The Annals of Statistics
2023, Vol. 51, No. 2, 717–737
https://doi.org/10.1214/23-AOS2267

ON ROBUSTNESS AND LOCAL DIFFERENTIAL PRIVACY

BY MENGCHU LI[a], THOMAS B. BERRETT[b] AND YI YU[c]

Department of Statistics, University of Warwick, [a]*mengchu.li@warwick.ac.uk,* [b]*tom.berrett@warwick.ac.uk,* [c]*yi.yu.2@warwick.ac.uk*

It is of soaring demand to develop statistical analysis tools that are robust against contamination as well as preserving individual data owners' privacy. In spite of the fact that both topics host a rich body of literature, to the best of our knowledge, we are the first to systematically study the connections between the optimality under Huber's contamination model and the local differential privacy (LDP) constraints.

In this paper, we start with a general minimax lower bound result, which disentangles the costs of being robust against Huber contamination and preserving LDP. We further study four concrete examples: a two-point testing problem, a potentially diverging mean estimation problem, a nonparametric density estimation problem and a univariate median estimation problem. For each problem, we demonstrate procedures that are optimal in the presence of both contamination and LDP constraints, comment on the connections with the state-of-the-art methods that are only studied under either contamination or privacy constraints, and unveil the connections between robustness and LDP via partially answering whether LDP procedures are robust and whether robust procedures can be efficiently privatised. Overall, our work showcases a promising prospect of joint study for robustness and local differential privacy.

1. Introduction. In modern data collection and analysis, the privacy of individuals is a key concern. There has been a surge of interest in developing data analysis methodologies that yield strong statistical performance without compromising individuals' privacy, largely driven by applications in modern technology, including in Google (e.g., Erlingsson, Pihur and Korolova (2014)), Apple (e.g., Tang et al. (2017)) and Microsoft (e.g., Ding, Kulkarni and Yekhanin (2017)), and by pressure from regulatory bodies (e.g., Forti (2021), Aridor, Che and Salz (2021)). The prevailing framework for the development of private methodology is that of differential privacy (Dwork et al. (2006)). Although this originates in cryptography, there is a growing body of statistical literature that aims to explore the constraints of this framework and provide procedures that make optimal use of available data (e.g., Wasserman and Zhou (2010), Duchi, Jordan and Wainwright (2018), Rohde and Steinberger (2020), Cai, Wang and Zhang (2021)). Work in this area is split between central models of privacy, where there is a third party trusted to collect and analyse data before releasing privatised results, and local models of privacy, where data are randomised before collection. We, in this paper, will consider the local differential privacy constraint, to be formally defined in Section 1.2. While classical methods for locally private analysis are restricted to the estimation of the parameter of a binomial distribution (Warner (1965)), modern research has resulted in mechanisms for many other statistical problems including various hypothesis testing problems (e.g., Kairouz, Oh and Viswanath (2016), Joseph et al. (2019), Berrett and Butucea (2020), Acharya et al. (2022), Lam-Weil, Laurent and Loubes (2022)), mean and median estimation (e.g., Duchi, Jordan and Wainwright (2018)), nonparametric estimation problems (e.g., Rohde and Steinberger (2020), Butucea et al. (2020), Berrett, Györfi and Walk (2021)), and change point analysis (e.g., Berrett and Yu (2021), Li, Berrett and Yu (2022)), to name but a few.

Received January 2022; revised February 2023.
MSC2020 subject classifications. 62C20.
Key words and phrases. Huber's contamination model, local differential privacy, minimax optimality.

In addition to preserving individuals' privacy, being robust to outliers and adversarial contamination is another desideratum for modern learning algorithms. The study of robust statistical procedures has sparked great interest among statisticians and there have been a number illuminating textbooks in this area (e.g., Huber and Ronchetti (2009), Hampel et al. (1986), Huber (2004), Maronna et al. (2019)). The focus of classical robust statistics is on the analysis of outliers' influence on statistical procedures, quantified through specific notions such as the breakdown point and influence function. Due to the demands of analysing more complex data types, the focus has, more recently, shifted towards providing nonasymptotic guarantees on convergence rates under various contamination models, including heavy-tailed models (e.g., Catoni (2012), Lugosi and Mendelson (2019)), different types of model misspecification (e.g., Huber (1968), Cherukuri and Hota (2021)), and strong contamination models (e.g., Diakonikolas et al. (2017), Lugosi and Mendelson (2021), Pensia, Jog and Loh (2020)). A number of computationally efficient algorithms that achieve (nearly) minimax rate-optimal rates under either one or more aforementioned models have also been proposed for various tasks; see Diakonikolas et al. (2019) for a recent survey.

The connections between differential privacy and robustness have been well studied in the central model of differential privacy, where there is a trusted data curator. A natural starting point for many differentially private estimators is a function of the data whose sensitivity to changes in single observations can be controlled (e.g., Dwork et al. (2006), Dwork and Lei (2009), Canonne et al. (2019), Cai, Wang and Zhang (2021)). This is also the case for robust statistics, where estimators are often constructed in order to be minimally sensitive to arbitrary changes in a small number of data points (e.g., Huber and Ronchetti (2009), Huber (2004)). See Avella-Medina (2020) for further discussion of these connections. There has been, recently, an increasing trend, mostly in the theoretical computer science literature, of developing algorithms that are simultaneously robust and privacy-preserving under the central model (e.g., Dimitrakakis et al. (2014), Ghazi et al. (2021), Esfandiari, Mirrokni and Narayanan (2021), Kothari, Manurangsi and Velingker (2021), Liu et al. (2021)). There is, however, little work on the connection between privacy and robustness in the local model of differential privacy. A key distinction between our work and the aforementioned robust procedures in the central model is that there it is possible to add noise after computing robust estimators, while in local privacy the requirement to add noise to each observation separately means that the approaches taken in the two models are fundamentally different. Note that some very recent works (Cheu, Smith and Ullman (2021), Acharya, Sun and Zhang (2021), Chhor and Sentenac (2022)) consider contamination after the privatisation step and the results therein feature an interaction of privacy level α and contamination level ε. In our work, we suppose that contamination happens before the data are sent for privatisation.

1.1. *A summary of our contributions*. This paper concerns the pursuit of answers to the questions:

Q1. *Can robust procedures be directly applied to privatised information and attain optimal performance?*

Q2. *Can locally private procedures be automatically robust?*

To address the aforementioned questions, in this work we will study a range of statistical problems, including hypothesis testing (Section 2), mean estimation (Section 3), nonparametric density estimation (Section 4) and median estimation (Section E of the Supplementary Material, Li, Berrett and Yu (2023)) with data assumed to be generated according to Huber's ε-contamination model (e.g., Huber (1992)), specified in (1).

Starting with Q1, when studying Huber's contamination model without privacy constraints, Chen, Gao and Ren (2016) developed a general theory and showed that a Scheffé

tournament method provides optimal estimators for many problems. Such a method discretises the parameter space and reduces estimation problems to hypothesis selection problems. However, it was shown by Gopi et al. (2020) that hypothesis selection is exponentially more difficult under local privacy constraints, prohibiting the use of this general learning scheme in many specific statistical learning tasks; we discuss this aspect in Section 2.3. Despite this negative answer to Q1 for this general robust procedure, for the specific problems considered in the sequel, we will show that optimal LDP procedures can be regarded as robust procedures applied to privatised data. See Sections 2.3, 3.3 and 4.3 for details.

As for Q2, it turns out that there are deep connections between locally private estimation and estimation within Huber's contamination model. Chen, Gao and Ren (2016) showed that the total variation modulus of continuity, defined in (5), controls the difficulty of a wide range of statistical problems studied under contamination. On the other hand, Rohde and Steinberger (2020) showed that this modulus is also the key quantity in understanding the difficulty of a general class of estimation problems under local privacy. We further explore this relationship and show that suitably chosen procedures for locally private estimation are also optimal when Huber contamination is introduced. More importantly, we see that in specific problems the costs of preserving privacy and contamination are separable, which matches the intuition behind our lower bound result in Proposition 1.

In this paper, we will show that for a range of statistical problems, we are able to find procedures that are simultaneously robust, privacy-preserving and statistically rate-optimal, in terms of the contamination proportion ε, the privacy parameter α, the sample size n and other model parameters that may occur in specific problems.

- Section 2 considers a simple hypothesis testing problem. When contamination is introduced, this becomes a composite hypothesis testing problem where the separation between the hypotheses depends on the level of contamination. We study a combination of the Scheffé test (Devroye and Lugosi (2001)) and the randomised response mechanism (Warner (1965)), that results in a test that has previously been used for hypothesis testing under local differential privacy constraint (e.g., Joseph et al. (2019), Gopi et al. (2020)). We obtain matching upper and lower bounds to prove that this procedure is optimal under Huber contamination and privacy constraints.
- In Section 3, we turn our attention to robust mean estimation, where the inlier distribution has bounded kth central moment for some fixed $k > 1$, and unknown mean in $[-D, D]$ for some potentially diverging $D \geq 1$. We propose a procedure that is minimax rate-optimal in terms of the mean upper bound D, the sample size n, the privacy parameter α and the contamination proportion ε. Previous work on mean estimation under local differential privacy (Duchi, Jordan and Wainwright (2018)) assumes that $D = 1$ and deploys a Laplace privacy mechanism, which we show is sub-optimal when D is large. Previous work has noted the difficulty of private estimation with unbounded parameter spaces, both in the central model of privacy (Brunel and Avella-Medina (2020), Karwa and Vadhan (2017), Kamath et al. (2021)) and the local model (Duchi, Jordan and Wainwright (2013)). It is shown that in the local model uniformly consistent estimation is impossible when $D = \infty$, even without contamination (see Appendix G, Duchi, Jordan and Wainwright (2013)). We derive a phase transition phenomenon, whereby there exists a boundary for D beyond which uniformly consistent estimation is impossible and below which a rate-optimal procedure is available.
- We study nonparametric density estimation problems in Section 4. The procedures we consider are the basis expansion procedures of Duchi, Jordan and Wainwright (2018) and Butucea et al. (2020). We give new analyses to show that these privacy procedures remain minimax rate-optimal when Huber contamination is introduced, for squared-L_2 and L_∞ losses, respectively.

- In Section E of the Supplementary Material (Li, Berrett and Yu (2023)), we study a univariate median estimation problem, deploying a locally private stochastic gradient descent method (e.g., Duchi, Jordan and Wainwright (2018)). Different from the privacy mechanism in the aforementioned three problems, the privacy mechanism here is sequentially interactive. We show that this method is robust against Huber contamination and minimax rate-optimal regarding all the model parameters.

1.2. *General setup.* We will now formally define the framework we study. Let $\mathcal{P}$ denote a class of distributions on the sample space $\mathcal{X}$ and let $\mathcal{G} \supset \mathcal{P}$ denote the set of all distributions on $\mathcal{X}$. Two key ingredients in this paper are: (a) robustness against Huber contamination and (b) privacy preservation in the sense of local differential privacy.

