{\max}]} \mathcal{R}(H^{s,R}_{\rho M_N}, T_N) \to \begin{cases} 0 \text{ for any } M_N^2 \gg \log\log(N)\log^{5/2}(N)\log(1/\delta) \\ 1 \text{ for any } M_N \to 0. \end{cases}$$

Furthermore, whenever ρ satisfies (7), there exists a sequence of distributed (ϵ, δ)-DP testing protocols T_N using only local randomness such that the above display holds as well.

We construct such adaptive distributed (ϵ, δ)-DP testing protocols in Section 4 and their resulting performance proofs the above theorem. The adaptive methods can be seen as extensions of the methods exhibited in Section 3 for when the smoothness is known. The adaptive methods can essentially be seen as a multiple testing extension of the known smoothness methods, testing along a grid of smoothness levels between $s_{\min}$ and $s_{\max}$. The strain on the privacy budget stemming from conducting multiple testing procedures is limited, due to the fact that the cardinality of this grid is order $\log(N)$. The Type I error control is assured by a Bonferroni correction, which leverages the exponential bounds on the Type I error of the individual "known smoothness tests".

3. Optimal differentially private testing procedures.

In this section, we construct distributed (ϵ, δ)-DP testing procedures that attain the minimax separation rates derived in Section 2.

The testing procedures are constructed in three steps. First, in Section 3.1, we construct a distributed (ϵ, δ)-DP testing procedure that uses only local randomness and that is optimal in the low privacy-budget regime described in the previous section. We refer to this procedure as T_{I}. Second, we construct two distributed (ϵ, δ)-DP testing procedures that use local randomness and shared randomness, respectively, and that are optimal in their respective high privacy-budget regimes. We refer to these procedures as T_{II} and T_{III} and describe them in Sections 3.2 and 3.3, respectively.

The testing procedures differ in terms of the testing strategy. In the low privacy-budget case where T_{I} is optimal, the testing strategy can be seen to consist of first computing a locally optimal private test statistic in each machine; that is, a test statistic that would result in the optimal private test using just the local data. The locally optimal test statistic is based on the squared Euclidean norm of the truncated observation. To deal with the nonlinearity of the Euclidean norm, the strategy appropriately restricts the domain of the clipped locally optimal test statistic, after which we employ a Lipschitz-extension to obtain a test statistic that is well-defined on the sample space and more robust to outliers than the Euclidean norm itself. The noisy version of this test statistic is locally optimal under privacy constraints, in the sense that a corresponding (strict) p-value test attains the lower bound rate (up to a logarithmic factor) as established by Theorem 2 for the case where $m = 1$. When $m > 1$, the final test statistic is obtained by averaging the locally optimal private test statistics.

In the large ϵ regime, instead of computing a locally optimal test statistic, both T_{II} and T_{III} are based on truncated, clipped and noisy versions of the local observations. The key difference between the two is that the latter uses the same random rotation of the local observations, which is made possible by the availability of shared randomness.

Together, the methods prove Theorem 4 below, which forms the "upper bound" part of the minimax separation rate described by Theorems 2 and 1. Unlike the formulation of the latter theorems, we note that the result is not asymptotic.

THEOREM 4. *Let $s, R > 0$ be given. For all $\alpha \in (0,1)$, there exists a constant $C_\alpha > 0$ such that if*

(8)
$$\rho^2 \geqslant C_\alpha \left(\frac{\sigma^2}{mn}\right)^{\frac{2s}{2s+1/2}} + \left(\frac{\sigma^2}{mn^2\epsilon^2}\right)^{\frac{2s}{2s+3/2}} \wedge \left(\left(\frac{\sigma^2}{\sqrt{mn}\sqrt{1 \wedge n\epsilon^2}} \right)^{\frac{2s}{2s+1/2}} + \left(\frac{\sigma^2}{mn^2\epsilon^2}\right) \right),$$

there exists a distributed (ϵ, δ)-DP testing protocol $T \equiv T_{m,n,s,\sigma}$ such that

$$\mathcal{R}(H^{s,R}_{\rho M_N}, T) \leqslant \alpha, \tag{9}$$

for all natural numbers m, N and $n = N/m$, $\sigma \in [1/N, \sigma_{\max}]$, $\epsilon \in (N^{-1}, 1]$, $\delta \leqslant N^{-(1+\omega)}$ for any constant $\omega > 0$, $\sigma_{\max} > 0$ and a nonnegative sequence $M_N^2 \gtrsim \log\log(N)\log^{3/2}(N)\log(1/\delta)$.

Similarly, for any $\alpha \in (0,1)$, there exists a constant $C_\alpha > 0$ such that if

(10)
$$\rho^2 \geqslant C_\alpha \left(\frac{\sigma^2}{mn}\right)^{\frac{2s}{2s+1/2}} + \left(\frac{\sigma^2}{mn^{3/2}\epsilon\sqrt{1 \wedge n\epsilon^2}}\right)^{\frac{2s}{2s+1}} \wedge \left(\left(\frac{\sigma^2}{\sqrt{mn}} \right)^{\frac{2s}{2s+1/2}} + \left(\frac{\sigma^2}{mn^2\epsilon^2}\right) \right),$$

we have that there exists a distributed (ϵ, δ)-DP shared randomness testing protocol $T \equiv T_{m,n,s,\sigma}$ such that

$$\mathcal{R}(H^{s,R}_{\rho M_N}, T) \leqslant \alpha, \tag{11}$$

for all natural numbers m, N and $n = N/m$, $\sigma \in [1/N, \sigma_{\max}]$, $\epsilon \in (N^{-1}, 1]$, $\delta \leqslant N^{-(1+\omega)}$ for any constant $\omega > 0$ and a nonnegative sequence $M_N^2 \gtrsim \log\log(N)\log^{3/2}(N)\log(1/\delta)$.

The proof of the theorem follows directly from the guarantees proven for each of the three testing protocols; we defer it to Section B in the supplement. Before giving the detailed construction of the three tests, we introduce some common notation. Let Π_L denote the projection of elements $\mathbb{R}^{\mathbb{N}}$ onto the first $d_L := \sum_{l=1}^{L} 2^l$ coordinates, where the elements as ordered and indexed as follows;

$$\Pi_L x = (x_{11}, \dots, x_{12}, x_{21}, \dots, x_{14}, \dots, x_{L1}, \dots, x_{L2^L}, 0, 0, 0, \dots).$$

We shall also use the notation $d_L := \sum_{l=1}^{L} 2^l$ and let $X_{L;i}^{(j)}$ denote vector in $\mathbb{R}^{d_L}$ formed by the first d_L coordinates of $\Pi_L X_i^{(j)}$ and let $X_L^{(j)} = (X_{L;i}^{(j)})_{i \in [n]}$. Furthermore, we recall that for $v = (v_1, \dots, v_n) \in \mathcal{X}^n$ for a vector space $\mathcal{X}$, $\bar{v}$ denotes the vector space average $n^{-1} \sum_{i=1}^{n} v_i$.

In order to obtain statistics with (uniformly) bounded sensitivity it is useful to bound quantities between certain thresholds. Formally, for $a, b, x \in \mathbb{R}$ with $a < b$, let $[x]_a^b$ denote x *clipped between* a *and* b, that is

$$[x]_a^b := \begin{cases} b & \text{if } x > b, \\ x & \text{if } a \leqslant x \leqslant b \\ a & \text{otherwise.} \end{cases} \tag{12}$$

The distributed privacy protocols under consideration in this paper can be seen as noisy versions of statistics of the data. Roughly put, the “amount” of noise added depends on the *sensitivity* of the statistics. This brings us to the concept of sensitivity. Formally, consider a metric d on a set $\mathcal{Y}$. Given n elements $x = (x_1, \dots, x_n)$ in a sample space $\mathcal{X}$, the d-*sensitivity at* x of a map $S : \mathcal{X}^n \to \mathcal{Y}$ is

$$\Delta_S(x) := \sup_{\check{x} \in \mathcal{X}^n : \mathrm{d}_H(x, \check{x}) \leqslant 1} \mathrm{d}\left(S(x), S(\check{x})\right),$$

where d_H is the Hamming distance on $\mathcal{X}^n$ (see Section 1.6 for a definition). The d-*sensitivity* of S is defined as $\Delta_S := \sup_x \Delta_S(x)$. In this paper, the main noise mechanism is the *Gaussian mechanism*. The Gaussian mechanism yields (ϵ, δ)-differentially private transcripts for statistics that have bounded L_2-sensitivity, with the noise variance scaling with the L_2-sensitivity. See [36] for a thorough treatment. We remark that for the rates in Regime 3 up until 6 in Table 1, $(\epsilon, 0)$-DP can be attained by employing a Laplace mechanism instead. That is, for the values of ϵ for which Regime 3 up until 6 in Table 1 are optimal, the test statistics in the sections have matching L_1- and L_2-sensitivity, so the Gaussian mechanism can be replaced by the Laplace mechanism instead in these regimes.

