The Annals of Statistics
2026, Vol. 54, No. 2, 692–717
https://doi.org/10.1214/25-AOS2583

VERSATILE DIFFERENTIALLY PRIVATE LEARNING FOR GENERAL LOSS FUNCTIONS

BY QILONG LU[1,a], SONG XI CHEN[2,b] AND YUMOU QIU[3,c]

[1]*Guanghua School of Management, Peking University,* [a]*lu_qilong@stu.pku.edu.cn*
[2]*Department of Statistics and Data Science, Tsinghua University,* [b]*sxchen@tsinghua.edu.cn*
[3]*School of Mathematical Sciences and Center for Statistical Science, Peking University,* [c]*qiuyumou@math.pku.edu.cn*

This paper aims to provide a versatile privacy-preserving release mechanism along with a unified approach for subsequent parameter estimation and statistical inference. We propose a privacy mechanism based on zero-inflated symmetric multivariate Laplace (ZIL) noise, which requires no prior specification of subsequent analysis tasks, allows for general loss functions under minimal conditions, imposes no limit on the number of analyses, and is adaptable to increasing data volume in online scenarios. We derive the trade-off function for the proposed ZIL mechanism, which characterizes its privacy protection level. Furthermore, to formalize the local differential privacy (LDP) property of the ZIL mechanism, we extend the classical ε-LDP to a more general f-LDP framework. To address scenarios where only individual attribute values require protection, we propose attribute-level differential privacy (ADP) and its local version. Within the M-estimation framework, we introduce a novel doubly random (DR) corrected loss for the ZIL mechanism, which yields consistent and asymptotically normal M-estimates under differential privacy constraints. The proposed approach is computationally efficient and does not require numerical integration or differentiation for noisy data. It applies to a broad class of loss functions, including nonsmooth ones. Two alternative estimators for smooth loss are also proposed with asymptotic properties. The cost of privacy in terms of estimation efficiency for these three estimators is evaluated both theoretically and numerically.

1. Introduction. Data privacy has become an increasing concern with the phenomenal growth in the amount of personal information stored in digital devices, such as health data, web search histories, and personal preferences (Erlingsson, Pihur and Korolova (2014), Apple Differential Privacy Team (2017), Ding, Kulkarni and Yekhanin (2017)). Analyzing data under privacy protection involves two roles: data providers and data analysts. The data provider stores the original data and releases privacy-preserved data or statistics to data analysts through a certain mechanism. The data analysts conduct analysis for various tasks based on the data or statistics released by the data provider (Dwork and Roth (2013)). Privacy-preserving mechanisms aim to protect the personal information in the original data while allowing data analysts to extract useful information from the outputs.

To quantify the privacy protection level of a data release mechanism, Dwork et al. (2006a) proposed the concept of (ε, δ)-differential privacy ((ε, δ)-DP) that measures the similarity of the distributions of the outputs when the recording of an arbitrary sample is changed while all other samples are kept the same. Based on the DP framework, various privacy measures have been developed for different goals, for example, protecting explicit specification of the information in the data (Kifer and Machanavajjhala (2014)) and edge privacy for social network data (Nissim, Raskhodnikova and Smith (2007)). To address scenarios where no trusted

Received July 2025.
MSC2020 subject classifications. Primary 62-11; secondary 68P27.
Key words and phrases. Differential privacy, M-estimation, symmetric multivariate Laplace distribution, zero-inflated symmetric multivariate Laplace distribution.

curator is available, Kasiviswanathan et al. (2011) and Duchi, Jordan and Wainwright (2013) introduced ε-local differential privacy (LDP), which ensures privacy by applying randomization directly on each user's data before collection. Dong, Roth and Su (2022) linked differential privacy with the type I and type II error in hypothesis testing and proposed the concept of f-DP for a trade-off function f.

Existing DP mechanisms can be broadly classified into three categories. One adds noise to data outputs, such as histograms (Lei (2011)), summary statistics, or estimators (Avella-Medina (2021)), which is called the sensitivity method (Dwork et al. (2006b)). Another category introduces noises within the computational procedure for a particular task. The noisy stochastic gradient descent (Noisy-SGD) is a representative of this category, which adds noise to the SGD procedure of a specific loss function (Rajkumar and Agarwal (2012), Bassily, Smith and Thakurta (2014)), and the objective function perturbation (Chaudhuri, Monteleoni and Sarwate (2011)) is another representative. For this category the data analysts and the data provider need to communicate in each step of the computation task, and the noise is added in each step by the data provider (Dong, Roth and Su (2022)).

These two categories of methods are not versatile, as the data privacy protection is task-specific, and the data analysts can only conduct a limited number of analyses using perturbed statistics or SGD. Specifically, we say a differential privacy mechanism is versatile if it is: (i) applicable to general estimation tasks, (ii) applicable to any number of analyses, and (iii) adaptive to increasing data volume in online scenarios.