As for the contamination, to be specific, we consider problems where data are generated not directly from $P \in \mathcal{P}$, but from a contaminated distribution $P_\varepsilon \in \mathcal{P}_\varepsilon(\mathcal{P})$, where $\mathcal{P}_\varepsilon(\mathcal{P})$ is defined as

$$\mathcal{P}_\varepsilon(\mathcal{P}) = \big\{P_\varepsilon = (1-\varepsilon)P + \varepsilon G : \varepsilon \in [0,1], P \in \mathcal{P}, G \in \mathcal{G}\big\}. \tag{1}$$

The class of distributions $\mathcal{P}_\varepsilon(\mathcal{P})$ is known as Huber's ε-contamination model (Huber (2004)) in the robust statistics literature.

As for the privacy, formally speaking, a privacy mechanism is a conditional distribution of the privatised data given the raw data, that is, $Q(\cdot|x_1,\dots,x_n)$, $\{x_i\}_{i=1}^n \subset \mathcal{X}$, when the raw data are $\{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n$. A privacy mechanism is said to be *α-differentially private*, if for all possible observations $\{x_i\}_{i=1}^n$ and $\{x_i'\}_{i=1}^n$ that differ in at most one coordinate, it holds that

$$\sup_A \frac{Q(A|x_1,\dots,x_n)}{Q(A|x_1',\dots,x_n')} \le e^\alpha, \tag{2}$$

where the supremum is taken over all measurable sets A.

For an α-differentially private mechanism to satisfy *the local privacy constraint*, the output must be of the form $\{Z_i\}_{i=1}^n \subset \mathcal{Z}$, where Z_1 is generated solely based on X_1, and for each $i \in \{2,\dots,n\}$, $n \ge 2$, Z_i is generated based on X_i and $\{Z_j\}_{j=1}^{i-1}$. The constraint (2) can therefore be written in terms of the conditional distributions Q_i that generate the Z_i, that is,

$$\max_{i=1,\dots,n} \sup_A \sup_{z_1,\dots,z_{i-1}\in\mathcal{Z}} \sup_{x,x'\in\mathcal{X}} \frac{Q_i(A|x,z_1,\dots,z_{i-1})}{Q_i(A|x',z_1,\dots,z_{i-1})} \le e^\alpha, \tag{3}$$

with the convention that $\{z_1,\dots,z_j\} = \varnothing$ if $j < 1$. Any conditional distribution Q that satisfies (3) is said to be an *α-locally differentially private (LDP) privacy mechanism* (see, e.g., Duchi, Jordan and Wainwright (2018)). We write $\mathcal{Q}_\alpha$ for the set of all α-LDP privacy mechanisms.

From (3), we can see that $\alpha > 0$ represents the desired level of privacy which is an input to the data analysis—the larger α is the less protected the raw data are. In this paper we focus on the high-privacy regime $0 < \alpha \le 1$, where α may be a function of the sample size n. As we will discuss in more detail later, we will require that $n\alpha^2$ diverges, as the sample size n grows unbounded.

With both the ingredients in hand, we define the *α-LDP minimax risk under contamination*, which is at the centre of the analysis in this paper; that is,

$$\mathcal{R}_{n,\alpha}\big(\theta(\mathcal{P}), \Phi\circ\rho, \varepsilon\big) = \inf_{Q\in\mathcal{Q}_\alpha} \inf_{\hat\theta} \sup_{P_\varepsilon\in\mathcal{P}_\varepsilon(\mathcal{P})} \mathbb{E}_{P_\varepsilon,Q}\big[\Phi\circ\rho\big\{\hat\theta, \theta(P)\big\}\big], \tag{4}$$

where:

- the population quantity of interest is $\theta(P) \in \Theta$, denoting a functional supported on $\mathcal{P}$;

- the bivariate function ρ is a semi-metric on the space Θ and $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a nondecreasing function with $\Phi(0) = 0$;
- the first infimum is taken over all possible α-LDP privacy mechanisms;
- the second infimum is over all measurable functions $\hat{\theta} = \hat{\theta}(Z_1, \ldots, Z_n)$ of the privatised data generated from privacy mechanism Q; and
- the loss function is of the form of an unconditional expectation, with respect to both the data generating mechanism P_ε and the privacy mechanism Q.

As detailed in (4), our goal is to understand the fundamental limits imposed by both the contamination and the privacy constraint. In Proposition 1 below, we show that such statistical tasks are at least as hard as either only preserving privacy at level α without contamination or only being robust against contamination without preserving privacy.

PROPOSITION 1. *For any $\varepsilon \in [0, 1)$, define the total variation modulus of continuity $\omega(\varepsilon)$ to be*

$$\omega(\varepsilon) = \sup\big\{\rho\big(\theta(R_0), \theta(R_1)\big) : \mathrm{TV}(R_0, R_1) \le \varepsilon/(1-\varepsilon),\, R_0, R_1 \in \mathcal{P}\big\}. \tag{5}$$

Define the α-LDP minimax risk without contamination to be

$$\mathcal{R}_{n,\alpha}\big(\theta(\mathcal{P}), \Phi \circ \rho\big) = \mathcal{R}_{n,\alpha}\big(\theta(\mathcal{P}), \Phi \circ \rho, 0\big),$$

with $\mathcal{R}_{n,\alpha}(\cdot,\cdot,\cdot)$ defined in (4). For any given $\alpha \in (0, \infty)$, it holds that

$$\mathcal{R}_{n,\alpha}\big(\theta(\mathcal{P}), \Phi \circ \rho, \varepsilon\big) \ge \mathcal{R}_{n,\alpha}\big(\theta(\mathcal{P}), \Phi \circ \rho\big) \vee \frac{\Phi(\omega(\varepsilon)/2)}{2}.$$

We remark that Theorem 5.1 in Chen, Gao and Ren (2018) provides a general lower bound under Huber's contamination model for nonprivate data. Proposition 1 extends it to a general case that accounts for the LDP constraint. A key aspect of the lower bound in Proposition 1 is that the cost of contamination is separated from that of privacy, that is, the level of privacy required does not increase the error introduced by the contamination. In the sequel, we will show that this lower bound is tight in the estimation problems we consider, while a similar decoupling of a slightly different form can be found for our testing problem. In the examples we consider, the difficulty of the combined problem of robustness and LDP is the difficulty of the harder one of the two individual problems. We see this disentanglement as a sign of the connection between privacy and robustness. In our examples, we have shown that it is possible to find optimal procedures that are simultaneously privacy-preserving and robust. We have not found any problems for which privacy precludes robustness, or vice versa.

The intuition behind Proposition 1 is that if the two distributions are indistinguishable on the raw data space $\mathcal{X}$, then no 'transformation' Q can distinguish them either. Note that a similar quantity to $\omega(\varepsilon)$ was used by Donoho and Liu (1991), where the Hellinger distance was considered instead of the total variance distance, to translate perturbations in the distribution to perturbations in the quantities of interest measured by the chosen loss function—$\rho(\theta(R_0), \theta(R_1))$.

1.3. *Notation.* For $a, b \in \mathbb{R}$, let $a \wedge b = \min(a, b)$, $a \vee b = \max(a, b)$ and $a_+ = a \vee 0$. For nonnegative real sequences $\{a_n\}_{n\in\mathbb{N}_+}$ and $\{b_n\}_{n\in\mathbb{N}_+}$, $a_n \ll b_n$ denotes that $\lim_{n\to\infty} a_n/b_n = 0$, $a_n \gg b_n$ denotes that $b_n \ll a_n$, $a_n \lesssim b_n$ denotes the existence of a constant $C > 0$ such that $\limsup_{n\to\infty} a_n/b_n \le C$, $a_n \gtrsim b_n$ denotes that $b_n \lesssim a_n$ and $a_n \asymp b_n$ denotes that $a_n \lesssim b_n \lesssim a_n$. Let $\mathbb{N}_+$ denote all positive integers and $\mathbb{R}_+$ denote the set of nonnegative real numbers. Let $|S|$ denote the cardinality of a set S. For two distributions P_a and P_b, their total variation distance is $\mathrm{TV}(P_a, P_b) = \sup_S |P_a(S) - P_b(S)|$, where the supremum is taken over all measurable

sets S, and their Kullback–Leibler divergence is $\mathrm{KL}(P_a, P_b) = \int \log(\mathrm{d}P_a/\mathrm{d}P_b)\,\mathrm{d}P_a$, if P_a is absolutely continuous with respect to P_b. Let $L_2[0,1] = \{f : [0,1] \to \mathbb{R} : \int_0^1 f^2(x)\,\mathrm{d}x < \infty\}$. A random variable X is sub-exponential with parameters (τ, b) if $\mathbb{E}[\exp\{\lambda(X - \mathbb{E}(X))\}] \le \exp(\lambda^2\tau^2/2)$, for $|\lambda| \le 1/b$.

2. Robust testing under local differential privacy. Under the general setup described in Section 1.2, assuming that $X_1, \ldots, X_n$ are i.i.d. random variables generated from P on $\mathcal{X}$, we consider the robust testing problem

$$\begin{aligned} &\mathrm{H}_0 : P \in \mathcal{P}_\varepsilon(P_0) = \{P_\varepsilon : (1-\varepsilon)P_0 + \varepsilon G, G \in \mathcal{G}\} \quad \text{vs.} \\ &\mathrm{H}_1 : P \in \mathcal{P}_\varepsilon(P_1) = \{P_\varepsilon : (1-\varepsilon)P_1 + \varepsilon G, G \in \mathcal{G}\}, \end{aligned} \tag{6}$$

where P_0 and P_1 are two fixed distributions supported on $\mathcal{X}$, and $\mathcal{G}$ is the set of all distributions supported on $\mathcal{X}$. In this section, we are interested in testing (6) under an α-LDP constraint. For a given α-LDP privacy mechanism Q, we let $\Phi_Q = \{\phi : \mathcal{Z}^n \to \{0,1\}\}$ denote the set of all $\{0,1\}$-valued measurable functions of privatised data $\{Z_i\}_{i=1}^n$ generated via the privacy mechanism Q. The α-LDP minimax testing risk can be written as

$$\mathcal{R}_{n,\alpha}(\varepsilon) = \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\phi \in \Phi_Q} \Big\{ \sup_{P \in \mathcal{P}_\varepsilon(P_0)} \mathbb{E}_{P,Q}(\phi) + \sup_{P' \in \mathcal{P}_\varepsilon(P_1)} \mathbb{E}_{P',Q}(1-\phi) \Big\}, \tag{7}$$

which corresponds to (4) with $\mathcal{P} = \{P_0, P_1\}$ and ρ being the 0-1 loss, that is, $\rho = 0$ if ϕ returns the correct hypothesis and $\rho = 1$ otherwise. Note that, by considering the trivial test that always rejects H_0, we have $\mathcal{R}_{n,\alpha}(\varepsilon) \le 1$.

To fully understand the hardness of (6), we construct lower and upper bounds on $\mathcal{R}_{n,\alpha}(\varepsilon)$ in Sections 2.1 and 2.2, respectively. We conclude in Section 2.3 with a discussion of the existing literature. In Section B.2 of the Supplementary Material (Li, Berrett and Yu (2023)), we show that very similar results hold when the LDP constraint is relaxed to a general class of local privacy constraints, namely Rényi local differential privacy.

2.1. *Lower bound.* In Proposition 2 below, we provide a lower bound on the α-LDP minimax testing risk $\mathcal{R}_{n,\alpha}(\varepsilon)$, in terms of the sample size n, the privacy constraint α, the total variation distance $\mathrm{TV}(P_0, P_1)$ and the Huber contamination proportion ε. As we will discuss later, this result also serves as an infeasibility result by providing necessary conditions on the testing problem (6) for the existence of a test with vanishing risk.