3.1. *Private testing procedure I: low privacy-budget strategy.*

In the classical setting without privacy constraints (and $m = 1$), a rate optimal test for the hypotheses of (3) is given by

$$\mathbb{1}\left\{ S_L^{(j)} > \kappa_\alpha \right\}, \text{ where } S_L^{(j)} := \frac{1}{\sqrt{d_L}} \left(\left\| \sigma^{-1} \sqrt{n} \overline{X_L^{(j)}} \right\|_2^2 - d_L \right), \tag{13}$$

where $d_L := \sum_{l=1}^{L} 2^l$ and the rate optimal choice of L is $L_* = \left\lceil \frac{1}{2s+1/2} \log_2(N) \right\rceil$. Under the null hypothesis, $S_{L_*}^{(j)}$ is Chi-square distributed degrees of freedom. Under the alternative hypothesis, the test statistic picks up a positive “bias” as $\|\sigma^{-1}\sqrt{n}\overline{X_{L_*}^{(j)}}\|^2 \sim \chi^2_{L_*}(\|\Pi_{L_*} f\|_2^2)$ under $\mathbb{P}_f$, which could surpass the critical value κ_α if $\sigma^{-2} n \|\Pi_{L_*} f\|_2^2$ is large enough. Consequently, the level of the test is controlled by setting κ_α appropriately large. For a proof of its rate optimality, see e.g. [43].

As is commonly the case for superlinear functions, the test statistic $S_{L;\tau}^{(j)}$ has poor sensitivity uniformly over the sample space, meaning that a change in just one datum can result in a large change in the test statistic. This means that it forms a poor candidate to base a privacy preserving transcript on. For example, one would need to add a substantial amount of noise guarantee DP for the statistic. To remedy this, we follow a similar strategy as proposed in [26] and improved upon by [66]. We construct a clipped and symmetrized version of the test statistic above, which has small sensitivity on a set $\mathcal{C}_{L;\tau}$, in which $X^{(j)}$ takes values with high probability. We define the test statistic explicitly on $\mathcal{C}_{L;\tau}$ only. By a version of the McShane–Whitney–Extension Theorem, we obtain a test statistic with the same sensitivity that is defined on the entire sample space.

Consider for $\tau > 0$, $L \in \mathbb{N}$, $d_L := \sum_{l=1}^{L} 2^l$ and $V_{L;\tau}^{(j)} \sim \chi^2_{d_L}$ independent of $X^{(j)}$ the random map from $(\mathbb{R}^{d_L})^n$ to $\mathbb{R}$ defined by

$$\tilde{S}_{L;\tau}^{(j)}(x) = \left[\frac{1}{\sqrt{d_L}} \left(\|\sigma^{-1}\sqrt{n}\overline{x}\|_2^2 - V_{L;\tau}^{(j)} \right) \right]_{-\tau}^{\tau} . \tag{14}$$

For any τ, this test statistic $\tilde{S}_{L;\tau}^{(j)}(X_L^{(j)})$ can be seen to have mean zero and bounded variance under the null hypothesis, by similar reasoning as for the test statistic in (13) (see the proof of Lemma 1 for details).

Loosely speaking, the test statistic $\tilde{S}_{L;\tau}^{(j)}(X_L^{(j)})$ retains the signal as long as $\tau > 0$ is chosen appropriately in comparison to the signal size (i.e. $\|\Pi_L f\|_2^2$) and has good sensitivity for "likely" values of $X^{(j)}$ under $\mathbb{P}_f$, but not uniformly over the sample space. We make the latter statement precise as follows.

Let $K_\tau = \lceil 2\tau D_\tau^{-1} \rceil$ and consider the set $\mathcal{C}_{L;\tau} = \mathcal{A}_{L;\tau} \cap \mathcal{B}_{L;\tau}$, where

(15)

$$\mathcal{A}_{L;\tau} = \left\{ (x_i) \in (\mathbb{R}^\infty)^n : \left| \|\sigma^{-1} \textstyle\sum_{i \in \mathcal{J}} \Pi_L x_i \|_2^2 - k d_L \right| \leqslant \tfrac{1}{8} k D_\tau n \sqrt{d_L} \;\; \forall \mathcal{J} \subset [n], |\mathcal{J}| = k \leqslant K_\tau \right\},$$

$$\mathcal{B}_{L;\tau} = \left\{ (x_i) \in (\mathbb{R}^\infty)^n : |\langle \sigma^{-1} \Pi_L x_i, \sigma^{-1} \textstyle\sum_{k \neq i} \Pi_L x_k \rangle| \leqslant \frac{1}{8} k D_\tau n \sqrt{d_L}, \;\; \forall i = 1, \dots, n \right\}.$$

Lemma 18 in the supplement shows that $X^{(j)}$ concentrates on $\mathcal{C}_{L;\tau}$ when the underlying signal is, roughly speaking, not too large compared to τ (in particular under the null hypothesis).

It can be shown that, on the set $\mathcal{C}_{L;\tau}$, $x \mapsto S^{(j)}(x)$ is D_τ-Lipschitz with respect to the Hamming distance, see Lemma 19 in the supplement. Lemma 20 in the supplement shows that there exists a measurable function $S_{L;\tau}^{(j)} : (\mathbb{R}^{d_L})^n \to \mathbb{R}$, D_τ-Lipschitz with respect to the Hamming distance, such that $S_{L;\tau}^{(j)}(X_L^{(j)}) = \tilde{S}_{L;\tau}^{(j)}(X_L^{(j)})$ whenever $X^{(j)} \in \mathcal{C}_{L;\tau}$. Lemma 20 is essentially the construction of McShane [64] for obtaining a Lipschitz extension with respect to the Hamming distance, but our lemma verifies in addition the Borel measurability of the resulting map.

The Lipschitz constant upper bounds the sensitivity of a test statistic that is Lipschitz continuous with respect to the Hamming distance. Specifically, we have that

$$\Delta_{S^{(j)}} = \sup_{x, \breve{x} \in \ell_2(\mathbb{N})^n : \mathrm{d}_H(x, \breve{x}) \leqslant 1} \left| S^{(j)}(x) - S^{(j)}(\breve{x}) \right| \leqslant D_\tau .$$

Using the Gaussian mechanism, the transcripts

$$Y_{L;\tau}^{(j)} = \gamma_\tau \breve{S}_{L;\tau}^{(j)}(X_L^{(j)}) + W_\tau^{(j)}, \quad \text{where } W_\tau^{(j)} \sim N(0,1) \text{ independent for } j \in [m], \tag{16}$$

$\gamma_\tau = \epsilon / (D_\tau \sqrt{2\mathfrak{c} \log(2/\delta)})$ and $\tau > 0$, are $(\epsilon/\sqrt{\mathfrak{c}}, \delta)$-differentially private for any $\epsilon > 0$ (see e.g. [36]). These transcripts are mean zero and have bounded variance under the null hypothesis, so a test of the form

$$\varphi_\tau := \mathbb{1} \left\{ \frac{1}{\sqrt{m}} \sum_{j=1}^{m} Y_{L;\tau}^{(j)} \geqslant \kappa(\gamma_\tau \vee 1) \right\} \tag{17}$$

has an arbitrarily small level for large enough $\kappa > 0$ (see Lemma 21 in the supplement). Furthermore, the lemma below shows that, if the signal size is large enough in the $\sum_{l=1}^{L} 2^l$ first coordinates, the above test enjoys a small Type II error probability as well.

LEMMA 1. *Consider the test φ_τ as defined by (17). If*

$$\tau/4 \leqslant \frac{n\|f_L\|_2^2}{\log(N)\sqrt{2\mathfrak{c}\log(2/\delta)}\sigma^2\sqrt{d}} \leqslant \tau/2 \tag{18}$$

and

$$\|\Pi_L f\|_2^2 \geqslant C_\alpha \kappa \log(N)\sqrt{\mathfrak{c}\log(1/\delta)}\left(\frac{\sqrt{2^L}}{\sigma^2\sqrt{N}\sqrt{n}(\sqrt{n}\epsilon \wedge 1)}\right) \bigvee \left(\frac{1}{\sigma^2 N n\epsilon^2}\right) \tag{19}$$

for $C_\alpha > 0$ large enough, it holds that $\mathbb{P}_f(1-\varphi_\tau) \leqslant \alpha$.

A proof of the above lemma is given in Section B.1 of the supplement. The above test is calibrated for the detection of signals size between $\tau/4$ and $\tau/2$. In order to detect signals of any size larger than the right-hand side of (19), we follow what is essentially a multiple testing procedure. For large signals, we need a larger clipping to detect them, as well as a larger set $\mathcal{C}_{L;\tau}$ to assure that the data is in $\mathcal{C}_{L;\tau}$ with high probability, as larger signals increase the probability of "outliers" from the perspective of the sensitivity of the L_2-norm.