The third category of method adds noise to the original data (Warner (1965), Duchi, Jordan and Wainwright (2018)) or synthetic data (Zhang et al. (2024)) directly. It is more versatile than the first two categories, as the noises are not tied to any analysis task and the corrupted data can be used repeatedly. This approach also offers stronger privacy protection, as the data are randomized before being transmitted, thus inherently satisfying LDP. Adding either the Laplace or Gaussian noise is a common practice (Wasserman and Zhou (2010), Dong, Roth and Su (2022)). However, the existing methods in this category require stronger conditions on the loss function for statistical inference. The subsequent analysis of the noisy data is closely related to the deconvolution method for the measurement error problems (Carroll and Hall (1988), Fan (1991)). To remove the effects of the added noise in the estimation, the deconvoluted loss function can be formulated, leading to a corrected loss (Stefanski (1989), Wang, Stefanski and Zhu (2012)). The forms of the corrected loss depend on the type of added noise. For Gaussian noise the corrected loss involves numerical integration because the inverse Fourier transform does not have a closed-form solution. For the componentwise independent Laplace noise, the corrected loss involves high-order differentiation of the underlying loss function. These suggest that applying both types of noise will impose strong conditions on the loss and bring inconvenience in the statistical estimation and inference.

Recent works have explored the estimation error under LDP. Duchi, Jordan and Wainwright (2018) considered the estimation error under ε-LDP for certain statistical tasks. Duchi and Ruan (2024) showed the same order of estimation error (at the order of $1/(n\varepsilon^2)$) in a more general framework under ε-LDP. Both results indicate that the estimation error induced by LDP is nonnegligible compared to the intrinsic estimation error. All these works designed their data release mechanisms according to specific tasks, which inherently limits the ability to meet various analytical demands for a dataset while preserving privacy. This limitation makes the data privacy procedure less versatile.

In addition to achieving the local differential privacy, we aim to provide a unified and versatile privacy-preserving mechanism that facilitates consistent parameter estimation and inference under the framework of M-estimation, without requiring the loss function to be smooth or using numerical Fourier inversion. As a result, the procedure would permit a wide range of inference tasks with even nonsmooth loss functions. To achieve these goals, we

consider the zero-inflated symmetric multivariate Laplace (ZIL) distribution as the noise distribution. The ZIL noise distribution is designed to simplify the subsequent parameter estimation and statistical inference, avoiding the need to compute derivatives of the loss function or to perform integrations in the deconvoluted loss. To quantify the privacy level of the ZIL mechanism, we extend the classical ε-LDP to a more general f-LDP framework.

Some recent works, including label-DP (Beimel, Nissim and Stemmer (2016), Wang and Xu (2019)) and componentwise LDP (CLDP) (Amorino and Gloter (2025)) attempted to relax the definition of differential privacy. We also propose a relaxation named f-attribute level DP (f-ADP) and its local version (f-LADP). In contrast to making hypothesis testing attacks on two different people difficult, f-ADP and f-LADP aim to make hypothesis testing attacks on two different attribute values difficult. The DP properties of the ZIL mechanism are developed under both the f-LDP and f-LADP framework by deriving the trade-off function and their asymptotic limits as the data dimension diverges. The connection between the ZIL trade-off function and the (ε, δ)-DP criteria is derived which allows interpretation of the DP level via the (ε, δ)-DP criteria and provides a guideline for selecting the noise level and the zero proportion of the ZIL distribution under a given privacy budget.

To facilitate consistent M-estimation with data released under the ZIL mechanism, we propose a doubly random (DR) procedure that additionally adds symmetric multivariate Laplace (SL) noises to the output of ZIL mechanism to obtain an unbiased corrected loss function. The proposed method is versatile for a general class of M-estimation that does not require differentiation or numerical Fourier integration and is free of tuning parameters, as would be for the case of the deconvolution density estimation. It works for nonsmooth loss functions, including quantile regression, classification using the hinge loss for support vector machines, and neural network models using the ReLU activation function.

We show that the proposed doubly random (DR) corrected loss estimator is consistent and asymptotically normal. The variance of the DR estimator is obtained, which can be easily estimated for inference purposes. Compared with the estimation procedures with the well-known Gaussian and Laplace mechanisms that add independent normal and Laplace errors, respectively, the proposed DR procedure is much simpler, avoiding integration (Gaussian noise case) and differentiation (Laplace noise case), and works for more general loss functions without requiring them to be smooth. If the data analysts' tasks are constrained to second-order smooth loss functions, we further propose a smoothed doubly random (SDR) corrected loss that utilizes the smoothness of the loss function for parameter estimation. Although the privacy guarantee achieved by our proposed mechanism is f-LDP rather than ε-LDP, we find that the estimation error of the SDR estimator is, in a certain sense, comparable to the results of Duchi, Jordan and Wainwright (2018) and Duchi and Ruan (2024). A thorough discussion on the trade-off between privacy protection level and estimation efficiency of the proposed estimators is presented as well.