PROPOSITION 2. *For $\alpha \in (0,1)$ and the robust testing problem defined in* (6)*, it holds that the α-LDP minimax testing risk defined in* (7) *satisfies*

$$\mathcal{R}_{n,\alpha}(\varepsilon) \ge \frac{1}{2}\exp\{-16\alpha^2 n\{\mathrm{TV}(P_0, P_1) - \varepsilon/(1-\varepsilon)\}_+^2\}.$$

Note that when $\mathrm{TV}(P_0, P_1) \le \varepsilon/(1-\varepsilon)$, we can actually show that $\mathcal{R}_{n,\alpha}(\varepsilon) = 1$. Indeed, this can be seen by a simpler argument since, in this case, according to Lemma A.1 in the Supplementary Material, there exists some $\widetilde{P} \in \mathcal{P}_\varepsilon(P_0) \cap \mathcal{P}_\varepsilon(P_1)$, and therefore

$$\mathcal{R}_{n,\alpha}(\varepsilon) \ge \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\phi \in \Phi_Q} \{\mathbb{E}_{\tilde{P},Q}[\phi] + \mathbb{E}_{\tilde{P},Q}[1-\phi]\} = 1.$$

In particular, whenever $\varepsilon \ge 1/2$ we have that $\varepsilon/(1-\varepsilon) \ge 1 \ge \mathrm{TV}(P_0, P_1)$, so that $\mathcal{R}_{n,\alpha}(\varepsilon) = 1$.

2.2. *Upper bound.* Given the lower bound in Proposition 2, we are to show that a folklore test based on the randomised response mechanism and the Scheffé set (e.g., Devroye and Lugosi (2001)) is minimax rate-optimal for the testing problem (6).

Step 1 (Privatisation). Given data $\{X_i\}_{i=1}^n$, let $Y_i = \mathbb{1}\{X_i \in A^c\}$, where A is the Scheffé set of P_0 and P_1 in (6), that is, $A = \arg\max_{S\subset\mathcal{X}}\{P_0(S) - P_1(S)\}$. Let the privatised data $\{Z_i\}_{i=1}^n$ be obtained via the randomised response mechanism. To be specific, let $\{U_i\}_{i=1}^n$ be independent Unif[0, 1] random variables that are independent of $\{X_i\}_{i=1}^n$. For $i \in \{1, \dots, n\}$, let

$$Z_i = \begin{cases} Y_i, & U_i \le e^\alpha/(1+e^\alpha), \\ 1 - Y_i, & \text{otherwise.} \end{cases} \tag{8}$$

Step 2 (Test construction). Let

$$\widehat{N}_0 = \sum_{i=1}^n \mathbb{1}\{Z_i = 0\} \quad \text{and} \quad \widetilde{N}_0 = \frac{e^\alpha + 1}{e^\alpha - 1}\left(\widehat{N}_0 - \frac{n}{e^\alpha + 1}\right).$$

The test is then defined as

$$\tilde{\phi} = \mathbb{1}\big\{\big|\widetilde{N}_0/n - P_0(A)\big| > \big|\widetilde{N}_0/n - P_1(A)\big|\big\} = \mathbb{1}\big\{2\widetilde{N}_0/n < P_0(A) + P_1(A)\big\}. \tag{9}$$

Note that the privacy mechanism defined in Step 1 satisfies the α-LDP constraint in (3) (e.g., Gopi et al. (2020)). In fact, the test $\tilde{\phi}$ and a nonprivate counterpart have been used in the nonrobust two-point testing problem with LDP constraints (e.g., Algorithm 6 in Joseph et al. (2019); Algorithm 4 in Gopi et al. (2020)) and the robust two-point testing problem without LDP constraints (e.g., Section 2 in Chen, Gao and Ren (2016)), respectively. It is known to be rate-optimal in both cases. The following theorem, combining previous analyses, shows that this test $\tilde{\phi}$ is still optimal under both the privacy constraint and the presence of contamination.

THEOREM 3. *For $\alpha \in (0, 1)$ and the robust testing problem defined in (6), assuming that $\mathrm{TV}(P_0, P_1) > 2\varepsilon$ with $\varepsilon \in [0, 1/2)$, the test defined in (8) and (9) satisfies that*

$$\sup_{P\in\mathcal{P}_\varepsilon(P_0)} \mathbb{E}_{P,Q}(\tilde{\phi}) + \sup_{P'\in\mathcal{P}_\varepsilon(P_1)} \mathbb{E}_{P',Q}(1-\tilde{\phi}) \le 2\exp\big[-C\alpha^2 n\big\{\mathrm{TV}(P_0, P_1) - 2\varepsilon\big\}^2\big],$$

where $C > 0$ is some absolute constant.

We first note that in Theorem 3, we require $\varepsilon < 1/2$, which, as discussed above, is necessary for the existence of nontrivial tests. Further, since $\varepsilon/(1-\varepsilon) \ge \varepsilon$, the lower bound in Proposition 2 implies that

$$\mathcal{R}_{n,\alpha}(\varepsilon) \ge \frac{1}{2}\exp\big\{-16\alpha^2 n\big\{\mathrm{TV}(P_0, P_1) - \varepsilon\big\}_+^2\big\}. \tag{10}$$

Comparing the upper bound in Theorem 3 and the lower bound in (10), up to constants, we see that the test $\tilde{\phi}$ is optimal in terms of the privacy constraint α, the sample size n, the separation $\mathrm{TV}(P_0, P_1)$ and the contamination proportion ε.

REMARK 1 (When ε is known). When ε is known, a modification of the test procedure defined in (8) and (9) achieves slightly better performance and shows that uniformly consistent testing is possible if and only if $\mathrm{TV}(P_0, P_1) > \varepsilon/(1-\varepsilon)$. With the same privacy mechanism as in (8), consider

$$\phi' = \mathbb{1}\big\{2\widetilde{N}_0/n < (1-\varepsilon)\big\{P_0(A) + P_1(A)\big\} + \varepsilon\big\},$$

which uses the same test statistic but compares it to a different critical value. Similar calculations to those carried out for Theorem 3 show that

$$\sup_{P\in\mathcal{P}_\varepsilon(P_0)} \mathbb{E}_{P,Q}(\phi') + \sup_{P'\in\mathcal{P}_\varepsilon(P_1)} \mathbb{E}_{P',Q}(1-\phi') \le 2\exp\bigl[-C'\alpha^2 n\bigl\{\mathrm{TV}(P_0, P_1) - \varepsilon/(1-\varepsilon)\bigr\}_+^2\bigr]$$

for some absolute constant $C' > 0$, provided $\alpha \in (0, 1]$. We now see that, when $\mathrm{TV}(P_0, P_1) \le \varepsilon/(1-\varepsilon)$ we have $\mathcal{R}_{n,\alpha}(\varepsilon) = 1$, and when $\mathrm{TV}(P_0, P_1) > \varepsilon/(1-\varepsilon) + 1/\sqrt{n\alpha^2}$ we have

$$-\log \mathcal{R}_{n,\alpha}(\varepsilon) \asymp \alpha^2 n\bigl\{\mathrm{TV}(P_0, P_1) - \varepsilon/(1-\varepsilon)\bigr\}_+^2.$$

Details of the calculations are given at the end of Section B.1 of the Supplementary Material (Li, Berrett and Yu (2023)).

REMARK 2 (Relation to Proposition 1). The derived minimax error rate does not match the form given in Proposition 1. This form appears to be most suitable for estimation problems, while for testing problems we must look at the error slightly differently. Indeed, it is common in the theory of hypothesis testing to look at conditions under which errors are below given thresholds, often through minimal separation rates, rather than errors themselves (e.g., Ingster and Suslina (2003)). In problem (6) the total variation modulus of continuity is given by

$$\begin{aligned}\omega(\varepsilon) &= \sup\bigl\{\mathbb{1}_{\{\theta(R_1)\neq\theta(R_0)\}} : \mathrm{TV}(R_0, R_1) \le \varepsilon/(1-\varepsilon), R_0, R_1 \in \{P_0, P_1\}\bigr\} \\ &= \mathbb{1}_{\{\mathrm{TV}(P_0,P_1)\le\varepsilon/(1-\varepsilon)\}},\end{aligned}$$

so we have $\omega(\varepsilon) \le 0.1$ if and only if $\mathrm{TV}(P_0, P_1) > \varepsilon/(1-\varepsilon)$. Moreover, we have seen that, up to constants, $\mathcal{R}_{n,\alpha}(0) \le 0.1$ if and only if $\mathrm{TV}(P_0, P_1)$ is larger than $(n\alpha^2)^{-1/2}$ and, on the other hand, $\mathcal{R}_{n,\alpha}(\varepsilon) \le 0.1$ if and only if $\mathrm{TV}(P_0, P_1) \ge \varepsilon/(1-\varepsilon) + (n\alpha^2)^{-1/2}$. Thus $\mathrm{TV}(P_0, P_1)$ is seen to characterise when each of $\omega(\varepsilon)$, $\mathcal{R}_{n,\alpha}(0)$, $\mathcal{R}_{n,\alpha}(\varepsilon)$ is below the (arbitrary) threshold 0.1. The level that $\mathrm{TV}(P_0, P_1)$ must exceed for $\mathcal{R}_{n,\alpha}(\varepsilon) \le 0.1$ is given by the *sum* of the levels required for $\omega(\varepsilon) \le 0.1$ and $\mathcal{R}_{n,\alpha}(0) \le 0.1$, and we see a decoupling of the contamination and the privacy.

2.3. *Discussion*. Our results in Proposition 2 and Theorem 3 are similar in spirit to Theorem 5.7 in Joseph et al. (2019), which states a minimax lower bound result in terms of the sample complexity and presents an algorithm using a Laplace mechanism that achieves the optimal sample complexity. In particular, they consider a compound hypothesis testing problem where H_0 and H_1 correspond to convex and compact sets of discrete distributions well separated in terms of total variation distance, whereas our result concerns Huber's contamination model, with different proof schemes.