It turns out that a sufficient range of clipping thresholds to consider (for detecting the signals $f \in \mathcal{B}^{s,R}_{p,q}$ under consideration in Lemma 2) is given by

$$\tau \in \mathrm{T}_L := \left\{ 2^{-k+2}\frac{n(1-2^{-s})^{2-2/q}R^2}{\sigma^2\sqrt{2^L}} : k = 1, \ldots, \lceil 1 + 2\log_2(NR/\sigma)\rceil \right\}. \tag{20}$$

The (ϵ,δ)-differentially private testing procedure T_{I} is now constructed as follows. For each $\tau \in \mathrm{T}_L$, the machine transfers (16) with $\mathfrak{c} = |\mathrm{T}_L|$. By the independence of the Gaussian noise added in (16) for each $\tau \in \mathrm{T}_L$, the transcript $Y^{(j)} = \{Y^{(j)}_{L;\tau} : \tau \in \mathrm{T}_L\}$ is (ϵ,δ)-differentially private (see e.g. Theorem A.1 in [36]).

The test

$$T_{\mathrm{I}} := \mathbb{1}\left\{ \max_{\tau\in\mathrm{T}_L} \frac{1}{\sqrt{m}}\sum_{j=1}^{m} Y^{(j)}_{L;\tau} \geqslant \kappa_\alpha \left(\frac{\epsilon}{D_\tau\sqrt{2|\mathrm{T}_L|\log(2/\delta)}} \vee 1\right)\sqrt{\log|\mathrm{T}_L|}\right\} \tag{21}$$

then satisfies $\mathbb{P}_0 T_{\mathrm{I}} \leqslant \alpha$ via a union bound and sub-exponential tail bound, we defer the reader to the proof of Lemma 2 for details. Furthermore, for $f \in \mathcal{B}^{s,R}_{p,q}$, we have $\|\Pi_L f\|_2 \leqslant \|f\|_2 \lesssim R$. If f in addition satisfies (19), there exists $\tau^* \in \mathrm{T}_L$ such that (18) is satisfied and consequently

$$\mathbb{P}_f(1-T_{\mathrm{I}}) \leqslant \mathbb{P}_f(1-\varphi_{\tau^*}) \leqslant \alpha/2.$$

The optimal choice of L depends on the regularity level of the signal f, balancing the approximation error $\|f - \Pi_L f\|_2^2$ and the right-hand side of (19), for which we defer the details to Section B.1 in the supplement. To summarize, we have obtained the following lemma.

LEMMA 2. *For all $R > 0$, $\alpha \in (0,1)$ there exists $\kappa_\alpha > 0$ and $C_\alpha > 0$ such that the test T_I defined in (21) satisfies $\mathbb{P}_0 T_I \leqslant \alpha$. Furthermore, if $f \in \mathcal{B}^{s,R}_{p,q}$ is such that for some L and $M_{N,\delta,\tau} = \log(N)\sqrt{\log\log(NR/\sigma)\log(NR/\sigma)\log(1/\delta)}$,*

$$\|\Pi_L f\|_2^2 \geqslant C_\alpha M_{N,\delta,\tau}\left(\frac{\sqrt{2^L}}{\sigma^2\sqrt{N}\sqrt{n}(\sqrt{n}\epsilon \wedge 1)}\right) \bigvee \left(\frac{1}{\sigma^2 N n\epsilon^2}\right),$$

we have that $\mathbb{P}_f(1-T_I) \leqslant \alpha$.

3.2. *Private testing procedure II: high privacy-budget strategy.* In the high-privacy budget regime, we construct a testing procedure that consists essentially of two steps. In the first step, the data is truncated, clipped and averaged over the coordinates, after which Gaussian noise is added to obtain a private summary of the original data. Then, as a second step, the transcripts are averaged, and based on this average, a test statistic that is reminiscent of a chi-square test is computed in the central server. This is in contrast to the strategy of the previous section, where each server computes a (private version of) a chi-square test statistic.

The approach taken here is to divide the servers equally over the first d_L coordinates (i.e. as uniformly as possible), where we recall the notation $d_L := \sum_{l=1}^{L} 2^l$. That is to say, for $L, K_L \in \mathbb{N}$, we partition the coordinates $\{1, \dots, d_L\}$ into approximately d_L/K_L sets of size K_L. The servers are then equally divided over each of these partitions and communicate the sum of the clipped $X^{(j)}_{L;i}$'s coefficients corresponding to their partition, were we also recall that the notation $X^{(j)}_{L;i}$ denotes the vector in $\mathbb{R}^{d_L}$ formed by the first d_L coordinates of $\Pi_L X^{(j)}_i$.

More formally, take $K_L = \lceil n\epsilon^2 \wedge d_L \rceil$ and consider sets $\mathcal{J}_{lk;L} \subset [m]$ for indexes $(l,k) \in \{l = 1, \dots, L, k = 1, \dots, 2^l\} =: I_L$, such that $|\mathcal{J}_{lk;L}| = \lceil \frac{mK_L}{d_L} \rceil$ and each $j \in \{1, \dots, m\}$ is in $\mathcal{J}_{lk;L}$ for at least K_L different indexes $k \in \{1, \dots, d_L\}$. For $(l,k) \in I_L$, $j \in \mathcal{J}_{lk;L}$, generate the transcripts according to

$$Y^{(j)}_{lk;L} | X^{(j)} \equiv Y^{(j)}_{lk;L}(X^{(j)}) = \gamma_L \sum_{i=1}^{n} [\sigma^{-1}(X^{(j)}_i)_{lk}]^{\tau}_{-\tau} + W^{(j)}_{lk} \tag{22}$$

with $\gamma_L = \epsilon/(2\sqrt{2K_L \log(2/\delta)}\tau)$, $\tau = \tilde{\kappa}_\alpha \sqrt{\log(N/\sigma)}$ and $(W^{(j)}_{lk})_{j\in[m],(l,k)\in I_L}$ i.i.d. standard Gaussian noise.

Since $x \mapsto \sum_{i=1}^{n} [\sigma(x^{(j)}_i)_{lk}]^{\tau}_{-\tau}$ has sensitivity bounded by 2τ, for $k = 1, \dots, K$, releasing

$$Y^{(j)}_L(X^{(j)}) = (Y^{(j)}_{L,l_1k_1}(X^{(j)}), \dots, Y^{(j)}_{L,l_{K_L}k_{K_L}}(X^{(j)}))$$

satisfies (ϵ, δ)-DP, see Lemma 23 in the supplement for details.

If the privacy budget were of no concern, submitting the above transcripts with $2^L \asymp N^{1/(2s+1/2)}$ would be sufficient to construct a test statistic that attains the unconstrained rate of $\rho^2 \asymp N^{-2s/(2s+1/2)}$. Under (more stringent) privacy constraints, however, the optimal number of coordinates to be transmitted should depend on the privacy budget. Whenever $\epsilon \lesssim 1/\sqrt{n}$, it turns out that submitting just one coordinate is in fact rate optimal. Sending more than one coordinate leads to worse rates as the noise overpowers the benefit of having a higher dimensional transcript. As ϵ increases, the optimal number of coordinates to be transmitted increases as well. Whenever $\epsilon \gtrsim \sigma^{-\frac{2}{4s+1}} m^{\frac{1}{4s+1}} n^{\frac{1/2-2s}{4s+1}}$, the optimal number of coordinates to be transmitted is $2^L \asymp N^{1/(2s+1/2)}$.

The test

$$T_{\mathrm{II}} = \mathbb{1}\left\{ \frac{1}{\sqrt{d_L}} \sum_{(l,k)\in I_L} \left[\left(\frac{1}{\sqrt{|\mathcal{J}_{lk;L}|}} \sum_{j\in \mathcal{J}_{lk;L}} Y^{(j)}_{lk;L} \right)^2 - \frac{n\epsilon^2}{4K_L\tau^2} - 1 \right] \geqslant \kappa_\alpha \left(\frac{n\epsilon^2}{4K_L\tau^2} \vee 1 \right) \right\} \tag{23}$$

satisfies $\mathbb{P}_0 T_{\mathrm{II}} \leqslant \alpha$ by Lemma 24 in the supplement whenever $\tilde{\kappa}_\alpha > 0$ and $\kappa_\alpha > 0$ are chosen large enough.

The power that the test attains depends on the signal size up until resolution level L, i.e. $\|\Pi_L f\|_2$. Specifically, the test Type II error $\mathbb{P}_f(1 - T_{\mathrm{II}}) \leqslant \alpha$ whenever

$$\|\Pi_L f\|_2^2 \geqslant C_\alpha \frac{\log\log(N)\log(N)\log(1/\delta) 2^{(3/2)L}}{mn^2\epsilon^2}. \tag{24}$$

The optimal choice of L for is determined by the trade-off between the approximation error $\|f - \Pi_L f\|_2^2$ and the right-hand side of (24). The proof of the following lemma is given in Section B.2 of the supplement.