The paper is organized as follows. Section 2 reviews the necessary concepts and properties of differential privacy and f-DP, and introduces f-ADP and the corresponding local versions, f-LDP and f-LADP. Section 3 describes the proposed ZIL mechanism and derives its trade-off function and the associated properties. Section 4 proposed the doubly random corrected loss method for parameter estimation. Section 5 establishes the theoretical results for the proposed DR estimator. Section 6 introduces two estimators tailored for second-order smooth loss functions, together with an analysis of their theoretical properties. Section 7 conducts simulation studies to verify the theoretical findings. All technical proofs and additional numerical results are relegated to the Supplementary Material (SM) (Lu, Chen and Qiu (2026)).

2. Background on differential privacy. In this paper we construct a differentially private mechanism that releases noisy data with a novel noise distribution and propose a new denoising approach to achieve versatile estimation and statistical inference for general analysis tasks chosen by analysts. We first introduce basic concepts of central DP and then the definitions of local mechanisms and local DP, which are necessary for versatile DP learning.

2.1. *Central DP.* Let $\mathcal{X} \subset \mathbb{R}^d$ denote the space of the d-dimensional data X from an individual and $\mathcal{X}^n$ denote the space of a dataset $\mathbf{X} = \{X_1, \ldots, X_n\}$ containing n individuals, where $X_i = (X_{i1}, \ldots, X_{id})^\top$, and the superscript n denotes n Cartesian products of $\mathcal{X}$. Let $\mathcal{X}_j$ denote the space of each component of X for $j = 1, \ldots, d$. A differential privacy (DP) mechanism Mech$(\cdot)$ is a randomized algorithm that releases some (randomized) statistics or noisy data of the input dataset. It is a randomized mapping defined on the space of datasets $\mathcal{X}^n$ to some abstract space $\mathcal{R}$. Randomized mapping means the data release mechanism would add noise into the input dataset to preserve privacy. For two datasets $\mathbf{X} = \{X_1, \ldots, X_n\}$ and $\mathbf{X}' = \{X'_1, \ldots, X'_n\}$, we define

$$\Delta_{\mathrm{I}}(\mathbf{X}, \mathbf{X}') = \big|\{i : X_i \neq X'_i, 1 \leq i \leq n\}\big| \tag{2.1}$$

as the number of individuals whose data records differ between the two datasets. Here, for a set $\mathcal{A}$, the notation $|\mathcal{A}|$ denotes its cardinality. A permutation-invariant definition of Δ_{I} is

$$\Delta_{\mathrm{I}}(\mathbf{X}, \mathbf{X}') = \min_{\pi \in \mathcal{S}_n} \big|\{i \in [n] : X_i \neq X'_{\pi(i)}\}\big|,$$

where $\mathcal{S}_n$ denotes the symmetric group on n elements, that is, the set of all bijections (permutations) $\pi : [n] \to [n]$. For each $\pi \in \mathcal{S}_n$, the notation $\pi(i)$ refers to the image of the index i under the permutation π. For notation simplicity we use the simpler form in (2.1). The subscript "I" denotes the individual-level difference to differentiate from the attribute-level difference in (2.3) where the subscript "A" will be used. A DP mechanism would make the distributions of Mech$(\mathbf{X})$ and Mech$(\mathbf{X}')$ being similar for any pairs of neighboring datasets $\mathbf{X}$ and $\mathbf{X}'$ with only one individual having different records so that the information of X_i is preserved for all i under this mechanism.

DEFINITION 2.1 ((ε, δ)-DP, Dwork et al. (2006b)). For any nonnegative ε and δ, a randomized algorithm Mech : $\mathcal{X}^n \to \mathcal{R}$ is (ε, δ)-differentially private if, for every pair of datasets $\mathbf{X}, \mathbf{X}' \in \mathcal{X}^n$ with $\Delta_{\mathrm{I}}(\mathbf{X}, \mathbf{X}') = 1$ and every measurable set $S \subseteq \mathcal{R}$,

$$\mathbb{P}\big(\mathrm{Mech}(\mathbf{X}) \in S\big) \leq e^{\varepsilon} \cdot \mathbb{P}\big(\mathrm{Mech}(\mathbf{X}') \in S\big) + \delta,$$

where the probability measure $\mathbb{P}$ is conditioned on the datasets $\mathbf{X}$, $\mathbf{X}'$ and is induced by the randomness of Mech$(\cdot)$ only. Particularly, $(\varepsilon, 0)$-DP is denoted as ε-DP.