An extension of the two-point testing problem defined in (6) is hypothesis selection, which is popular in both the computer science and statistics literatures (e.g., Gopi et al. (2020), Bun et al. (2021), Yatracos (1985), Devroye and Lugosi (2001), Chen, Gao and Ren (2016)). To be specific, for a fixed but unknown distribution $P \in \mathcal{P}$, given a set of $k_0 \in \mathbb{N}_+$ distributions $\mathcal{Q} = \{q_1, \ldots, q_{k_0}\} \subset \mathcal{G}$, one seeks an element in $\mathcal{Q}$ that is closest to P in total variation distance. In particular, taking $\mathcal{Q}$ to be a δ-covering set ($\delta > 0$) of $\mathcal{P}$, with k_0 being the δ-covering number, this hypothesis selection problem is similar to an estimation problem of the distribution P. Based on this setup, Chen, Gao and Ren (2016) shows that applying a tournament procedure with nonprivate counterparts of $\tilde{\phi}$, to a δ-covering set of $\mathcal{P}$, is minimax rate-optimal for estimating $P \in \mathcal{P}_\varepsilon(\mathcal{P})$ in terms of the total variation metric. Their procedure returns an element $\widehat{P} \in \mathcal{Q}$, with high probability, satisfying that

$$\mathrm{TV}(\widehat{P}, P) \lesssim \bigl\{\sqrt{\log(k_0)/n + \delta^2}\bigr\} \vee \varepsilon.$$

With $\delta = \inf\{\delta_1 : n \geq \log(k_0)/\delta_1^2\}$, it holds that $\mathrm{TV}(\widehat{P}, P) \lesssim \delta \vee \varepsilon$, and $n \geq \log(k_0)/\delta^2$ is called the sample complexity of the procedure. Based on the same setup, Bun et al. (2021) considers the problem under the central privacy model but without contamination, that is, $P \in \mathcal{P}$, and they develop an algorithm that guarantees $\mathrm{TV}(\widehat{P}, P) \lesssim \delta$ with the sample complexity $n \gtrsim \log(k_0)/\delta^2 + \log(k_0)/(\delta\alpha)$. Note that in both these two problems, k_0 appears in the sample complexity through its logarithm $\log(k_0)$.

However, under local privacy constraints, Gopi et al. (2020) establishes a lower bound (cf. Theorem 2 in Gopi et al. (2020)) on the sample complexity of $n \gtrsim k_0/(\delta^2\alpha^2)$, which shows that the cost of this general estimation method induced by a δ-covering set is exponentially higher in the local privacy setting compared to the central privacy and nonprivate settings. The exponential gap in the sample complexity directly leads to a suboptimal rate for estimation tasks. We now state an example illustrating the suboptimality of hypothesis selection approaches to estimation problems, even in simple parametric problems, which motivates our proposals of *problem-specific* estimation procedures studied in the rest of this paper.

EXAMPLE. Consider $\{X_i\}_{i=1}^n$ to be i.i.d. random variables from $\mathcal{N}(\mu, 1)$, $\mu \in [-D, D]$. Estimating μ is a special case of the robust mean estimation problem in Section 3. For this problem, we have $k_0 \asymp D/\delta$. Theorem 2 in Gopi et al. (2020) implies that the sample complexity of the associated hypothesis selection problem is

$$n \gtrsim \frac{k_0}{\delta^2\alpha^2} \asymp \frac{D/\delta}{\delta^2\alpha^2} = \frac{D}{\delta^3\alpha^2} \quad \text{i.e., } \delta \gtrsim \left(\frac{D}{n\alpha^2}\right)^{1/3}.$$

Since $\mathrm{TV}(\mathcal{N}(\mu, 1), \mathcal{N}(\mu', 1)) \asymp |\mu - \mu'|$ (Devroye, Mehrabian and Reddad (2018)), this implies that any estimator $\hat{\mu}$ obtained based on the δ-covering set would have convergence rate measured by $|\hat{\mu} - \mu|$ bounded below by $\{D/(n\alpha^2)\}^{1/3}$. However, since Gaussian distributions have moments of all orders, we will see that we can attain a near-parametric upper bound by applying Theorem 5 below with arbitrarily large k.

As for the questions Q1 and Q2 raised in Section 1.1, in this two-point testing problem, we see that:

- there exists a procedure optimal against contamination (Section 2 in Chen, Gao and Ren (2016)) that can be properly privatised to achieve optimal performance; and
- there exists an α-LDP procedure (Joseph et al. (2019)) that is automatically robust and minimax rate-optimal.

3. Robust mean estimation under local differential privacy. Recalling the general setup in Section 1.2, in this section we consider distributions supported on the real line, that is, $\mathcal{X} = \mathbb{R}$, with finite kth ($k > 1$) central moments and possibly diverging expectation, as the sample size grows unbounded. We are interested in estimating the expectation, that is, $\theta(P) = \mathbb{E}_{X\sim P}(X)$. To be specific, we let

$$\mathcal{P} = \mathcal{P}_k = \big\{P : \mu = \mathbb{E}_{X\sim P}(X) \in [-D, D], \sigma^k = \mathbb{E}_{X\sim P}\big[|X - \mu|^k\big] \leq 1\big\}, \tag{11}$$

where k is considered fixed but arbitrary and $D \geq 1$ may be a function of the sample size.

We again assume the data are generated from Huber's contamination model (1). Let $\{X_i\}_{i=1}^n$ be i.i.d. random variables with distribution $P_{k,\varepsilon} \in \mathcal{P}_\varepsilon(\mathcal{P}_k)$ and suppose that we are interested in estimating the expectation of the inlier distribution μ.

The α-LDP minimax risk defined in (4) takes its specific form in the robust mean estimation problem as follows:

$$\mathcal{R}_{n,\alpha}(\varepsilon) = \mathcal{R}_{n,\alpha}\big(\theta(\mathcal{P}_k), (\cdot)^2, \varepsilon\big) = \inf_{Q\in\mathcal{Q}_\alpha} \inf_{\hat{\mu}} \sup_{P_{k,\varepsilon}\in\mathcal{P}_\varepsilon(\mathcal{P}_k)} \mathbb{E}_{P_{k,\varepsilon},Q}\big\{(\hat{\mu} - \mu)^2\big\}, \tag{12}$$

where the metric of interest is the squared loss and the infimum over $\hat{\mu}$ is taken over all measurable functions of the privatised $\{X_i\}_{i=1}^n$ via some privacy mechnism $Q \in \mathcal{Q}_\alpha$.

3.1. *Lower bound.* We first provide in Proposition 4 a lower bound that comprises three terms. If $\log(D) \geq 32n\alpha^2$, then $\mathcal{R}_{n,\alpha}(\varepsilon) \gtrsim 1$. This implies that uniformly consistent estimation of μ is impossible. When $\log(D) \ll n\alpha^2$, the lower bound simplifies to $(n\alpha^2)^{1/k-1} \vee \varepsilon^{2-2/k}$, which is the minimax rate of the robust mean estimation problem with fixed D. We show in Theorem 5 that a matching upper bound can be achieved by a novel noninteractive procedure under minimal conditions.

PROPOSITION 4. *Let $\{X_i\}_{i=1}^n$ be i.i.d. random variables from $P_{k,\varepsilon} \in \mathcal{P}_\varepsilon(\mathcal{P}_k)$, with $\mathcal{P}_k$ defined in (11). For $\alpha \in (0, 1]$ and $n \geq n_0$ with a large enough absolute constant $n_0 \in \mathbb{N}_+$, it holds that the α-LDP minimax estimation risk defined in (12) satisfies*

$$\mathcal{R}_{n,\alpha}(\varepsilon) \gtrsim \left(n\alpha^2\right)^{1/k-1} \vee \varepsilon^{2-2/k} \vee \frac{D^2}{\exp\left(64n\alpha^2\right)}.$$

Proposition 4 explains the hardness of this estimation problem by isolating the cost of preserving privacy, contamination and estimating a possibly diverging mean respectively. If any of these three becomes too hard, this estimation problem becomes infeasible. In terms of the preservable level of privacy, we require $\alpha \gg n^{-1/2}$ for consistency. In terms of the Huber contamination proportion, we require $\varepsilon \ll 1$. In terms of D, the upper bound on the absolute mean, we require $D \ll \exp(32n\alpha^2)$.

The dependence on D is a rather interesting finding. When $k = 2$ and $D = \infty$, it is shown in Appendix G of Duchi, Jordan and Wainwright (2013) that $\mathcal{R}_{n,\alpha}(0) = \infty$. To interpolate between the case of an unbounded parameter space and the case where the mean takes values in a fixed compact set, we prove in Lemma C.1 (Li, Berrett and Yu (2023)) that, for any $D \in [0, \infty)$ and $n \in \mathbb{N}_+$, the lower bound

$$\mathcal{R}_{n,\alpha}(0) \geq \frac{D^2}{32\exp\left(64n\alpha^2\right)}$$

holds.

Another interesting aspect roots in the derivation of the term $(n\alpha^2)^{1/k-1}$, which unveils a somewhat deeper connection between locally private estimation and estimation within Huber's contamination model. To derive the term $(n\alpha^2)^{1/k-1}$, we apply Corollary 3.1 in Rohde and Steinberger (2020), which crucially depends on the following quantity:

$$\omega'(\eta) = \omega\big(\eta/(1+\eta)\big) = \sup\big\{|\mu_0 - \mu_1| : \mathrm{TV}(R_0, R_1) \leq \eta, R_0, R_1 \in \mathcal{P}_k\big\},$$

where μ_0 and μ_1 denote the means of R_0 and R_1, respectively. It is almost identical to $\omega(\cdot)$ defined in (5), up to a change of variable, and has been shown to play a central role in establishing minimax rates for locally private estimation problems (Rohde and Steinberger (2020)). Therefore, we see that the total variation modulus of continuity $\omega(\cdot)$, a single unifying quantity, can quantify the costs of both privacy and contamination in Huber's contamination model.

3.2. *Upper bound.* We are now to show that the lower bound in Proposition 4 is indeed tight by providing a novel noninteractive estimator that is adaptive to D. We split the data into four folds, each of which is privatised separately. The procedure has two main steps:

- First, for a pre-specified $M > 0$, the first fold is used to construct a private histogram with bin width $M/3$ and to identify bins containing a proportion of the contaminated distribution exceeding a threshold. This allows us to find a larger bin of width M of the form $[\{S + (L-1)/3\}M, \{S + 1 + (L-1)/3\}M]$, with $S \in \mathbb{N}_+$ and $L \in \{0, 1, 2\}$, such that only a negligible proportion of the distribution lies outside this interval. This interval can be thought of as a crude, initial estimate of the location of the distribution.

- Second, for each $\ell \in \{0, 1, 2\}$ and each data point X_i in fold $\ell + 2$, we divide $X_i - (\ell - 1)M/3$ by M and use a Laplace mechanism to privatise the remainder. The privatised data from fold $L + 2$ can then be used to pinpoint the mean within the interval $[\{S + (L - 1)/3\}M, \{S + 1 + (L - 1)/3\}M]$, by adding the privatised remainders to its left-hand end point.

Crucially, the value of L is not needed to privatise the data from the final folds and is only used when constructing the final estimator. This can be done because there are only three possible values of L and we may reserve a fold for each eventuality. One fold is incorporated into the estimator and the other two are discarded.

We now formalise the procedure described above.

Step 1 (Privatisation). For some absolute constant $c > 0$, let $T = \exp(cn\alpha^2)$. Let M be a tuning parameter such that $T/M \in \mathbb{N}$. Let

$$A_j = \big[(j-1)M/3, jM/3\big), \quad j \in \mathcal{J} = \{-3T/M, -3T/M + 1, \ldots, 3T/M, 3T/M + 1\}.$$

Let the data be $\{X_i\}_{i=1}^{4n}$. Generate independent standard Laplace variables $(W_{ij})_{i\in[n], j\in\mathcal{J}}$, $(W_i^{(\ell)})_{i=(\ell+1)n+1}^{(\ell+2)n}$ for $\ell = 0, 1, 2$. For $i \in [n]$ and $j \in \mathcal{J}$ set

$$Z_{ij} = \mathbb{1}_{\{X_i \in A_j\}} + \frac{2}{\alpha} W_{ij}. \tag{13}$$

For $\ell = 0, 1, 2$ and $i = (\ell + 1)n + 1, \ldots, (\ell + 2)n$ set

$$R_i^{(\ell)} = \min\big\{X_i - (j-1)M/3 : j \in \mathcal{J}, j \equiv \ell \,(\mathrm{mod}\, 3), X_i \geq (j-1)M/3\big\}$$

and

$$Z_i^{(\ell)} = \big[R_i^{(\ell)}\big]_0^M + \frac{M}{\alpha} W_i^{(\ell)}, \tag{14}$$

where $[\cdot]_0^M = \min\{\max\{\cdot, 0\}, M\}$.