LEMMA 3. *Take $\alpha \in (0,1)$. Suppose f satisfies (24) and that $\epsilon \geqslant \frac{2^{L+1}}{\sqrt{mn}}$ for some $L \in \mathbb{N}$. Then, the distributed (ϵ,δ)-DP testing protocol T_{II} of level α has Type II error $\mathbb{P}_f(1-T) \leqslant \alpha$ for a large enough constant $C_\alpha > 0$ and $\tilde{\kappa}_\alpha > 0$, depending only on α.*

3.3. *Private testing procedure III: high privacy-budget shared randomness strategy.* In this section, we construct a testing procedure that is based on the same principles as the one in the previous section, but with the difference that the servers a source of randomness. The transcripts are still based on the clipped and averaged coordinates of the truncated data, but instead of dividing the servers across the coordinates, we apply the same random rotation across the servers.

Next, we describe the testing procedure in detail. Consider for $L \in \mathbb{N}$ the quantities $d_L = \sum_{l=1}^L 2^l$ and $K_L = \lceil n\epsilon^2 \wedge d_L \rceil$ and let U_L denote a random rotation uniformly drawn (i.e. from the Haar measure) on the group of random orthonormal $d_L \times d_L$-matrices.

For $I_L := \{(l,k) : l = 1, \ldots, \lceil \log_2(K_L) \rceil,\, k = 1, \ldots, 2^l\}$, $(l,k) \in I_L$ and $j = 1, \ldots, m$, generate the transcripts according to

$$Y_{lk;L}^{(j)} | (X^{(j)}, U) = \gamma_L \sum_{i=1}^n [(UX_{L;i}^{(j)})_{lk}]_{-\tau}^{\tau} + W_{lk}^{(j)}, \tag{25}$$

with $\gamma_L = \frac{\epsilon}{2\sqrt{2K_L \log(2/\delta)\log(N)\tau}}$, $\tau = \tilde{\kappa}_\alpha \sqrt{\log(N/\sigma)}$, $\tilde{\kappa}_\alpha > 0$ and $(W_l^{(j)})_{j,l}$ i.i.d. centered standard Gaussian noise. By an application of Lemma 25, the transcript $Y_L^{(j)} := (Y_{lk;L}^{(j)})_{(l,k)\in I_L}$ is (ϵ,δ)-differentially private.

In the shared randomness strategy above, we essentially only send the first $\sum_{l=1}^{\lceil \log_2(K_L) \rceil} 2^l$ coordinates. The random rotation U_L ensures that, roughly speaking, a sufficient amount of the signal is present in these first coordinates, with high probability.

We then construct the test

$$T_{\text{III}} = \mathbb{1}\left\{ \frac{1}{\sqrt{K_L}} \sum_{(l,k)\in I_L} \left[\left(\frac{1}{\sqrt{m}} \sum_{j=1}^m Y_{lk;L}^{(j)} \right)^2 - n\gamma_L^2 - 1 \right] \geqslant \kappa_\alpha \left(n\gamma_L^2 \vee 1 \right) \right\}, \tag{26}$$

which satisfies $\mathbb{P}_0 \varphi \leqslant \alpha/2$ by Lemma 26 in the supplement, for $\kappa_\alpha > 0$ large enough. The lemma below is proven in Section B of the supplement, and yields that the Type II error of the test satisfies $\mathbb{P}_f(1 - T_{\text{III}}) \leqslant \alpha$ whenever the coordinates up to resolution level L are of sufficient size. The optimal value for L depends on the truncation level s, and is chosen by balancing the approximation error $\|f - \Pi_L f\|_2^2$ and the right-hand side of (27), we defer the reader to Section B.3 of the supplement for details.

LEMMA 4. *The testing protocol T_{III}, with level α and has corresponding Type II error probability $\mathbb{P}_f(1 - T_{III}) \leqslant \alpha$ whenever*

$$\|\Pi_L f\|_2^2 \geqslant C_\alpha \frac{2^L \log(1/\delta)\log(N)}{mn\sqrt{n\epsilon^2 \wedge 2^L}\sqrt{n\epsilon^2 \wedge 1}} \tag{27}$$

for constant $C_\alpha > 0$ and $\tilde{\kappa}_\alpha > 0$ depending only on α.

4. Adaptive tests under DP constraints. In the previous section we have derived methods that match (up to logarithmic factors) the theoretical lower bound established in Section 2. The proposed tests, however, depend on the regularity parameter s of the functional parameter of interest f.

In this section we derive an distributed (ϵ, δ)-DP testing protocol that adapts to the regularity when it is unknown. This method attains the optimal rate of Theorem 3, and consequently proves the aforementioned theorem.

The adaptive procedure builds on the tests constructed in Section 3 and combines them using essentially a multiple testing strategy. Roughly speaking, the method consists of taking approximately a $1/\log N$-mesh-size grid in the regularity interval $[s_{\min}, s_{\max}]$, constructing optimal tests for each of the grid points and combining them using a type of Bonferroni's correction. By design, the tests constructed in Section 3 are based on sub-exponential private test statistics, which allows a combination of the test statistics with a Bonferroni correction of the order of $\log\log N$.

Combining $\log N$ many (ϵ', δ)-differentially private transcripts using Gaussian mechanisms, results in a (ϵ, δ)-differentially private protocol, with $\epsilon = \epsilon'\sqrt{\log N}$. This means that the erosion of the privacy budget by conducting a test for each grid-point is limited to a logarithmic factor, means the method greatly improves over the potentially polynomially worse rate of a non-adaptive method.

The detailed adaptive testing procedures are given as follows. Let ρ_s equal the right-hand side of (7) in case there is access to local randomness only, or the right-hand side of (6) in case shared randomness is available. Let $L_s = \lfloor s^{-1}\log_2(1/\rho_s)\rfloor \vee 1$ and define furthermore $\mathcal{S} := \{L_{s_{\min}}, \dots, L_{s_{\max}}\}$ such that $L_s \in \mathcal{S}$ for all $s \in [s_{\min}, s_{\max}]$. Furthermore, we note that the resulting "collection of resolution levels" satisfies $|\mathcal{S}| \leqslant C_{s_{\max}} \log N$ for some constant $C_{s_{\max}} > 0$ depending only on $s_{\max}$.

Consider the first the case without access to shared randomness. We partition the collection of resolution levels $\mathcal{S}$, depending on the model characteristics, as follows.

$$\mathcal{S}_{\texttt{LR}}^{\texttt{LOW}} = \left\{ L \in \mathcal{S} : 2^L \leqslant \epsilon\sqrt{mn}(1 + \sqrt{n}\mathbb{1}_{\{\sqrt{n}\epsilon > 1\}}) \right\}, \quad \mathcal{S}_{\texttt{LR}}^{\texttt{HIGH}} = \mathcal{S} \backslash \mathcal{S}_{\texttt{LR}}^{\texttt{LOW}}. \tag{28}$$

If the true regularity s_0 is such that $L_{s_0} \in \mathcal{S}^{\texttt{LOW}}$, the low privacy-budget test of Section 3.1 (with $L = L_{s_0}$) is a rate optimal strategy. If $L_{s_0} \in \mathcal{S}^{\texttt{HIGH}}$, the high privacy-budget test of Section 3.2 is rate optimal.

For the case of shared randomness, the phase transitions occur for different values of $s \in [s_{\min}, s_{\max}]$, or their respective resolution levels L_s. So in this case, we partition the collection of resolution levels as

$$\mathcal{S}_{\texttt{SHR}}^{\texttt{LOW}} = \left\{ L \in \mathcal{S} : 2^L \leqslant \epsilon^2 mn \right\}, \quad \mathcal{S}_{\texttt{SHR}}^{\texttt{HIGH}} = \mathcal{S} \backslash \mathcal{S}_{\texttt{SHR}}^{\texttt{LOW}}. \tag{29}$$

Consider some $\mathcal{S}' \subset \mathcal{S}$. The "adaptive version" of the low privacy-budget test defined in (21) takes the form

$$T_{\mathrm{I}}^{S} := \mathbb{1}\left\{ \max_{L \in \mathcal{S}', \tau \in \mathrm{T}_L} \frac{1}{\sqrt{m}\,(\gamma_L \vee 1)\sqrt{\log|\mathrm{T}_L||\mathcal{S}'|}} \sum_{j=1}^{m} Y_{L;\tau}^{(j)} \geqslant \kappa_\alpha \right\}, \tag{30}$$

where T_L is as defined in (20) and $Y_L^{(j)} = \{Y_{L;\tau}^{(j)} : \tau \in \mathrm{T}_L\}$ is generated according to (16) for $L \in S$ with

$$\gamma_\tau = \frac{\epsilon}{2D_\tau\sqrt{|\mathrm{T}_L||\mathcal{S}'|\log(4/\delta)}}.$$

The above choice of ϵ yields that $(Y_L^{(j)})_{L \in S}$ is $(\epsilon/2, \delta/2)$-DP due to the Gaussian mechanism. The enlargement of the critical region, which is now effectively rescaled by $\sqrt{\log|\mathrm{T}|_L|\mathcal{S}'|}$

instead of $\sqrt{\log |\mathrm{T}|_L}$, accounts for the potentially larger set of test statistics over which the maximum is taken. In the case of having access only to local sources of randomness, we set $\mathcal{S}' = \mathcal{S}_{\text{LR}}^{\texttt{LOW}}$. If $\mathcal{S}_{\text{LR}}^{\texttt{LOW}}$ is empty, we set $T_{\text{I}} = 0$ instead, which forms an $(0, 0)$-differentially private protocol.