As differential privacy is measured by the similarity between the conditional distributions of Mech$(\mathbf{X})$ and Mech$(\mathbf{X}')$, given $\mathbf{X}$ and $\mathbf{X}'$, it can be characterized from the perspective of testing the hypotheses

$$H_0 : \text{the original dataset is } \mathbf{X} \quad \text{vs.} \quad H_1 : \text{the original dataset is } \mathbf{X}' \tag{2.2}$$

based on the released results from the mechanism Mech. Let P and Q be two generic notations for hypotheses testing, representing the distributions under the null and alternative hypotheses, respectively. The probabilities of type I and type II errors of a test function ϕ are, respectively, $\alpha_\phi := \mathbb{E}_P[\phi]$ and $\beta_\phi := 1 - \mathbb{E}_Q[\phi]$, where $\mathbb{E}_P[\phi] := \int \phi(x)\,\mathrm{d}P(x)$ and $\mathbb{E}_Q[\phi] := \int \phi(x)\,\mathrm{d}Q(x)$. Below is the definition of the trade-off function.

DEFINITION 2.2. For any two probability distributions P and Q on the same probability space, define the trade-off function $T(P, Q) : [0, 1] \to [0, 1]$ as

$$T(P, Q)(\alpha) = \inf_{\phi}\{\beta_\phi : \alpha_\phi \le \alpha\},$$

where the infimum is taken over all (measurable) test functions.

The following lemma from Dong, Roth and Su (2022) provides the necessary and sufficient condition for trade-off functions.

LEMMA 2.1. *A function* $f : [0, 1] \to [0, 1]$ *is a trade-off function if and only if* f *is convex, continuous, nonincreasing, and* $f(x) \le 1 - x$ *for* $x \in [0, 1]$.

For the hypotheses in (2.2), to simplify notations, we also use $\mathrm{Mech}(\mathbf{X})$ and $\mathrm{Mech}(\mathbf{X}')$ to denote the conditional distributions of the released results, given $\mathbf{X}$ and $\mathbf{X}'$, respectively, when there is no confusion. The trade-off function $T(\mathrm{Mech}(\mathbf{X}), \mathrm{Mech}(\mathbf{X}'))(\alpha)$ fully characterizes the optimal test for distinguishing the original data being $\mathbf{X}$ or $\mathbf{X}'$. The definition of f-DP is based on a trade-off function f, presented in the following.

DEFINITION 2.3 (f-DP (Dong, Roth and Su (2022))). Let f be a trade-off function. A mechanism Mech is said to be f-differentially private (f-DP) if for all $\alpha \in [0, 1]$

$$T\big(\mathrm{Mech}(\mathbf{X}), \mathrm{Mech}(\mathbf{X}')\big)(\alpha) \ge f(\alpha)$$

for all neighboring datasets $\mathbf{X}, \mathbf{X}' \in \mathcal{X}^n$ with $\Delta_{\mathrm{I}}(\mathbf{X}, \mathbf{X}') = 1$.

In Definition 2.3 the measure of privacy protection is reflected by the function f. For each $\alpha \in (0, 1)$, if the probability of the analyst's type I error for distinguishing two adjacent datasets is less than α, then the probability of the type II error must be above $f(\alpha)$. For two trade-off functions f_1 and f_2, if $f_1(\alpha) \ge f_2(\alpha)$ for all $\alpha \in [0, 1]$, higher level of privacy is protected by an f_1-DP mechanism. However, statistical inference would be more difficult as larger noises need to be added to this mechanism.

Let $G_\mu(\alpha) := \Phi(\Phi^{-1}(1-\alpha) - \mu)$ for $\alpha \in [0, 1]$, where $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ denote the cumulative distribution function and quantile function of the standard normal distribution, respectively. A special case of f-DP is the μ-GDP (Gaussian differential privacy) with $f = G_\mu$, built upon the standard Gaussian noise. As $G_\mu(\alpha)$ monotonically decreases as μ increases for a given α, a smaller μ indicates higher privacy protection level of a μ-GDP mechanism.

Let $\mathrm{Proc} : \mathcal{R} \to \mathcal{Z}$ denote a randomized algorithm that maps the released result $\mathrm{Mech}(\mathbf{X}) \in \mathcal{R}$ of an f-DP mechanism Mech to some space $\mathcal{Z}$, yielding a new mechanism denoted by $\mathrm{Proc} \circ \mathrm{Mech}$. The following two propositions of an f-DP mechanism are from Wasserman and Zhou (2010) and Dong, Roth and Su (2022).

PROPOSITION 2.1. *A mechanism* Mech *is* (ε, δ)*-DP if and only if the* Mech *is* $f_{\varepsilon,\delta}$*-DP, where* $f_{\varepsilon,\delta}(\alpha) := \max\{0, 1 - \delta - e^{\varepsilon}\alpha, e^{-\varepsilon}(1 - \delta - \alpha)\}$ *for* $\alpha \in [0, 1]$.