Step 2 (Estimator construction). With

$$\delta = T^{-2}\big(n\alpha^2\big)^{-1} \quad \text{and} \quad \tau = \epsilon + (1-\epsilon)(6/M)^k + 4\sqrt{2\log\big(12T/(M\delta)\big)/\big(n\alpha^2\big)},$$

define

$$\widehat{\mathcal{J}} = \left\{ j \in \mathcal{J} : \frac{1}{n}\sum_{i=1}^{n} Z_{ij} \geq \tau \right\}.$$

If $\widehat{\mathcal{J}} = \varnothing$, then output $\hat{\mu} = 0$. Otherwise, let

$$J = \max \widehat{\mathcal{J}} - 1 \quad \text{and} \quad L = \begin{cases} 0 & \text{if } J \equiv 0 \,(\mathrm{mod}\, 3), \\ 1 & \text{if } J \equiv 1 \,(\mathrm{mod}\, 3), \\ 2 & \text{if } J \equiv 2 \,(\mathrm{mod}\, 3). \end{cases}$$

The estimator is defined as

$$\hat{\mu} = \frac{1}{n} \sum_{i=(L+1)n+1}^{(L+2)n} Z_i^{(L)} + (J-1)M/3. \tag{15}$$

The privatisation in (13) and (14) is noninteractive and satisfies α-LDP, as shown in Lemma C.2 (Li, Berrett and Yu (2023)). The statistical guarantees of the estimator $\hat{\mu}$ defined in (15) are collected in Theorem 5 below.

THEOREM 5. *Suppose we are given i.i.d. random variables* $\{X_i\}_{i=1}^{4n}$ *with distribution* $P_{k,\varepsilon} = (1-\varepsilon)P_k + \varepsilon G$, *where* $P_k \in \mathcal{P}_k$ *is defined in* (11) *and* G *is any arbitrary distribution supported on* $\mathbb{R}$. *The estimator* $\hat{\mu}$ *defined in* (15), *with inputs satisfying that* (*i*) $T \geq D$, (*ii*) $(n\alpha^2)^{-1}\log(12T^3n\alpha^2/M) \leq \min(\alpha^{-2}/2, 1/512)$ *and* (*iii*) $\varepsilon + (1-\varepsilon)(6/M)^k \leq 1/12$, *satisfies that*

$$\mathbb{E}\{(\hat{\mu}-\mu)^2\} \lesssim T^2\exp(-n\alpha^2/512) + \frac{M^2}{n\alpha^2} + \varepsilon^2M^2 + M^{-2(k-1)}.$$

In particular, choosing $T = \exp(n\alpha^2/3072)$ *and* $M \asymp \varepsilon^{-1/k} \wedge (n\alpha^2)^{1/(2k)}$, *when* $\alpha \leq 1$, $D \leq T$ *and* $\min\{\varepsilon, (n\alpha^2)^{-1}\} \leq c_0$ *for some small constant* $c_0 > 0$, *we have that*

$$\mathbb{E}\{(\hat{\mu}-\mu)^2\} \lesssim (n\alpha^2)^{1/k-1} \vee \varepsilon^{2-2/k}. \tag{16}$$

Note that the constant hidden in (16) depends on k but is independent of all other parameters. With a properly chosen truncation tuning parameter M, Theorem 5 gives an upper bound on the mean squared error of $\hat{\mu}$ of the order $(n\alpha^2)^{1/k-1} \vee \varepsilon^{2-2/k}$. Comparing with Proposition 4, we see that $\hat{\mu}$ is minimax rate-optimal in terms of the dependence on n, α and ε. It is notable that it remains rate-optimal even for a diverging D, provided that it is not growing faster than a certain exponential rate in $n\alpha^2$ (more precisely, we require $D \leq T = \exp(n\alpha^2/3072)$). In fact, this requirement is essentially unavoidable, as Proposition 4 shows that uniformly consistent estimation is impossible if $\log D \gg n\alpha^2$. Lastly, Theorem 5 shows that the error of our estimator has an upper bound independent of D, provided $D \leq T$, which is in contrast to the performance of the private mean estimator based on the standard Laplace mechanism studied in Duchi, Jordan and Wainwright (2018) (see Proposition C.1 in the Supplementary Material (Li, Berrett and Yu (2023))).

3.3. *Discussion*. Our focus is on cases where both contamination and privacy constraints are present. In this section, we discuss some related results in the literature that only consider either contamination or the local privacy constraint.

Without the local privacy constraint, Prasad, Balakrishnan and Ravikumar (2019) proposes a two-step robust mean estimator (cf. Algorithm 2 therein) that achieves optimality in the model (11) with $k \geq 2$ and $D = \infty$, under mild conditions (cf. Lemma 3 therein). With probability at least $1-\delta$, their estimator $\hat{\mu}_{\mathrm{Pra}}$ satisfies that

$$|\hat{\mu}_{\mathrm{Pra}} - \mu|^2 \lesssim \frac{\log(1/\delta)}{n} \vee \varepsilon^{2-2/k}, \tag{17}$$

for $k \geq 4$, which is known to be information-theoretically optimal (e.g., Diakonikolas et al. (2019)). For $k \geq 2$, (17) also holds as long as $\log(1/\delta) \gtrsim \log(n)$ (cf. Lemma 3 therein). Recalling that our results are in the form of expectations, to compare (17) with Theorem 5, we hence ignore the logarithmic term. Regarding the term involving ε, we see that in Theorem 5, it is completely isolated from the privacy level α; and in both (17) and Theorem 5, it is of the order $\varepsilon^{2-2/k}$. As for preserving privacy at level α, we see the effects are two-fold (see also Section 3.2.1 in Duchi, Jordan and Wainwright (2013)): (1) a reduction of the effective sample size from n to $n\alpha^2$; and (2) instead of n^{-1} in the nonprivate case, we have in Theorem 5 the term $n^{1/k-1}$, that is, the heavy-tailedness of $P \in \mathcal{P}_k$ has no effect on the nonprivate convergence rate in (17), whereas a loss of $1/k$ in the exponent is incurred in the private setting. Given that we have shown in Proposition 4 that the rate we achieved in Theorem 5 is optimal, the loss in the rate $n^{1/k}$ is unavoidable due to the privacy constraint—similar phenomena have also been observed in the nonparametric estimation literature (e.g., Berrett and Butucea (2019), Berrett and Yu (2021)) as well as in Section 4.

In view of Q1 in Section 1.1, we claim that the estimator $\hat{\mu}_{\mathrm{Pra}}$ is similar to our estimator $\hat{\mu}$ in a broad sense. As detailed in Algorithm 2 in Prasad, Balakrishnan and Ravikumar (2019), $\hat{\mu}_{\mathrm{Pra}}$ first constructs a shortest interval initial estimator using half of the data, which is similar to finding J using the first n samples in our construction serving as a crude estimate. The remaining data are then used to refine this crude estimate. One may, therefore, see that our estimator $\hat{\mu}$ as a noninteractive and private version of an optimal robust mean estimator.

In view of Q2 in Section 1.1, Duchi, Jordan and Wainwright (2018) studies the problem (11) with $D = 1$ and $\varepsilon = 0$ (cf. Section 3.2.1 therein). They estimate the mean by averaging the privatised data obtained by adding Laplace noise to the truncated data. This mechanism guarantees the α-LDP constraint, but in the case when D is large and $\varepsilon > 0$, it is sub-optimal.

To summarise, in this robust mean estimation problem, we see that:

- there exists a procedure (Prasad, Balakrishnan and Ravikumar (2019)) optimal against contamination that can be properly privatised to achieve optimal performance; and
- a standard α-LDP procedure is not automatically robust and minimax rate-optimal when the parameter space grows.

4. Robust density estimation under local differential privacy. Recalling the general setup in Section 1.2, in this section, we consider distributions supported on $\mathcal{X} = [0, 1]$, belonging to the Sobolev class $\mathcal{P} = \mathcal{F}_\beta$, defined below.

DEFINITION 6 (Sobolev class). Let $\beta > 1/2$ and $\{\gamma_t\}_{t\in\mathcal{T}}$ be an orthonormal basis of $L_2[0, 1]$ indexed by a countable family $\mathcal{T}$. For a given coefficient sequence $\{a_t\}_{t\in\mathcal{T}}$ associated with $\{\gamma_t\}_{t\in\mathcal{T}}$, the Sobolev class $\mathcal{F}_\beta$ is defined as

$$\begin{aligned}\mathcal{F}_\beta = \Bigg\{ f : [0,1] \to \mathbb{R}_+ \,\Big|\, & \int_{[0,1]} f(x)\,\mathrm{d}x = 1, \\ & \sum_{t\in\mathcal{T}} |a_t|^{2\beta} \left| \int_{[0,1]} f(x)\gamma_t(x)\,\mathrm{d}x \right|^2 = \sum_{t\in\mathcal{T}} |a_t|^{2\beta} |f_t|^2 < \infty \Bigg\}.\end{aligned} \tag{18}$$

We again assume that the data are generated from Huber's contamination model (1). When $\{X_i\}_{i=1}^n$ are i.i.d. random variables with distribution $P_{f_\varepsilon} \in \mathcal{P}_\varepsilon(\mathcal{F}_\beta)$, we are interested in estimating the density of the inlier distribution $\theta(P) = f$.

The α-LDP minimax risk defined in (4) takes its specific form in the robust density estimation problem as follows:

$$\mathcal{R}_{n,\alpha}(\mathcal{F}_\beta, \Phi\circ\rho, \varepsilon) = \inf_{Q\in\mathcal{Q}_\alpha} \inf_{\tilde{f}} \sup_{P_{f_\varepsilon}\in\mathcal{P}_\varepsilon(\mathcal{F}_\beta)} \mathbb{E}_{P_{f_\varepsilon},Q}\big\{\Phi\circ\rho(\tilde{f}, f)\big\}, \tag{19}$$

where the infimum over $\tilde{f}$ is taken over all measurable functions of the privatised data generated from some $Q \in \mathcal{Q}_\alpha$. We consider two loss functions, both of which are commonly used in the nonparametric estimation literature (e.g., Tsybakov (2009)). To be specific, we consider $\Phi\circ\rho$ to be the squared-L_2 loss and the L_∞ loss, separately. For any two density functions g_1 and g_2 supported on $[0, 1]$, let the squared-L_2-loss and the L_∞ loss be

$$\|g_1 - g_2\|_2^2 = \int_{[0,1]} \big\{g_1(x) - g_2(x)\big\}^2\,\mathrm{d}x \quad \text{and} \quad \|g_1 - g_2\|_\infty = \sup_{x\in[0,1]} \big|g_1(x) - g_2(x)\big|, \tag{20}$$

respectively.