In the case of having access to local sources of randomness only; if $\mathcal{S}_{\text{LR}}^{\texttt{HIGH}}$ is non-empty, the adaptive version of the high privacy-budget test defined in (B.3) is given by

(31)

$$T_{\text{II}} = \mathbb{1}\left\{ \max_{L \in \mathcal{S}_{\text{LR}}^{\texttt{HIGH}}} \frac{1}{\sqrt{d_L}\,(\eta_L \vee 1)} \sum_{(l,k)\in I_L} \left[\left(\frac{1}{\sqrt{|\mathcal{J}_{lk;L}|}} \sum_{j \in \mathcal{J}_{lk;L}} Y_{lk;L}^{(j)} \right)^2 - \eta_L - 1 \right] \geqslant \kappa_\alpha \sqrt{\log |\mathcal{S}^{\texttt{HIGH}}|} \right\},$$

where the transcripts are generated according to (S.74) for $L \in \mathcal{S}^{\texttt{HIGH}}$, with $\gamma_L = \epsilon / (4\sqrt{|\mathcal{S}^{\texttt{HIGH}}| K_L \log(4/\delta)}\tau)$, $\eta_L = \frac{n\epsilon^2}{4K_L\tau^2}$, $\tau = \tilde{\kappa}_\alpha \sqrt{\log(N/\sigma)}$. Due to the Gaussian mechanism, the transcripts satisfy an $(\epsilon/2, \delta/2)$-DP constraint. As before, if $\mathcal{S}^{\texttt{HIGH}}$ is empty, set $T_{\text{II}} = 0$ instead.

In the case of having access to local randomness only, the adaptive testing procedure then consists of computing the tests $T_{\text{I}}^{\mathcal{S}^{\texttt{LOW}}}$ and T_{II}, for which the released transcripts satisfy (ϵ, δ)-DP. The final test is then given by

$$T = T_{\text{I}}^{\mathcal{S}^{\texttt{LOW}}} \vee T_{\text{II}}. \tag{32}$$

In Section B in the supplement, it is shown that this test is adaptive and rate optimal (up to logarithmic factors), proving the first part of Theorem 3.

In case of shared randomness, the adaptive version of the high privacy-budget test defined in (S.84) is given by

(33)

$$T_{\text{III}} = \mathbb{1}\left\{ \max_{L \in S_{\text{SHR}}^{\texttt{HIGH}}} \frac{1}{\sqrt{K_L}\left(n\gamma_L^2 \vee 1\right)} \sum_{(l,k)\in I_L} \left[\left(\frac{1}{\sqrt{m}} \sum_{j=1}^{m} Y_{lk;L}^{(j)} \right)^2 - n\gamma_L^2 - 1 \right] \geqslant \kappa_\alpha \sqrt{\log |\mathcal{S}|} \right\},$$

where the transcripts are generated according to (S.83) for $L \in S_{\text{SHR}}^{\texttt{HIGH}}$, $\gamma_L = \frac{\epsilon}{4\sqrt{K_L |S_{\text{SHR}}^{\texttt{HIGH}}| \log(4/\delta) \log(N)}\tau}$, $\tau = \tilde{\kappa}_\alpha \sqrt{\log(N/\sigma)}$. By similar reasoning as earlier, the transcripts $\{Y_L^{(j)} : L \in \S_{\text{SHR}}^{\texttt{HIGH}}\}$ are $(\epsilon/2, \delta/2)$-DP. If $S_{\text{SHR}}^{\texttt{HIGH}}$ is empty, we set $T_{\text{III}} = 0$ instead.

The adaptive testing procedure in the case of shared randomness then consists of computing the tests $T_{\text{I}}^{S^{\texttt{LOW}}}$ and T_{III}, for which the released transcripts satisfy (ϵ, δ)-DP. The final test is then given by

$$T = T_{\text{I}}^{S_{\text{SHR}}^{\texttt{LOW}}} \vee T_{\text{III}}. \tag{34}$$

In the supplement's Section B, we prove that this test is adaptive, attaining the optimal rate for shared randomness protocols (up to logarithmic factors), giving us the second statement Theorem 3.

5. The minimax private testing lower bound. In this section, we present a single theorem outlining the lower bound for the detection threshold for distributed testing protocols that adhere to DP constraints, with and without the use of shared randomness. The theorem directly yields the "lower bound part" of Theorems 2 and 1 presented in Section 2. In conjunction with Theorem 4, the theorem shows that the tests constructed in Section 3 are rate optimal up to logarithmic factors.

THEOREM 5. *Let $s, R > 0$ be given. For all $\alpha \in (0,1)$, there exists a constant $c_\alpha > 0$ such that if*

$$\rho^2 \leqslant c_\alpha \left(\frac{\sigma^2}{mn}\right)^{\frac{2s}{2s+1/2}} + \left(\frac{\sigma^2}{mn^2\epsilon^2}\right)^{\frac{2s}{2s+3/2}} \wedge \left(\left(\frac{\sigma^2}{\sqrt{mn}\sqrt{1 \wedge n\epsilon^2}}\right)^{\frac{2s}{2s+1/2}} + \left(\frac{\sigma^2}{mn^2\epsilon^2}\right)\right), \tag{35}$$

it holds that

$$\inf_{T \in \mathscr{T}^{(\epsilon,\delta)}} \mathcal{R}(H_\rho^{s,R}, T) > 1 - \alpha, \tag{36}$$

for all natural numbers m, N and $n = N/m$, $\sigma > 0$, $\epsilon \in (N^{-1}, 1]$ and $\delta \leqslant N^{-(1+\omega)}$ for any constant $\omega > 0$.

Similarly, for any $\alpha \in (0,1)$, there exists a constant $c_\alpha > 0$ such that if

$$\rho^2 \leqslant c_\alpha \left(\frac{\sigma^2}{mn}\right)^{\frac{2s}{2s+1/2}} + \left(\frac{\sigma^2}{mn^{3/2}\epsilon\sqrt{1 \wedge n\epsilon^2}}\right)^{\frac{2s}{2s+1}} \wedge \left(\left(\frac{\sigma^2}{\sqrt{mn}}\right)^{\frac{2s}{2s+1/2}} + \left(\frac{\sigma^2}{mn^2\epsilon^2}\right)\right), \tag{37}$$

we have that there exists a distributed (ϵ,δ)-DP shared randomness testing protocol $T \equiv T_{m,n,s,\sigma}$ such that

$$\inf_{T \in \mathscr{T}_{SHR}^{(\epsilon,\delta)}} \mathcal{R}(H_\rho^{s,R}, T) > 1 - \alpha, \tag{38}$$

for all natural numbers m, N and $n = N/m$, $\sigma > 0$, $\epsilon \in (N^{-1}, 1]$ and $\delta \leqslant N^{-(1+\omega)}$ for any constant $\omega > 0$.

The theorem states that, whenever the signal-to-noise ratio ρ is below a certain threshold times the minimax separation rate, no distributed testing protocol can achieve a combined Type I and Type II error rate below α. Its proof is lengthy and involves a combination of various techniques. We defer the full details of the proof to Section A of the supplement, but provide an overview of the main steps below.

For Steps 1, 2 and 3, there is no distinction between local and shared randomness. We use the same notation for distributed protocols in these steps, but simply assume U is degenerate in the case of local randomness.

Step 1: The first step is standard in minimax testing analysis: we lower bound the testing risk by a Bayes risk,

$$\inf_{T \in \mathscr{T}} \mathcal{R}(H_\rho, T) \geqslant \inf_{T \in \mathscr{T}} \sup_{\pi} \left(\mathbb{P}_0(T(Y) = 1) + \int \mathbb{P}_f(T(Y) = 0) d\pi(f) - \pi(H_\rho^c)\right), \tag{39}$$

where $\mathscr{T}$ denotes either the class of local randomness or shared randomness (ϵ,δ)-DP protocols. This inequality allows the prior π to be chosen adversarially to the distribution of the transcripts. This turns out to be crucial in the context of local randomness protocols, as is further highlighted in Step 4. The specific prior distribution is chosen to be a centered Gaussian distribution, with a finite rank covariance, where the rank is of the order 2^L, for some $L \in \mathbb{N}$. This covariance is constructed in a way that it puts most of its mass in the dimensions in which the privacy protocol is the least informative, whilst at the same time it assures that the probability mass outside of the alternative hypothesis $\pi(H_\rho^c)$ is small. The particular choice for a Gaussian prior (instead of e.g. the two point prior in [45]) is motivated by Step 3.