PROPOSITION 2.2. (*i*) *If a mechanism* Mech *is* f*-DP, then its postprocessing* $\mathrm{Proc} \circ \mathrm{Mech}$ *is also* f*-DP.* (*ii*) *For any two distributions* P *and* Q*, we have* $T(\mathrm{Proc}(P), \mathrm{Proc}(Q)) \ge T(P, Q)$.

Proposition 2.1 shows the equivalence between f-DP and (ε, δ)-DP. Proposition 2.2 shows that postprocessing a mechanism does not compromise the privacy guarantees already provided by the mechanism. This property ensures that the privacy protection level of an estimator computed from the output of an f-DP mechanism is preserved.

The measures of differential privacy in Definitions 2.1 and 2.3 depend on the definition of the distance for neighboring datasets. Under the distance $\Delta_{\mathrm{I}}(\mathbf{X}, \mathbf{X}')$ in (2.1), if an analyst cannot distinguish any pair of datasets $\mathbf{X}, \mathbf{X}' \in \mathcal{X}^n$ with $\Delta_{\mathrm{I}}(\mathbf{X}, \mathbf{X}') = 1$, the analyst would not know whether any particular individual is part of the original dataset based on the output of the privacy mechanism. However, in some scenarios it is unnecessary to protect the privacy of all variables of an individual as a whole (Kifer and Machanavajjhala (2014)). For example, in survey sampling of yearly income, we may not need to preserve the information of which individual is sampled but only to ensure that each attribute of each individual cannot be inferred from the mechanism's output. We refer to this relaxed version of DP as attribute-level DP (ADP). Let

$$\Delta_{\mathrm{A}}(\mathbf{X}, \mathbf{X}') = \big|\{(i, j) : X_{ij} \neq X'_{ij}, 1 \leq i \leq n, 1 \leq j \leq d\}\big| \tag{2.3}$$

be the attribute-level distance between two datasets $\mathbf{X}$ and $\mathbf{X}'$. In the following we formally define ADP under the attribute-level distance $\Delta_{\mathrm{A}}(\mathbf{X}, \mathbf{X}')$, which relaxes the f-DP in Definition 2.3 under the global distance $\Delta_{\mathrm{I}}(\mathbf{X}, \mathbf{X}')$.

DEFINITION 2.4 (Attribute differential privacy). Let f be a trade-off function. A mechanism Mech is said to be f-attribute differentially private (f-ADP) if for all $\alpha \in [0, 1]$

$$T\big(\mathrm{Mech}(\mathbf{X}), \mathrm{Mech}(\mathbf{X}')\big)(\alpha) \geq f(\alpha)$$

for all neighboring datasets $\mathbf{X}, \mathbf{X}' \in \mathcal{X}^n$ with $\Delta_{\mathrm{A}}(\mathbf{X}, \mathbf{X}') = 1$. Furthermore, a mechanism is said to be (ε, δ)-ADP if it is $f_{\varepsilon,\delta}$-ADP.

An f-ADP mechanism is characterized by the hypotheses to distinguish two neighboring datasets which only differ in one attribute of one individual under the attribute-level distance $\Delta_{\mathrm{A}}(\mathbf{X}, \mathbf{X}')$. The edge differential privacy for social network data (Nissim, Raskhodnikova and Smith (2007), Chang et al. (2024)), which aims to protect the information of whether each edge exists or not in a network, is a special form of ADP. Compared to f-DP that prevents data analysts from distinguishing any difference in each individual based on the output of a mechanism, f-ADP prevents data analysts from distinguishing any difference in each attribute of each individual. It also requires adding a perturbation to every attribute in the data. Since f-ADP is simply f-DP based on a different definition of neighboring datasets, it inherits all the properties of f-DP. When the original data for each individual is one-dimensional, f-ADP is equivalent to f-DP. However, when the dimension $d > 1$, f-ADP is a more relaxed measure of differential privacy than f-DP and allows for higher data utility.

Compared to label-DP (Beimel, Nissim and Stemmer (2016), Wang and Xu (2019)), which only perturbs response variables and releases the original covariates, ADP protects the privacy of all attributes in a dataset. Nevertheless, the concept of ADP can be extended to the case of only protecting a subset of variables $\mathcal{J}$ and hence aligns with the concept of label DP. Given variables $X_{\mathcal{J}} = (X_j : j \in \mathcal{J})$, we can extend Definition 2.4 of ADP by defining random mechanisms $\mathrm{Mech}(\cdot)$ solely on $\mathbf{X}_{\mathcal{J}} = \{X_{i,\mathcal{J}}\}_{i=1}^n$ and the attribute-level distance $\Delta_{\mathrm{A}}(\mathbf{X}_{\mathcal{J}}, \mathbf{X}'_{\mathcal{J}}) = |\{(i, j) : X_{ij} \neq X'_{ij}, 1 \leq i \leq n, j \in \mathcal{J}\}|$ on $\mathcal{J}$.