4.1. *Lower bound.* In Proposition 7 below, we present lower bounds on $\mathcal{R}_{n,\alpha}(\mathcal{F}_\beta, \Phi \circ \rho, \varepsilon)$, with $\Phi \circ \rho$ taken to be the squared-L_2 loss and L_∞ loss. As we shall discuss later, Proposition 7 is an application of the general lower bound result Proposition 1.

PROPOSITION 7. *Let $\{X_i\}_{i=1}^n$ be i.i.d. random variables with distribution $P_{f_\varepsilon} \in \mathcal{P}_\varepsilon(\mathcal{F}_\beta)$. For $\alpha \in (0,1)$, it holds that the α-LDP minimax estimation risks defined in (19), equipped with the squared-L_2 loss and the L_∞ loss defined in (20), are*

$$\mathcal{R}_{n,\alpha}\big(\mathcal{F}_\beta, \|\cdot\|_2^2, \varepsilon\big) \gtrsim \big(n\alpha^2\big)^{-\frac{2\beta}{2\beta+2}} \vee \varepsilon^{\frac{4\beta}{2\beta+1}} \tag{21}$$

and

$$\mathcal{R}_{n,\alpha}\big(\mathcal{F}_\beta, \|\cdot\|_\infty, \varepsilon\big) \gtrsim \left\{\frac{\log(n\alpha^2)}{n\alpha^2}\right\}^{\frac{2\beta-1}{4\beta+2}} \vee \varepsilon^{\frac{2\beta-1}{2\beta+1}}, \tag{22}$$

respectively.

In view of Proposition 1, the results in Proposition 7 are obtained by separately controlling the lower bounds on (i) the nonrobust cases $\mathcal{R}_{n,\alpha}(\mathcal{F}_\beta, \|\cdot\|_2^2, 0)$ and $\mathcal{R}_{n,\alpha}(\mathcal{F}_\beta, \|\cdot\|_\infty, 0)$ and (ii) the total variation modulus of continuity $\omega(\varepsilon)$ under different loss functions.

As for (i), due to the Sobolev embedding theorem for Besov space (e.g., Proposition 4.3.20 in Giné and Nickl (2016)), it can be seen as a special case of Corollary 2.1 in Butucea et al. (2020), which is a result on minimax rates for estimating density functions under L_r loss ($r \geq 1$) without contamination but with an α-LDP constraint.

As for (ii), we construct a lower bound on $\omega(\varepsilon)$ by considering a pair of density functions with the help of wavelet basis functions. We note that Theorem 3 in Uppal, Singh and Poczos (2020) studies a robust density estimation problem in the Besov space, under the Besov integral probability metrics but without privacy constraints.

4.2. *Upper bounds.* In this subsection, we study the robustness properties of two α-LDP estimators of the density function, subject to the squared-L_2 loss and the L_∞ loss, in Theorems 8 and 9, respectively. Both estimators are of the form of linear projection estimators, but constructed based on different choices of orthonormal basis functions and privacy mechanisms.

The squared-L_2 loss. To obtain an upper bound on $\mathcal{R}_{n,\alpha}(\mathcal{F}_\beta, \|\cdot\|_2^2, \varepsilon)$, we analyse a projection estimator based on the orthonormal basis for $L^2[0,1]$ consisting of trigonometric functions, that is,

$$\varphi_1(x) = 1, \qquad \varphi_{2j}(x) = \sqrt{2}\cos(2\pi j x) \quad \text{and} \quad \varphi_{2j+1}(x) = \sqrt{2}\sin(2\pi j x), \quad j \in \mathbb{N}_+. \tag{23}$$

With the choices $a_j = \lfloor (j+1)/2 \rfloor$ in (18), the Sobolev space in Definition 6 is characterised by

$$\sum_{j=1}^{\infty} j^{2\beta}\theta_j^2 = \sum_{j=1}^{\infty} j^{2\beta}\left\{\int_0^1 f(x)\varphi_j(x)\,\mathrm{d}x\right\}^2 \leq r^2 < \infty, \tag{24}$$

where $r > 0$ is some universal constant controlling the radius of the ellipsoid. Condition (24) can be translated into smoothness conditions on functions having βth order (weak) derivative in $L^2[0,1]$ (e.g., Proposition 1.14 in Tsybakov (2009)). Our projection estimator $\hat{f}$ is then constructed as follows.

Step 1 (Privatisation). Given data $\{X_i\}_{i=1}^n$, for any $i \in \{1, \dots, n\}$, let $v = [\varphi_j(X_i)]_{j=1}^k \in \mathbb{R}^k$, where $k \in \mathbb{N}_+$ is a pre-specified tuning parameter. Let the corresponding privatised data be $Z_i \in \mathbb{R}^k$ satisfying that $\mathbb{E}_Q[Z_{i,j}|X_i] = \varphi_j(X_i)$ and $Z_{i,j} = \{-B, B\}$ with $B \lesssim \sqrt{k}/\alpha$, where $\mathbb{E}_Q$ denotes the expectation under the privacy mechanism. See Section D.2 of the Supplementary Material (Li, Berrett and Yu (2023)) for full details of this privacy mechanism.

Step 2 (Estimator construction). Let

$$\hat{f} = \sum_{j=1}^k \overline{Z}_j \varphi_j = \sum_{j=1}^k \left(\frac{1}{n} \sum_{i=1}^n Z_{i,j} \right) \varphi_j. \tag{25}$$

In the construction of $\hat{f}$, we choose the trigonometric functions, which satisfy the bounded basis function condition with

$$\max_{1 \le j \le k} \|\varphi_j(\cdot)\|_\infty \le \sqrt{2}. \tag{26}$$

We remark that one may also choose other basis functions, provided that all basis functions are upper bounded by an absolute constant in the function supremum norm. This is to guarantee that the α-LDP constraint (3) holds (Duchi, Jordan and Wainwright (2018)). Another input is the truncation number k, which essentially truncates this infinite-dimensional nonparametric problem to a k-dimensional estimation problem. In fact, the estimator (25) is considered in Section 5.2.2 of Duchi, Jordan and Wainwright (2018), as a nonparametric density estimator, which is shown to be minimax rate optimal in terms of the squared-L_2 norm under local differential privacy without contamination (cf. Corollary 7 in Duchi, Jordan and Wainwright (2018)). Together with Theorem 8 below, we will show that this private procedure is automatically robust, in view of Q2 in Section 1.1.

We remark that $\hat{f}$ defined in (25) is a privatised version of the widely used nonparametric projection estimator (e.g. Chapter 1 in Tsybakov (2009)), defined as

$$\tilde{f} = \sum_{j=1}^k \left\{ \frac{1}{n} \sum_{i=1}^n \varphi_j(X_i) \right\} \varphi_j,$$

which attains the optimal minimax rate under squared-L_2 loss when applied to nonparametric regression problems for functions belonging to the Sobolev class (24) (cf. Theorem 1.9 in Tsybakov (2009)).

THEOREM 8 (The squared-L_2 norm case). *Given i.i.d. random variables $\{X_i\}_{i=1}^n$ with distribution $P_{f_\varepsilon} = (1-\varepsilon)P_f + \varepsilon G$, where P_f denotes the distribution with the density function $f \in \mathcal{F}_\beta$ defined in (18), and G is an arbitrary distribution supported on $[0,1]$, for $\alpha \in (0,1)$ and $\varepsilon \in [0,1]$, the estimator $\hat{f}$ defined in (25) satisfies that*

$$\mathbb{E}_{P_{f_\varepsilon}, Q}\big[\|\hat{f} - f\|_2^2\big] \lesssim \varepsilon^{\frac{4\beta}{2\beta+1}} \vee \big(n\alpha^2\big)^{-\frac{2\beta}{2\beta+2}},$$

when the tuning parameter k is chosen to be $k \asymp \varepsilon^{-\frac{2}{2\beta+1}} \wedge (n\alpha^2)^{\frac{1}{2\beta+2}}$.

Comparing with the lower bound in (21), Theorem 8 shows that under a suitable choice of k, the estimator $\hat{f}$ is minimax rate-optimal in the presence of contamination and a local privacy constraint.

The L_∞ loss. Our next goal is to obtain an upper bound on $\mathcal{R}_{n,\alpha}(\mathcal{F}_\beta, \|\cdot\|_\infty, \varepsilon)$, and it turns out that the trigonometric basis used in the previous analysis under the squared-L_2

risk is not flexible enough to obtain the optimal rate in the L_∞ case. One reason is that it prevents the use of Laplace mechanism to construct optimal private estimator. As discussed in the last paragraph of Section 5.2.2 in Duchi, Jordan and Wainwright (2018), even under L_2 loss, adding Laplace noise to the trigonometric basis leads to sub-optimal rate of order $(n\alpha^2)^{-2\beta/(2\beta+3)}$, which is worse than the corresponding optimal rate $(n\alpha^2)^{-2\beta/(2\beta+2)}$ as that in Theorem 8. We instead exploit the wavelet basis, which allows the use of a simple Laplace mechanism to efficiently privatise the empirical wavelet coefficients (Butucea et al. (2020)).

Given a father wavelet $\phi : [-A, A] \to \mathbb{R}$ and a mother wavelet $\psi : [-A, A] \to \mathbb{R}$, where $A > 0$ is an absolute constant (see, e.g., Section 4.2.1 in Giné and Nickl (2016)), satisfying

$$\int_{\mathbb{R}} \psi(x)\,\mathrm{d}x = 0, \qquad \|\psi\|_\infty < \infty \quad \text{and} \quad \|\phi\|_\infty < \infty, \tag{27}$$

a wavelet basis of $L^2(\mathbb{R})$ can be formed as

$$\{\phi_k = \phi(\cdot - k) : k \in \mathbb{Z}\} \cup \{\psi_{jk} = 2^{j/2}\psi(2^j(\cdot) - k) : j \in \mathbb{N}_+ \cup \{0\}, k \in \mathbb{Z}\}.$$

We denote ϕ_k as ψ_{-1k} to simplify the notation. Given such a basis, we have that for any $f \in L^2(\mathbb{R})$,

$$f(x) = \sum_{j\geq -1}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(x) = \sum_{j\geq -1}\sum_{k\in\mathbb{Z}} \biggl\{\int_{\mathbb{R}} f(x')\psi_{jk}(x')\,\mathrm{d}x'\biggr\}\psi_{jk}(x). \tag{28}$$

Note that one can periodise a wavelet basis of $L^2(\mathbb{R})$ to construct a basis on $L^2[0, 1]$ or apply boundary correction to a wavelet basis for nonperiodic functions supported on $[0, 1]$ while regular wavelet properties including (27) still hold. We refer to Giné and Nickl (2016) and Daubechies (1992) for more detailed introduction on the theory of wavelets.

Since the density functions and wavelet basis are assumed to have compact supports on $[0, 1]$ and $[-A, A]$ respectively, the representation (28) implies that for each resolution level $j \geq -1$, $|\mathcal{N}_j| = |\{k : \beta_{jk} \neq 0\}| \leq 2^j + 2A + 1$. With the choice $a_j = 2^j$ in (18), the Sobolev space in Definition 6 can be characterised by

$$\sum_{j\geq -1} (2^j)^{2\beta}\|\beta_{j\cdot}\|_2^2 = \sum_{j\geq -1} (2^j)^{2\beta}\biggl(\sum_{k\in\mathcal{N}_j} \beta_{jk}^2\biggr) < \infty. \tag{29}$$

Condition (29) can also be translated to smoothness conditions on functions having βth order (weak) derivative in $L^2(\mathbb{R})$ (e.g., Proposition 4.3.20 in Giné and Nickl (2016)).