Step 2: In this step, we approximate the distribution of the transcripts with another distribution that results in approximately the same testing risk, but has two particular favorable properties for our purposes.

- Whenever the distribution of a transcript satisfies an (ϵ, δ)-DP constraint with $\delta > 0$, the transcript's density can be unbounded on a set of small probability mass (proportional to δ). Consequently, the local likelihoods of the transcripts can have erratic behavior in the tails. To remedy this, we consider approximations to the transcripts that have bounded likelihoods, that differ only on a set that is negligible in terms of its impact on the testing risk. Furthermore, these approximating transcripts satisfy a $(\epsilon, 2\delta)$-DP privacy constraint. These bounded likelihoods enable the argument of Step 5.
- Similarly, whenever $\delta > 0$, the distribution of the data conditionally on the transcript potentially has an unbounded density. For the argument employed in Step 3, we require a uniform abound on the density of the distribution of $X|Y$. We mitigate this by approximating the distribution of the transcripts by a distribution that induces a bounded density for the data conditionally on the transcript. This approximation of the original transcript satisfies a $(\epsilon, 3\delta)$-DP constraint.

Furthermore, we show that both approximations can be done in a way that the approximating transcript distribution is $(\epsilon, 6\delta)$-DP.

Step 3: By standard arguments, on can further lower bound the testing risk in (39) for a particular transcript distribution $\mathbb{P}^{Y|X,U} = \bigotimes_{j=1}^{m} \mathbb{P}^{Y^{(j)}|X^{(j)},U}$ and prior distribution π by a quantity depending on the chi-square divergence between $\mathbb{P}_{\pi}^{Y|U=u}$ and $\mathbb{P}_{0}^{Y|U=u}$;

$$1 - \left(\sqrt{(1/2) \int \mathbb{E}_0^{Y|U=u} \left(\left(\frac{d\mathbb{P}_{\pi}^{Y|U=u}}{d\mathbb{P}_{0}^{Y|U=u}} \right)^2 - 1 \right)^2 d\mathbb{P}^U(u) + \pi(H_\rho^c)} \right). \tag{40}$$

The likelihood ratio of the transcripts depends on the privacy protocol, and is difficult to analyze directly. We employ the technique developed in [77]. Specifically, Lemma 10.1 in [77], which states, roughly speaking, that the inequality

$$\mathbb{E}_0^{Y|U=u} \left(\frac{d\mathbb{P}_{\pi}^{Y|U=u}}{d\mathbb{P}_{0}^{Y|U=u}} \right)^2 \leqslant G \prod_{j=1}^{m} \mathbb{E}_0^{Y^{(j)}|U=u} \left(\frac{d\mathbb{P}_{\pi}^{Y^{(j)}|U=u}}{d\mathbb{P}_{0}^{Y^{(j)}|U=u}} \right)^2 \tag{41}$$

holds for a finite constant $0 < G < \infty$ and equality with the smallest possible G is attained whenever the conditional distribution of the data given the transcripts is Gaussian in an appropriate sense (we defer the details here to Section A.3 in the supplement). This result is a type of Brascamp-Lieb inequality [16, 60]. There is an existing literature on Brascamp-Lieb inequality in relation to information theoretical problems, in relation to mutual information [28, 61, 62], in addition to the communication constraint testing problem in [77]. That (41) has a "Gaussian maximizer" allows tractable analysis of the chi-square divergence in (40), yielding that the latter display is further lower bounded by

$$1 - \sqrt{(1/2) \int (\mathrm{A}_u^{\pi} \mathrm{B}_u^{\pi} - 1) \, d\mathbb{P}^U(u) + \pi(H_\rho^c)}, \tag{42}$$

where

$$\mathrm{A}_u^{\pi} := \int e^{f^\top \sum_{j=1}^{m} \Xi_u^j g} d(\pi \times \pi)(f, g), \quad \mathrm{B}_u^{\pi} := \prod_{j=1}^{m} \mathbb{E}_0^{Y^{(j)}|U=u} \left(\frac{d\mathbb{P}_{\pi}^{Y^{(j)}|U=u}}{d\mathbb{P}_{0}^{Y^{(j)}|U=u}} \right), \tag{43}$$

where Ξ_u^j denotes the covariance of (a subset of) the data $X_L^{(j)}$ (defined as in (44)) conditionally on the transcript $Y^{(j)}$ and $U = u$;

$$(44)\quad \Xi_u^j := \mathbb{E}_0^{Y^{(j)}|U=u} \mathbb{E}_0 \left[\sum_{i=1}^n \sigma^{-1} X_{L;i}^{(j)} \middle| Y^{(j)}, U = u \right] \mathbb{E}_0 \left[\sum_{i=1}^n \sigma^{-1} X_{L;i}^{(j)} \middle| Y^{(j)}, U = u \right]^\top .$$

Whilst the quantities A_u and B_u^π are still not fully tractable, sharp bounds for both are possible and form the content of Steps 4 and 5, respectively.

Step 4: As remarked earlier, class of local randomness protocols is a strictly smaller class. To attain the sharper (i.e. larger) lower bound for local randomness protocols, we exploit the fact that Step 2 allows us to choose the prior adversarially to the distribution of the transcripts. In particular, since U is degenerate in the case of local randomness only, this means that the covariance of π to be more diffuse in the directions in which Ξ_u^j is the smallest. When considering shared randomness protocols, U is not degenerate, and the lower bound follows by taking the covariance of π to be an order 2^L-rank approximation of the identity map on $\ell_2(\mathbb{N})$. The bounds for A_u^π are

$$(45)\qquad \mathrm{A}_u^\pi = \exp\left(C \frac{\rho^4}{c_\alpha 2^{3L}} \mathrm{Tr}\left(\Xi_u\right)^2 \right) \quad \text{and} \quad \mathrm{A}_u^\pi \leqslant \exp\left(C \frac{\rho^4}{c_\alpha 2^{2L}} \|\Xi_u\| \mathrm{Tr}\left(\Xi_u\right) \right)$$

for local randomness protocols and shared randomness protocols, respectively.

Step 5: So far, Steps 1-4 have not used the fact that the transcripts are necessarily less informative than the original data, as a consequence of the transcripts being (ϵ, δ)-DP. In this step, we exploit the privacy constraint to argue that A_u^π and B_u^π are small at the detection boundary for ρ.

In order to capture the information loss due to privacy in A_u^π, it suffices to bound the trace and operator norm of Ξ_u. The quantity Ξ_u can be seen as the Fisher information of the finite dimensional submodel spanned by the covariance of π. This quantity, loosely speaking, captures how much information the transcript contains on the original data. In order to analyze Ξ_u, we rely on a "score attack" type of technique, as employed in [23, 20].

The quantity B_u^π corresponds to the (product of) the local likelihoods of the transcripts. Whenever $\epsilon \geqslant 1/\sqrt{n}$, it suffices to consider the trivial bound

$$(46)\qquad \mathbb{E}_0^{Y^{(j)}|U=u} \left(\frac{d\mathbb{P}_\pi^{Y^{(j)}|U=u}}{d\mathbb{P}_0^{Y^{(j)}|U=u}} \right) \leqslant \mathbb{E}_0^{X^{(j)}|U=u} \left(\frac{d\mathbb{P}_\pi^{X^{(j)}}}{d\mathbb{P}_0^{X^{(j)}}} \right),$$

and further bounding the right-hand side without privacy specific arguments. Whenever $\epsilon < 1/\sqrt{n}$, more sophisticated methods are need to capture the effect of privacy. Our argument uses a coupling method, which, combined with the fact that the likelihoods of the transcripts are bounded in our construction, allows us to obtain a sharp bound for B_u^π. After obtaining the bounds in terms of the rank 2^L and ρ, the proof is finished by choosing L such that the second and third term in (40) are balanced (minimizing their sum).

6. Discussion.

The findings in this paper highlight the trade-off between statistical accuracy and privacy in federated goodness-of-fit testing under differential privacy (DP) constraints. We characterize the problem in terms of the minimax separation rate, which quantifies the difficulty of the testing problem based on the regularity of the underlying function, the sample size, the degree of data distribution, and the stringency of the DP constraint. The minimax separation rate varies depending on whether the distributed testing protocol has access to local or shared randomness. Furthermore, we construct data-driven adaptive testing procedures that achieve the same optimal performance, up to logarithmic factors, even when the regularity of the functional parameter is unknown.