2.2. *Local DP*. While the differential privacy (DP) provides privacy guarantees by adding noise to the output of a global statistic, such a centralized approach may not be suitable for designing versatile mechanisms that ensure privacy at the individual input while accommodating flexible downstream uses. It is natural to consider mechanisms that perturb the data locally to achieve versatile learning, before any aggregation or analysis takes place. This motivates the locally differential privacy (LDP), a stricter notion of privacy in which each data provider randomizes their own data independently. To give a clear definition of LDP, we first introduce the notion of the noninteractive local mechanism.

DEFINITION 2.5 (Local Mechanism). A mechanism Mech defined on $\mathcal{X}^n$ is said to be (noninteractive) local if there exist n mutually independent mechanisms $\{\text{Mech}_i\}_{i=1}^n$, each defined on $\mathcal{X}$, such that, for every $\mathbf{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$,

$$\text{Mech}(\mathbf{X}) = \big(\text{Mech}_1(X_1), \ldots, \text{Mech}_n(X_n)\big),$$

and $\{\text{Mech}_i\}_{i=1}^n$ are referred to as the independent local components of Mech.

In this paper, as we propose to add noise to each observation separately, it suffices to consider noninteractive local mechanisms. Kasiviswanathan et al. (2011) and Duchi, Jordan and Wainwright (2013) introduced ε-LDP. The following definition extends it to f-LDP.

DEFINITION 2.6 (f-LDP). A mechanism Mech satisfies f-LDP if it is local and the n independent local components $\{\text{Mech}_i\}_{i=1}^n$ satisfies for all $\alpha \in [0, 1]$,

$$T\big(\text{Mech}_i(X), \text{Mech}_i(X')\big)(\alpha) \geq f(\alpha)$$

for all $X, X' \in \mathcal{X}$ and $i = 1, \ldots, n$.

Proposition 2.1 establishes that ε-DP is equivalent to $f_{\varepsilon,0}$-DP. Consequently, the ε-LDP of Duchi, Jordan and Wainwright (2013) is equivalent to $f_{\varepsilon,0}$-LDP. The f-LDP is a more general description of the privacy level of local mechanisms. Note that Definition 2.6 of local privacy can be naturally extended to varying privacy levels across individuals, where $T(\text{Mech}_i(X), \text{Mech}_i(X'))(\alpha) \geq f_i(\alpha)$ for possibly distinct trade-off functions $\{f_i(\alpha)\}$. Particularly, if $f_i(\alpha) = 0$ for all $\alpha \in [0, 1]$, the ith observation has no privacy protection, and X_i becomes publicly available data. A DP learning with public data was considered in Bie, Kamath and Singhal (2022).

The following theorem presents the relationship between DP and LDP.

THEOREM 2.1. *A mechanism is f-LDP if and only if it is local and satisfies f-DP.*

The proof is deferred to Section S1 in the SM. From Theorem 2.1 the notion of LDP can be decomposed into two aspects: one on the mechanism being local and the other on the level of DP the mechanism. Similar to Definition 2.6, we extend the definition of f-ADP to f-local attribute level DP (f-LADP).

DEFINITION 2.7 (f-LADP). A mechanism is f-LADP if it is local and the n independent local components $\{\text{Mech}_i\}_{i=1}^n$ satisfies that, for all $\alpha \in [0, 1]$,

$$T\big(\text{Mech}_i(X), \text{Mech}_i(X')\big)(\alpha) \geq f(\alpha)$$

for all $X, X' \in \mathcal{X}$ with $\Delta_{\text{A}}(X, X') = 1$ and $i = 1, \ldots, n$.

The concept of f-LADP is related to that of componentwise local DP (CLDP). Amorino and Gloter (2025) proposed CLDP for scenarios where different attributes of an individual are independently and privately released by different entities. If the jth released attribute of a CLDP mechanism satisfies ε_j-LDP, this mechanism is said to be $(\varepsilon_1, \ldots, \varepsilon_d)$-CLDP. If a mechanism satisfies $(\varepsilon_1, \ldots, \varepsilon_d)$-CLDP, it is $f_{\varepsilon_{\max},0}$-LADP, where $\varepsilon_{\max} = \max_i\{\varepsilon_i\}$. Therefore, LADP characterizes the weakest DP level across all attributes. However, CLDP requires all components of noise to be independent, while LADP allows componentwise dependent noises for each individual.

The following theorem presents the relationship between ADP and LADP, whose proof is similar to that of Theorem 2.1 and is thus omitted.