The projection estimator is then constructed as follows.

Step 1 (Privatisation). Given data $\{X_i\}_{i=1}^n$, for any $i \in \{1, \ldots, n\}$, $j \in \{-1, 0, \ldots, J\}$ and $k \in \mathcal{N}_j$, where $J \in \mathbb{N}_+$ is a pre-specified tuning parameter, let W_{ijk}'s be independent standard Laplace random variables which are also independent of X_i's, the noise parameter σ_J be

$$\sigma_J = C2^{J/2}/\alpha \quad \text{with } C = \bigl(8\lceil A\rceil + 4\bigr)\|\psi\|_\infty\sqrt{2}(\sqrt{2} - 1)^{-1}, \tag{30}$$

and the privatised empirical wavelet coefficients be

$$\hat{\beta}_{jk} = \frac{1}{n}\sum_{i=1}^n Z_{ijk} = \frac{1}{n}\sum_{i=1}^n \{\psi_{jk}(X_i) + \sigma_J W_{ijk}\}.$$

Step 2 (Estimator construction). Let the final estimator be

$$\hat{f}_{\mathrm{Lap}} = \sum_{j=-1}^J \sum_{k\in\mathcal{N}_j} \hat{\beta}_{jk}\psi_{jk}. \tag{31}$$

The tuning parameter J serves as a truncation parameter, reducing an infinite-dimensional nonparametric estimation problem to a finite-dimensional problem with dimensionality being $\sum_{j=-1}^{J} |\mathcal{N}_j|$. The noise level σ_J is chosen to guarantee the α-LDP constraint (Proposition 3.1 in Butucea et al. (2020)).

The estimator (31) is previously studied in Butucea et al. (2020), without the presence of contamination but shown to be optimal in terms of estimating the L_∞-loss, under α-LDP constraint. We remark that Butucea et al. (2020) studies a more general space and a wider range of loss functions. Together with Theorem 9 below, we will show that this private procedure is also automatically robust, again in view of Q2 in Section 1.1.

THEOREM 9 (The L_∞ norm case). *Given i.i.d. random variables $\{X_i\}_{i=1}^n$ with distribution $P_{f_\varepsilon} = (1-\varepsilon)P_f + \varepsilon G$, where P_f denotes the distribution with the density function $f \in \mathcal{F}_\beta$, defined in (18), and G is an arbitrary distribution supported on $[0,1]$, for $\alpha \in (0,1)$ and $\varepsilon \in [0,1]$, the estimator $\hat{f}_{\mathrm{Lap}}$ defined in (31) satisfies that*

$$\mathbb{E}_{P_{f_\varepsilon}, Q}\big(\|\hat{f}_{\mathrm{Lap}} - f\|_\infty\big) \lesssim \left\{\frac{\log(n\alpha^2)}{n\alpha^2}\right\}^{\frac{2\beta-1}{4\beta+2}} \vee \varepsilon^{\frac{2\beta-1}{2\beta+1}},$$

with the tuning parameter J satisfying

$$2^J \asymp \left\{\frac{\log(n\alpha^2)}{n\alpha^2}\right\}^{-\frac{1}{2\beta+1}} \wedge \varepsilon^{-\frac{2}{2\beta+1}}.$$

Comparing with the lower bound in (22), Theorem 9 shows that under a suitable choice of J, the estimator $\hat{f}_{\mathrm{Lap}}$ is minimax rate optimal.

4.3. *Discussion.* In the classical nonparametric density estimation literature (e.g., Tsybakov (2009)), it is known that without the presence of contamination or privacy constraints,

$$\mathcal{R}_{n,\infty}\big(\mathcal{F}_\beta, \|\cdot\|_2^2, 0\big) \asymp n^{-\frac{2\beta}{2\beta+1}} \quad \text{and} \quad \mathcal{R}_{n,\infty}\big(\mathcal{F}_\beta, \|\cdot\|_\infty, 0\big) \asymp \left\{\frac{\log(n)}{n}\right\}^{\frac{2\beta}{4\beta+2}}. \tag{32}$$

With the presence of both contamination and privacy constraints, in this section, we have shown that

$$\mathcal{R}_{n,\alpha}\big(\mathcal{F}_\beta, \|\cdot\|_2^2, \varepsilon\big) \asymp \big(n\alpha^2\big)^{-\frac{2\beta}{2\beta+2}} \vee \varepsilon^{\frac{4\beta}{2\beta+1}} \quad \text{and}$$

$$\mathcal{R}_{n,\alpha}\big(\mathcal{F}_\beta, \|\cdot\|_\infty, \varepsilon\big) \asymp \left\{\frac{\log(n\alpha^2)}{n\alpha^2}\right\}^{\frac{2\beta-1}{4\beta+2}} \vee \varepsilon^{\frac{2\beta-1}{2\beta+1}}.$$

Comparing our results with the classical rates in (32), we see that, similar to the mean estimation problem studied in Section 3, the cost of preserving privacy is manifested through a reduction of the effective sample size from n to $n\alpha^2$ and a loss in the exponent of convergence rate, both of which have been observed in the literature (Duchi, Jordan and Wainwright (2013), Duchi, Jordan and Wainwright (2018), Butucea et al. (2020)). It is, however, interesting to observe that the privacy constraint and the contamination proportion are completely isolated in terms of the fundamental limits. We also remark that the condition of bounded basis function (26) and (27) are critical for both privatising the data and being robust to contamination.

On the other hand—with contamination but without LDP constraints—Uppal, Singh and Poczos (2020) studies a nonprivate version of $\hat{f}_{\mathrm{Lap}}$, which is linear, and a nonlinear wavelet thresholding estimator for robust estimation of densities belonging to Besov space. It is shown

that the wavelet thresholding estimator is optimal for a wide range of loss functions, but for the squared-L_2 loss and L_∞ loss we considered here, it suffices to use linear estimators to achieve optimality in the presence of contamination. (Similar phenomena were observed in Butucea et al. (2020) when estimating functions in Besov space under local privacy constraint.) Therefore, we may again view $\hat{f}_{\text{Lap}}$ as a properly privatised version of a robust estimator that yields optimal performance, the success of which depends crucially on the choice of basis function.

In the robust statistics literature (e.g., Chen, Gao and Ren (2016), Chen, Gao and Ren (2018), Uppal, Singh and Poczos (2020)), an interesting quantity to investigate is the maximum proportion of contamination such that $\mathcal{R}_{n,\alpha}(\theta(\mathcal{P}), \Phi \circ \rho, \varepsilon) \asymp \mathcal{R}_{n,\alpha}(\theta(\mathcal{P}), \Phi \circ \rho, 0)$, that is the maximum proportion of contamination that the estimators can tolerate to obtain the optimal rate without contamination. In the private setting, denoting this quantity as ε^*_α, we have that $\varepsilon^*_\alpha \asymp (n\alpha^2)^{-(2\beta+1)/(4\beta+4)}$ in the squared-L_2 case and $\varepsilon^*_\alpha \asymp \{\log(n\alpha^2)/(n\alpha^2)\}^{1/2}$ in the L_∞ case. Comparing to the nonprivate setting, denoting this quantity as ε^*, where $\varepsilon^* \asymp n^{-1/2}$ in the squared-L_2 case and $\varepsilon^* \asymp \{\log(n\alpha^2)/(n\alpha^2)\}^{\beta/(2\beta-1)}$ in the L_∞ case, we see that private algorithms can tolerate more contamination but at the price of converging at a slower rate, due to the presence of privacy constraints.

We conjecture that the condition that the density function of interest is supported on a compact set is critical in terms of achieving optimality under both contamination and LDP constraints. As a consequence of this compact support condition, we require bounded basis functions—see (26) and (27)—which facilitate our proofs. Another direct consequence of this compact support condition is the following summary.

As for the questions Q1 and Q2 raised in Section 1.1, in this robust density estimation problem, we see that:

- there exist procedures optimal against contamination (Uppal, Singh and Poczos (2020)) that can be properly privatised to achieve optimal performance; and
- there are existing α-LDP procedures (Duchi, Jordan and Wainwright (2018), Butucea et al. (2020)) that are automatically robust and minimax rate optimal.

5. Conclusions. In this paper, we have studied various statistical problems under both Huber's ε-contamination model (1) and LDP constraints (3). For the four problems concerned in this paper (with one left in the Supplementary Material (Li, Berrett and Yu (2023))), we made an attempt to answer Q1 and Q2 in Section 1.1; that is, being aware of the deep connections between robustness and LDP, what we can say about the ability of robust procedures to work with privatised data and about the robustness of private procedures. For all four problems that we studied, we find procedures that are simultaneously robust, privacy-preserving and statistically rate-optimal. We commented on the connections between our methods to those which are used only under contamination or only under LDP constraints, and provided partial answers to these two questions in specific cases.

The optimality of our procedures mostly relies on the knowledge of ε—an upper bound on the contamination level. This is an assumption commonly used in the literature (e.g., Huber (1992), Lai, Rao and Vempala (2016), Prasad, Balakrishnan and Ravikumar (2019), Lugosi and Mendelson (2021)) though impractical, since it is impossible to estimate ε when the contamination distribution is not specified. An overly large input of ε leads to an inflated error bound while a conservative input of ε leads to unjustified error controls. There has been some recent work on general approaches to robust methodology when ε is unknown (e.g., Jain, Orlitsky and Ravindrakumar (2022)). Applied to our procedures, the theoretical guarantees are not immediate since our error bounds hold in expectation rather than almost surely. Nevertheless, it would be interesting to adapt these ideas to further develop private, robust and optimal procedures, which are also adaptive to ε.

We also note that Proposition 1 is a markedly general result, the potential of which is by no means fully exploited in this paper. Based on the current work, which demonstrates the promise of jointly studying robustness and local differential privacy, we will continue working on understanding the interplay between these two areas in other settings, in particular for problems in high dimensions, with Proposition 1 providing a minimax lower bound to start with.

Acknowledgements. We are thankful to the anonymous referees for detailed comments and suggestions, which greatly improved the paper.

Funding. The second author was supported by Engineering and Physical Sciences Reseach Council (EPSRC) New Investigator Award EP/W016117/1.

SUPPLEMENTARY MATERIAL

Supplementary material: On robustness and local differential privacy (DOI: 10.1214/23-AOS2267SUPP; .pdf). All the technical details and some additional results are provided in the Supplementary Material (Li, Berrett and Yu (2023)).

REFERENCES

ACHARYA, J., CANONNE, C. L., LIU, Y., SUN, Z. and TYAGI, H. (2022). Interactive inference under information constraints. *IEEE Trans. Inf. Theory* **68** 502–516. MR4395425 https://doi.org/10.1109/tit.2021.3123905

ACHARYA, J., SUN, Z. and ZHANG, H. (2021). Robust testing and estimation under manipulation attacks. In *International Conference on Machine Learning* 43–53. PMLR.