One possible extension of this work is to consider a more general distribution of the privacy budget across the servers. Our current analysis supports differing budgets to the extent that $\epsilon_j \asymp \epsilon_k$, $\delta_j \asymp \delta_k$, and $n_j \asymp n_k$. However, one could explore more heterogeneous settings where severs differ significantly in their differential privacy constraints and number of observations. Although this would complicate the presentation of results, the techniques developed in this paper could, in principle, be extended to such settings.

Another interesting direction is to consider multiple testing problems, where the goal is to test multiple hypotheses simultaneously. We anticipate that the framework, insights, and theoretical results provided in the current paper will serve as valuable resources for future studies in this domain.

Regarding adaptation, not much is known about the cost of privacy outside the local DP setting (i.e., one observation per server; $n = 1$ in our context). Interestingly, the cost of adaptation is minimal in the privacy setting considered in this paper. It remains an open question whether this minimal cost is a general phenomenon, whether it can be characterized exactly, or whether the cost of adaptation is more severe in other settings. We leave these questions for future research.

SUPPLEMENTARY MATERIAL

Supplementary Material to "Federated Nonparametric Hypothesis Testing with Differential Privacy Constraints: Optimal Rates and Adaptive Tests"
In this supplement, we present the detailed proofs for the main results in the paper "Federated Nonparametric Hypothesis Testing with Differential Privacy Constraints: Optimal Rates and Adaptive Tests".

Funding. The research was supported in part by NIH grants R01-GM123056 and R01-GM129781.

REFERENCES

[1] ACHARYA, J., BONAWITZ, K., KAIROUZ, P., RAMAGE, D. and SUN, Z. (2020). Context Aware Local Differential Privacy. In *Proceedings of the 37th International Conference on Machine Learning* (H. D. III and A. SINGH, eds.). *Proceedings of Machine Learning Research* **119** 52–62. PMLR.

[2] ACHARYA, J., CANONNE, C., FREITAG, C. and TYAGI, H. (2019). Test without Trust: Optimal Locally Private Distribution Testing. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics* (K. CHAUDHURI and M. SUGIYAMA, eds.). *Proceedings of Machine Learning Research* **89** 2067–2076. PMLR.

[3] ACHARYA, J., CANONNE, C. L., FREITAG, C., SUN, Z. and TYAGI, H. (2021). Inference Under Information Constraints III: Local Privacy Constraints. *IEEE Journal on Selected Areas in Information Theory* **2** 253-267. https://doi.org/10.1109/JSAIT.2021.3053569

[4] ACHARYA, J., CANONNE, C. L. and TYAGI, H. (2020). Inference Under Information Constraints I: Lower Bounds From Chi-Square Contraction. *IEEE Transactions on Information Theory* **66** 7835-7855. https://doi.org/10.1109/TIT.2020.3028440

[5] ACHARYA, J., CANONNE, C. L. and TYAGI, H. (2020). Inference under information constraints II: Communication constraints and shared randomness. *IEEE Transactions on Information Theory* **66** 7856–7877.

[6] ACHARYA, J., LIU, Y. and SUN, Z. (2023). Discrete Distribution Estimation under User-level Local Differential Privacy. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics* (F. RUIZ, J. DY and J.-W. VAN DE MEENT, eds.). *Proceedings of Machine Learning Research* **206** 8561–8585. PMLR.

[7] ACHARYA, J., SUN, Z. and ZHANG, H. (2018). Differentially private testing of identity and closeness of discrete distributions. *Advances in Neural Information Processing Systems* **31**.

[8] ACHARYA, J., SUN, Z. and ZHANG, H. (2018). Differentially Private Testing of Identity and Closeness of Discrete Distributions. In *Advances in Neural Information Processing Systems* (S. BENGIO, H. WALLACH, H. LAROCHELLE, K. GRAUMAN, N. CESA-BIANCHI and R. GARNETT, eds.) **31**. Curran Associates, Inc.

[9] ALABI, D. and VADHAN, S. (2022). Hypothesis Testing for Differentially Private Linear Regression. In *Advances in Neural Information Processing Systems* (S. KOYEJO, S. MOHAMED, A. AGARWAL, D. BELGRAVE, K. CHO and A. OH, eds.) **35** 14196–14209. Curran Associates, Inc.

[10] AN, K. (1933). Sulla determinazione empirica di una legge didistribuzione. *Giorn Dell'inst Ital Degli Att* **4** 89–91.

[11] ARACHCHIGE, P. C. M., BERTOK, P., KHALIL, I., LIU, D., CAMTEPE, S. and ATIQUZZAMAN, M. (2019). Local differential privacy for deep learning. *IEEE Internet of Things Journal* **7** 5827–5842.

[12] BASSILY, R., SMITH, A. and THAKURTA, A. (2014). Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th annual symposium on foundations of computer science* 464–473. IEEE.

[13] BEAUFAYS, F., RAO, K., MATHEWS, R. and RAMASWAMY, S. (2019). Federated learning for emoji prediction in a mobile keyboard.

[14] BERRETT, T. and BUTUCEA, C. (2020). Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. *Advances in Neural Information Processing Systems* **33** 3164–3173.

[15] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration inequalities: a nonasymptotic theory of independence*, 1st ed ed. Oxford University Press, Oxford. OCLC: ocn818449985.

[16] BRASCAMP, H. J. and LIEB, E. H. (1976). Best constants in Young's inequality, its converse, and its generalization to more than three functions. *Advances in Mathematics* **20** 151–173.

[17] BROWN, G., HOPKINS, S. and SMITH, A. (2023). Fast, Sample-Efficient, Affine-Invariant Private Mean and Covariance Estimation for Subgaussian Distributions. In *Proceedings of Thirty Sixth Conference on Learning Theory* (G. NEU and L. ROSASCO, eds.). *Proceedings of Machine Learning Research* **195** 5578–5579. PMLR.

[18] BUTUCEA, C., DUBOIS, A., KROLL, M. and SAUMARD, A. (2020). Local differential privacy: Elbow effect in optimal density estimation and adaptation over Besov ellipsoids. *Bernoulli* **26** 1727 – 1764. https://doi.org/10.3150/19-BEJ1165

[19] BUTUCEA, C., ROHDE, A. and STEINBERGER, L. (2023). Interactive versus noninteractive locally differentially private estimation: Two elbows for the quadratic functional. *Annals of Statistics* **51**. https://doi.org/10.1214/22-AOS2254

[20] CAI, T. T., CHAKRABORTY, A. and VUURSTEEN, L. (2023). Optimal federated learning for nonparametric regression with heterogenous distributed differential privacy constraints. *preprint*.

[21] CAI, T. T., CHAKRABORTY, A. and VUURSTEEN, L. (2024). Supplement to "Federated nonparametric hypothesis testing with differential privacy constraints: Optimal rates and adaptive tests".

[22] CAI, T. T., WANG, Y. and ZHANG, L. (2021). The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *The Annals of Statistics* **49** 2825–2850.

[23] CAI, T. T., WANG, Y. and ZHANG, L. (2023). Score attack: A lower bound technique for optimal differentially private learning. *arXiv preprint arXiv:2303.07152*.

[24] CAI, T. T., XIA, D. and ZHA, M. (2024). Optimal differentially private PCA and estimation for spiked covariance matrices. *arXiv preprint arXiv:2401.03820*.

[25] CANONNE, C. L., KAMATH, G., MCMILLAN, A., SMITH, A. and ULLMAN, J. (2019). The structure of optimal private tests for simple hypotheses. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* 310–321.

[26] CANONNE, C. L., KAMATH, G., MCMILLAN, A., ULLMAN, J. and ZAKYNTHINOU, L. (2020). Private Identity Testing for High-Dimensional Distributions. In *Advances in Neural Information Processing Systems* (H. LAROCHELLE, M. RANZATO, R. HADSELL, M. F. BALCAN and H. LIN, eds.) **33** 10099–10111. Curran Associates, Inc.

[27] CANONNE, C. L. and SUN, Y. (2023). Private Distribution Testing with Heterogeneous Constraints: Your Epsilon Might Not Be Mine. *arXiv preprint arXiv:2309.06068*.

[28] CARLEN, E. A. and CORDERO-ERAUSQUIN, D. (2008). Subadditivity of the entropy and its relation to Brascamp-Lieb type inequalities. *arXiv:0710.0870 [math]*. arXiv: 0710.0870.

[29] COHEN, A., DAUBECHIES, I. and VIAL, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and computational harmonic analysis*.

[30] CRAMÉR, H. (1928). On the composition of elementary errors: First paper: Mathematical deductions. *Scandinavian Actuarial Journal* **1928** 13–74.

[31] DAUBECHIES, I. (1992). *Ten lectures on wavelets*. SIAM.