THEOREM 2.2. *A mechanism is f-LADP if and only if it is local and f-ADP.*

3. Zero-inflated multivariate Laplace mechanism and its privacy guarantee. We outline a new privacy protection mechanism that adds the zero-inflated symmetric multivariate Laplace (ZIL) noises. Adding noise directly to the data makes the ZIL mechanism versatile, as it requires no prior specification of subsequent analysis tasks, imposes no limit on the number of analyses, and is adaptable to the increasing data volume in online scenarios. The extent of the differential privacy guarantee will be studied as well.

3.1. *Zero-inflated symmetric multivariate Laplace distribution.* We first introduce the symmetric multivariate Laplace (SL) distribution $\mathcal{SL}_d(\Sigma)$ of dimension d with covariance matrix Σ defined via its characteristic function $\Psi_\Sigma^{\mathcal{SL}}(t) = (1 + t^\top \Sigma t/2)^{-1}$, which reduces to the Laplace distribution when $d = 1$. It is noted that, for $d > 1$, $\mathcal{SL}_d(I_d)$ does not correspond to the distribution whose marginal distributions are i.i.d. Laplace random variables with variance 1, where I_d denotes the d-dimensional identity matrix. Let $f_\Sigma^{\mathcal{SL}}$ denote the density of $\mathcal{SL}_d(\Sigma)$. A random vector $S \sim \mathcal{SL}_d(\Sigma)$ can be generated by $S = \sqrt{W}X$, where W and X are independent, W follows the exponential distribution Exp(1), and $X \sim N(0, \Sigma)$ (Kotz, Kozubowski and Podgórski (2001)).

For $\delta \in (0, 1)$ and a covariance matrix Σ, we define the zero-inflated symmetric multivariate Laplace (ZIL) distribution $\mathrm{ZIL}(\delta, \Sigma)$ as a mixture of the point mass at 0 and the symmetric multivariate Laplace distribution $\mathcal{SL}_d(\Sigma)$. Let $S \sim \mathcal{SL}_d(\Sigma)$ and ξ follow the binary distribution $\mathrm{Bern}(1, \delta)$. The ZIL random variable $Z \sim \mathrm{ZIL}(\delta, \Sigma)$ is generated by $Z = S\mathbf{1}(\xi = 0)$, which equals to 0 with probability δ and equals to S with probability $1 - \delta$, where $\mathbf{1}(\cdot)$ denotes the indicator function. The characteristic function $\Psi_{\delta,\Sigma}^{\mathrm{ZIL}}(t)$ of $\mathrm{ZIL}(\delta, \Sigma)$ is

$$\Psi_{\delta,\Sigma}^{\mathrm{ZIL}}(t) = \delta + (1 - \delta)\Psi_\Sigma^{\mathcal{SL}}(t) = \big(1 + \delta t^\top \Sigma t/2\big)\big(1 + t^\top \Sigma t/2\big)^{-1}.$$

3.2. *ZIL mechanism and privacy guarantee.* We introduce the ZIL mechanism for differential privacy and derive its trade-off function and the associated properties as follows.

DEFINITION 3.1 (ZIL mechanism). Suppose that $\{X_i\}_{i=1}^n$ are d-dimensional random vectors with a compact support $\mathcal{X}$. The ZIL mechanism, which is a randomized algorithm, is

$$\mathrm{Mech}_{\mathrm{ZIL}}\big(\{X_i\}_{i=1}^n; \delta, \lambda\big) = \{X_i + Z_i\}_{i=1}^n,$$

where $Z_1, \ldots, Z_n \overset{\mathrm{IID}}{\sim} \mathrm{ZIL}(\delta, \lambda^2 I_d)$ for a $\lambda > 0$ and $\delta \in (0, 1)$.

Similarly, the symmetric multivariate Laplace (SL) mechanism adds noise from the SL distribution $\mathcal{SL}_d(\Sigma)$ to the original data, which is formally defined in Definition S2.1 in the SM. To derive the trade-off function for the ZIL mechanism, we first introduce that for the SL mechanism. The following hypothesis is related to the privacy protection level of the SL mechanism:

$$H_0 : P = \mathbf{0}_d + \mathcal{SL}_d(I_d) \quad \text{vs.} \quad H_1 : Q = (c, 0, \ldots, 0)^\top + \mathcal{SL}_d(I_d), \tag{3.1}$$

whose trade-off function can be abbreviated as

$$T_{d,c}(\alpha) := T\big(\mathcal{SL}_d(I_d), (c, 0, \ldots, 0)^\top + \mathcal{SL}_d(I_d)\big)(\alpha). \tag{3.2}$$

It is shown in Theorem S2.1 in the SM that the SL mechanism satisfies $T_{d,c}$-LDP for some c. Let

$$T_{d,c,\delta}(\alpha) = \begin{cases} 0 & \text{if } \alpha > 1 - \delta, \\ (1 - \delta)T_{d,c}\left(\dfrac{\alpha}{1 - \delta}\right) & \text{if } 0 \le \alpha \le 1 - \delta \end{cases} \tag{3.3}$$

for $\delta \in (0, 1)$. Lemma S2.3 in the SM shows that $T_{d,c,\delta}(\alpha)$ is the trade-off function between the distributions $\mathrm{ZIL}(\delta, I_d)$ and $(c, 0, \ldots, 0)^\top + \mathrm{ZIL}(\delta, I_d)$. Define the diameter of a set A as $\mathrm{diam}(A) = \sup_{x,x' \in A} \|x - x'\|_2$. The following theorem shows the privacy protection level of the ZIL mechanism.