ARIDOR, G., CHE, Y.-K. and SALZ, T. (2021). The effect of privacy regulation on the data industry: Empirical evidence from GDPR. In *Proceedings of the* 22*nd ACM Conference on Economics and Computation* 93–94.

AVELLA-MEDINA, M. (2020). The role of robust statistics in private data analysis. *Chance* **33** 37–42.

BERRETT, T. and BUTUCEA, C. (2019). Classification under local differential privacy. *Ann. I.S.U.P.* **63** 191–205.

BERRETT, T. B. and BUTUCEA, C. (2020). Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. *Adv. Neural Inf. Process. Syst.* **33** 3164–3173.

BERRETT, T. B., GYÖRFI, L. and WALK, H. (2021). Strongly universally consistent nonparametric regression and classification with privatised data. *Electron. J. Stat.* **15** 2430–2453. MR4255329 https://doi.org/10.1214/21-ejs1845

BERRETT, T. B. and YU, Y. (2021). Locally private online change point detection. *Adv. Neural Inf. Process. Syst.* **34**. 3425–3437.

BRUNEL, V.-E. and AVELLA-MEDINA, M. (2020). Propose, test, release: Differentially private estimation with high probability. Preprint. Available at arXiv:2002.08774.

BUN, M., KAMATH, G., STEINKE, T. and WU, Z. S. (2021). Private hypothesis selection. *IEEE Trans. Inf. Theory* **67** 1981–2000. MR4282337 https://doi.org/10.1109/TIT.2021.3049802

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. MR4091090 https://doi.org/10.3150/19-BEJ1165

CAI, T. T., WANG, Y. and ZHANG, L. (2021). The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *Ann. Statist.* **49** 2825–2850. MR4338894 https://doi.org/10.1214/21-aos2058

CANONNE, C. L., KAMATH, G., MCMILLAN, A., SMITH, A. and ULLMAN, J. (2019). The structure of optimal private tests for simple hypotheses. In *STOC'*19*—Proceedings of the* 51*st Annual ACM SIGACT Symposium on Theory of Computing* 310–321. ACM, New York. MR4003341

CATONI, O. (2012). Challenging the empirical mean and empirical variance: A deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1148–1185. MR3052407 https://doi.org/10.1214/11-AIHP454

CHEN, M., GAO, C. and REN, Z. (2016). A general decision theory for Huber's ϵ-contamination model. *Electron. J. Stat.* **10** 3752–3774. MR3579675 https://doi.org/10.1214/16-EJS1216

CHEN, M., GAO, C. and REN, Z. (2018). Robust covariance and scatter matrix estimation under Huber's contamination model. *Ann. Statist.* **46** 1932–1960. MR3845006 https://doi.org/10.1214/17-AOS1607

CHERUKURI, A. and HOTA, A. R. (2021). Consistency of distributionally robust risk- and chance-constrained optimization under Wasserstein ambiguity sets. *IEEE Control Syst. Lett.* **5** 1729–1734. MR4213934

CHEU, A., SMITH, A. and ULLMAN, J. (2021). Manipulation attacks in local differential privacy. In 2021 *IEEE Symposium on Security and Privacy* (*SP*) 883–900. IEEE.

CHHOR, J. and SENTENAC, F. (2022). Robust estimation of discrete distributions under local differential privacy. Preprint. Available at arXiv:2202.06825.

DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **61**. SIAM, Philadelphia, PA. MR1162107 https://doi.org/10.1137/1.9781611970104

DEVROYE, L. and LUGOSI, G. (2001). *Combinatorial Methods in Density Estimation*. *Springer Series in Statistics*. Springer, New York. MR1843146 https://doi.org/10.1007/978-1-4613-0125-7

DEVROYE, L., MEHRABIAN, A. and REDDAD, T. (2018). The total variation distance between high-dimensional Gaussians. Preprint. Available at arXiv:1810.08693.

DIAKONIKOLAS, I., KAMATH, G., KANE, D., LI, J., MOITRA, A. and STEWART, A. (2019). Robust estimators in high-dimensions without the computational intractability. *SIAM J. Comput*. **48** 742–864. MR3945261 https://doi.org/10.1137/17M1126680

DIAKONIKOLAS, I., KAMATH, G., KANE, D. M., LI, J., MOITRA, A. and STEWART, A. (2017). Being robust (in high dimensions) can be practical. In *International Conference on Machine Learning* 999–1008.

DIMITRAKAKIS, C., NELSON, B., MITROKOTSA, A. and RUBINSTEIN, B. I. P. (2014). Robust and private Bayesian inference. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **8776** 291–305. Springer, Cham. MR3295543 https://doi.org/10.1007/978-3-319-11662-4_21

DING, B., KULKARNI, J. and YEKHANIN, S. (2017). Collecting telemetry data privately. *Adv. Neural Inf. Process. Syst*. **30**.

DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence, II. *Ann. Statist*. **19** 633–667. MR1105839 https://doi.org/10.1214/aos/1176348114

DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Local privacy, data processing inequalities, and statistical minimax rates. Preprint. Available at arXiv:1302.3203.

DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2018). Minimax optimal procedures for locally private estimation. *J. Amer. Statist. Assoc*. **113** 182–201. MR3803452 https://doi.org/10.1080/01621459.2017.1389735

DWORK, C. and LEI, J. (2009). Differential privacy and robust statistics. In *STOC'*09*—Proceedings of the* 2009 *ACM International Symposium on Theory of Computing* 371–380. ACM, New York. MR2780083

DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*. *Lecture Notes in Computer Science* **3876** 265–284. Springer, Berlin. MR2241676 https://doi.org/10.1007/11681878_14

ERLINGSSON, Ú., 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* 1054–1067.

ESFANDIARI, H., MIRROKNI, V. and NARAYANAN, S. (2021). Tight and robust private mean estimation with few users. Preprint. Available at arXiv:2110.11876.

FORTI, M. (2021). The deployment of artificial intelligence tools in the health sector: Privacy concerns and regulatory answers within the GDPR. *Eur. J. Leg. Stud*. **13** 29.

GHAZI, B., KUMAR, R., MANURANGSI, P. and NGUYEN, T. (2021). Robust and private learning of halfspaces. In *International Conference on Artificial Intelligence and Statistics* 1603–1611.

GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. *Cambridge Series in Statistical and Probabilistic Mathematics* **40**. Cambridge Univ. Press, New York. MR3588285 https://doi.org/10.1017/CBO9781107337862

GOPI, S., KAMATH, G., KULKARNI, J., NIKOLOV, A., WU, Z. S. and ZHANG, H. (2020). Locally private hypothesis selection. In *Conference on Learning Theory* 1785–1816.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics*: *The Approach Based on Influence Functions*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0829458

HUBER, P. J. (1968). Robust confidence limits. *Z. Wahrsch. Verw. Gebiete* **10** 269–278. MR0242330 https://doi.org/10.1007/BF00531848

HUBER, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in Statistics* 492–518.

HUBER, P. J. (2004). *Robust Statistics*. *Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0606374

HUBER, P. J. and RONCHETTI, E. M. (2009). *Robust Statistics*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2488795 https://doi.org/10.1002/9780470434697

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

JAIN, A., ORLITSKY, A. and RAVINDRAKUMAR, V. (2022). Robust estimation algorithms don't need to know the corruption level. Preprint. Available at arXiv:2202.05453.

JOSEPH, M., MAO, J., NEEL, S. and ROTH, A. (2019). The role of interactivity in local differential privacy. In 2019 *IEEE* 60*th Annual Symposium on Foundations of Computer Science* 94–105. IEEE Comput. Soc. Press, Los Alamitos, CA. MR4228162 https://doi.org/10.1109/FOCS.2019.00015

KAIROUZ, P., OH, S. and VISWANATH, P. (2016). Extremal mechanisms for local differential privacy. *J. Mach. Learn. Res.* **17** Paper No. 17, 51 pp. MR3491111

KAMATH, G., MOUZAKIS, A., SINGHAL, V., STEINKE, T. and ULLMAN, J. (2021). A private and computationally-efficient estimator for unbounded Gaussians. Preprint. Available at arXiv:2111.04609.

KARWA, V. and VADHAN, S. (2017). Finite sample differentially private confidence intervals. Preprint. Available at arXiv:1711.03908.

KOTHARI, P. K., MANURANGSI, P. and VELINGKER, A. (2021). Private robust estimation by stabilizing convex relaxations. Preprint. Available at arXiv:2112.03548.

LAI, K. A., RAO, A. B. and VEMPALA, S. (2016). Agnostic estimation of mean and covariance. In 57*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2016 665–674. IEEE Computer Soc., Los Alamitos, CA. MR3631029

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. MR4337717 https://doi.org/10.3150/21-bej1358

LI, M., BERRETT, T. B. and YU, Y. (2022). Network change point localisation under local differential privacy. *Adv. Neural Inf. Process. Syst.* **35**. 15013–15026.

LI, M., BERRETT, T. B. and YU, Y. (2023). Supplement to "On robustness and local differential privacy." https://doi.org/10.1214/23-AOS2267SUPP

LIU, X., KONG, W., KAKADE, S. and OH, S. (2021). Robust and differentially private mean estimation. Preprint. Available at arXiv:2102.09159.

LUGOSI, G. and MENDELSON, S. (2019). Sub-Gaussian estimators of the mean of a random vector. *Ann. Statist.* **47** 783–794. MR3909950 https://doi.org/10.1214/17-AOS1639

LUGOSI, G. and MENDELSON, S. (2021). Robust multivariate mean estimation: The optimality of trimmed mean. *Ann. Statist.* **49** 393–410. MR4206683 https://doi.org/10.1214/20-AOS1961

MARONNA, R. A., MARTIN, R. D., YOHAI, V. J. and SALIBIÁN-BARRERA, M. (2019). *Robust Statistics: Theory and Methods* (*with R*). *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR3839299

PENSIA, A., JOG, V. and LOH, P.-L. (2020). Robust regression with covariate filtering: Heavy tails and adversarial contamination. Preprint. Available at arXiv:2009.12976.

PRASAD, A., BALAKRISHNAN, S. and RAVIKUMAR, P. (2019). A unified approach to robust mean estimation. Preprint. Available at arXiv:1907.00927.

ROHDE, A. and STEINBERGER, L. (2020). Geometrizing rates of convergence under local differential privacy constraints. *Ann. Statist.* **48** 2646–2670. MR4152116 https://doi.org/10.1214/19-AOS1901

TANG, J., KOROLOVA, A., BAI, X., WANG, X. and WANG, X. (2017). Privacy loss in Apple's implementation of differential privacy on MacOS 10.12. Preprint. Available at arXiv:1709.02753.

TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794

UPPAL, A., SINGH, S. and POCZOS, B. (2020). Robust density estimation under Besov IPM losses. *Adv. Neural Inf. Process. Syst.* **33**.

WARNER, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* **60** 63–69.

WASSERMAN, L. and ZHOU, S. (2010). A statistical framework for differential privacy. *J. Amer. Statist. Assoc.* **105** 375–389. MR2656057 https://doi.org/10.1198/jasa.2009.tm08651

YATRACOS, Y. G. (1985). Rates of convergence of minimum distance estimators and Kolmogorov's entropy. *Ann. Statist.* **13** 768–774. MR0790571 https://doi.org/10.1214/aos/1176349553