[32] DING, B., KULKARNI, J. and YEKHANIN, S. (2017). Collecting telemetry data privately. *Advances in Neural Information Processing Systems* **30**.

[33] DUBOIS, A., BERRETT, T. and BUTUCEA, C. (2023). Goodness-of-Fit Testing for Hölder Continuous Densities Under Local Differential Privacy. In *Foundations of Modern Statistics*. *Springer Proceedings*

in Mathematics & Statistics **PROMS-425** 53-119. Springer International Publishing. https://doi.org/10.1007/978-3-031-30114-8_2

[34] DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science* 429–438. IEEE.

[35] DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2018). Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association* **113** 182–201.

[36] DWORK, C., ROTH, A. et al. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* **9** 211–407.

[37] DWORK, C. and SMITH, A. (2010). Differential privacy for statistics: What we know and what we want to learn. *Journal of Privacy and Confidentiality* **1**.

[38] DWORK, C., SMITH, A., STEINKE, T. and ULLMAN, J. (2017). Exposed! a survey of attacks on private data. *Annual Review of Statistics and Its Application* **4** 61–84.

[39] DWORK, C., TALWAR, K., THAKURTA, A. and ZHANG, L. (2014). Analyze gauss: optimal bounds for privacy-preserving principal component analysis. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing* 11–20.

[40] ERLINGSSON, U., PIHUR, V. and KOROLOVA, A. (2014). RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. *CCS '14* 1054–1067. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/2660267.2660348

[41] ERMAKOV, M. (1990). Asymptotically minimax tests for nonparametric hypotheses concerning the distribution density. *Journal of Soviet Mathematics* **52** 2891–2898.

[42] GABOARDI, M., LIM, H., ROGERS, R. and VADHAN, S. (2016). Differentially private Chi-squared hypothesis testing: Goodness of fit and independence testing. In *Proceedings of The 33rd International Conference on Machine Learning* (M. F. BALCAN and K. Q. WEINBERGER, eds.). *Proceedings of Machine Learning Research* **48** 2111–2120. PMLR, New York, New York, USA.

[43] GINE, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781107337862

[44] HARD, A., RAO, K., MATHEWS, R., RAMASWAMY, S., BEAUFAYS, F., AUGENSTEIN, S., EICHNER, H., KIDDON, C. and RAMAGE, D. (2018). Federated learning for mobile keyboard prediction. *arXiv preprint arXiv:1811.03604*.

[45] INGSTER, Y. and SUSLINA, I. A. (2003). *Nonparametric goodness-of-fit testing under Gaussian models* **169**. Springer Science & Business Media.

[46] INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Math. Methods Statist* **2** 85–114.

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

[48] JOHNSTONE, I. M. (2019). *Function Estimation and Gaussian Sequence Models*. Unpublished manuscript.

[49] KAMATH, G., LI, J., SINGHAL, V. and ULLMAN, J. (2019). Privately learning high-dimensional distributions. In *Conference on Learning Theory* 1853–1902. PMLR.

[50] KAMATH, G., SINGHAL, V. and ULLMAN, J. (2020). Private mean estimation of heavy-tailed distributions. In *Conference on Learning Theory* 2204–2235. PMLR.

[51] KARWA, V. and VADHAN, S. (2017). Finite sample differentially private confidence intervals. *arXiv preprint arXiv:1711.03908*.

[52] KONEČNÝ, J., MCMAHAN, H. B., RAMAGE, D. and RICHTÁRIK, P. (2016). Federated optimization: Distributed machine learning for on-device intelligence. *arXiv preprint arXiv:1610.02527*.

[53] KROLL, M. (2021). On density estimation at a fixed point under local differential privacy. *Electronic Journal of Statistics* **15** 1783 – 1813. https://doi.org/10.1214/21-EJS1830

[54] LALANNE, C., GARIVIER, A. and GRIBONVAL, R. (2023). About the Cost of Central Privacy in Density Estimation. *Transactions on Machine Learning Research*.

[55] LAM-WEIL, J., LAURENT, B. and LOUBES, J.-M. (2022). Minimax optimal goodness-of-fit testing for densities and multinomials under a local differential privacy constraint. *Bernoulli* **28** 579–600.

[56] LEPSKII, O. (1992). Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates. *Theory of Probability & Its Applications* **36** 682–697.

[57] LEVY, D., SUN, Z., AMIN, K., KALE, S., KULESZA, A., MOHRI, M. and SURESH, A. T. (2021). Learning with user-level privacy. *Advances in Neural Information Processing Systems* **34** 12466–12479.

[58] LI, M., TIAN, Y., FENG, Y. and YU, Y. (2024). Federated Transfer Learning with Differential Privacy. *arXiv preprint arXiv:2403.11343*.

[59] LI, T., SAHU, A. K., TALWALKAR, A. and SMITH, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine* **37** 50–60.

[60] LIEB, E. H. (1990). Gaussian kernels have only Gaussian maximizers. *Inventiones Mathematicae* **102** 179–208. Publisher: Springer New York. https://doi.org/10.1007/BF01233426

[61] LIU, J., COURTADE, T. A., CUFF, P. and VERDU, 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

[62] LIU, J., COURTADE, T. A., CUFF, P. and VERDU, S. (2016). Smoothing Brascamp-Lieb Inequalities and Strong Converses for Common Randomness Generation. *arXiv:1602.02216 [cs, math]*. arXiv: 1602.02216.

[63] LIU, Y., SURESH, A. T., YU, F. X. X., KUMAR, S. and RILEY, M. (2020). Learning discrete distributions: user vs item-level privacy. *Advances in Neural Information Processing Systems* **33** 20965–20976.

[64] MCSHANE, E. J. (1934). Extension of range of functions.

[65] NARAYANAN, S. (2022). Private high-dimensional hypothesis testing. In *Conference on Learning Theory* 3979–4027. PMLR.

[66] NARAYANAN, S. (2022). Private High-Dimensional Hypothesis Testing. In *Proceedings of Thirty Fifth Conference on Learning Theory* (P.-L. LOH and M. RAGINSKY, eds.). *Proceedings of Machine Learning Research* **178** 3979–4027. PMLR.

[67] NARAYANAN, S., MIRROKNI, V. and ESFANDIARI, H. (2022). Tight and robust private mean estimation with few users. In *International Conference on Machine Learning* 16383–16412. PMLR.

[68] NGUYEN, A., DO, T., TRAN, M., NGUYEN, B. X., DUONG, C., PHAN, T., TJIPUTRA, E. and TRAN, Q. D. (2022). Deep federated learning for autonomous driving. In *2022 IEEE Intelligent Vehicles Symposium (IV)* 1824–1830. IEEE.

[69] PETROV, V. V. (2022). Sums of independent random variables. In *Sums of Independent Random Variables* De Gruyter.

[70] RODRIGUEZ, I. M., SEXTON12, W. N., SINGER, P. E. and VILHUBER, L. The modernization of statistical disclosure limitation at the US Census Bureau.

[71] SART, M. (2023). Density estimation under local differential privacy and Hellinger loss. *Bernoulli* **29** 2318 – 2341. https://doi.org/10.3150/22-BEJ1543

[72] SHEFFET, O. (2018). Locally private hypothesis testing. In *International Conference on Machine Learning* 4605–4614. PMLR.

[73] SMIRNOV, N. (1948). Table for Estimating the Goodness of Fit of Empirical Distributions. *The Annals of Mathematical Statistics* **19** 279 – 281. https://doi.org/10.1214/aoms/1177730256

[74] SMITH, A. (2011). Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the forty-third annual ACM symposium on Theory of computing* 813–822.

[75] SPOKOINY, V. G. (1996). Adaptive hypothesis testing using wavelets. *The Annals of Statistics* **24**. https://doi.org/10.1214/aos/1032181163

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

[77] SZABÓ, B., VUURSTEEN, L. and VAN ZANTEN, H. (2023). Optimal high-dimensional and nonparametric distributed testing under communication constraints. *The Annals of Statistics* **51** 909 – 934. https://doi.org/10.1214/23-AOS2269

[78] TEAM, A. D. P. (2017). Learning with Privacy at Scale.

[79] THORISSON, H. (2000). *Coupling, Stationarity, and Regeneration*. *Probability and Its Applications*. Springer New York.

[80] TRIEBEL, H. (1992). *Theory of Function Spaces II*. *Monographs in mathematics*. Springer.

[81] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer.

[82] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak convergence*. Springer.

[83] VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*, 1 ed. Cambridge University Press. https://doi.org/10.1017/9781108231596

[84] VON MISES, R. (1928). Statistik und wahrheit. *Julius Springer* **20**.

[85] YE, M. and BARG, A. (2018). Optimal Schemes for Discrete Distribution Estimation Under Locally Differential Privacy. *IEEE Transactions on Information Theory* **64** 5662-5676. https://doi.org/10.1109/TIT.2018.2809790