THEOREM 3.1. *The ZIL mechanism in Definition 3.1 satisfies $T_{d,c_A,\delta}$-LADP and $T_{d,c_I,\delta}$-LDP for $c_A = \max_{j\in\{1,\ldots,d\}} \mathrm{diam}(\mathcal{X}_j)/\lambda$ and $c_I = \mathrm{diam}(\mathcal{X})/\lambda$, respectively.*

The proof is deferred to Section S2.2 in the SM. As the noise-addition step can be executed locally on each individual's device, the ZIL mechanism is local, and a curator is not required for its implementation. From Theorems 2.1 and 2.2, the ZIL mechanism also satisfies $T_{d,c_A,\delta}$-ADP and $T_{d,c_I,\delta}$-DP. Note that the ZIL noise is uncorrelated but dependent across the d dimensions. Therefore, the ZIL mechanism does not satisfy CLDP in Amorino and Gloter (2025).

Theorem 3.1 suggests that the privacy level of the ZIL mechanism depends on $T_{d,c,\delta}(\alpha)$ in (3.3), which is a shrinkage of $T_{d,c}(\alpha)$ in (3.2) toward zero. In the following we investigate the properties of $T_{d,c}$, including its monotonicity with respect to c, d, and α, which will be naturally inherited by $T_{d,c,\delta}$.

PROPOSITION 3.1. (*i*) *For any d, if $c_1 \geq c_2 > 0$, $T_{d,c_1}(\alpha) \leq T_{d,c_2}(\alpha)$ for $\alpha \in [0, 1]$.*
(*ii*) *For any $c > 0$, if $d_1 \geq d_2$, then $T_{d_1,c}(\alpha) \leq T_{d_2,c}(\alpha)$ for $\alpha \in [0, 1]$.*
(*iii*) *For any d and $c > 0$, $T_{d,c}(\alpha)$ is decreasing with respect to α, convex and continuous. Furthermore, $T_{d,c}(\alpha) \leq 1 - \alpha$ for $\alpha \in [0, 1]$ and is symmetric about the 45-degree line such that $T_{d,c}(\alpha) = T_{d,c}^{-1}(\alpha)$ for $\alpha \in [0, 1]$, where $T_{d,c}^{-1}(\alpha) = \inf\{t \in [0, 1] : T_{d,c}(t) \leq \alpha\}$.*
(*iv*) *Part (i)–(iii) also holds for $T_{d,c,\delta}(\alpha)$.*

The proof is deferred to Section S2.3 in the SM. Note that $T_{d,c}(\alpha)$ is the type II error of the most powerful test for the hypotheses in (3.1) at the significance level α. Given a data point $s = (s_1, \ldots, s_d)^\top$, we need to determine whether it originates from the distribution P or Q in (3.1). Recall that $f_{I_d}^{\mathcal{SL}}(x)$ is the density of $\mathcal{SL}_d(I_d)$. According to the Neyman–Pearson Lemma, the most powerful test ϕ^{ump} at the significance level α is

$$\phi^{\mathrm{ump}}(s; k) = \begin{cases} 1 & \text{if } f_{I_d}^{\mathcal{SL}}\big(s - (c, 0, \ldots, 0)^\top\big) > k f_{I_d}^{\mathcal{SL}}(s), \\ 0 & \text{if } f_{I_d}^{\mathcal{SL}}\big(s - (c, 0, \ldots, 0)^\top\big) \leq k f_{I_d}^{\mathcal{SL}}(s), \end{cases}$$

where k satisfies that $\mathbb{E}_{S\sim P}\phi^{\mathrm{ump}}(S; k) = \alpha$. The probabilities of type I and type II errors are $a(k) = \mathbb{E}_{S\sim P}\{\phi^{\mathrm{ump}}(S; k)\} = \alpha$ and $b(k) = 1 - \mathbb{E}_{S\sim Q}\{\phi^{\mathrm{ump}}(S; k)\}$, respectively. Thus, the trade-off function $T_{d,c}(\alpha) = T(P, Q)(\alpha) = b(a^{-1}(\alpha))$. Deriving a closed-form expression for the trade-off function for $d > 1$ is challenging. This is because, according to Kotz, Kozubowski and Podgórski (2001), the density of $\mathcal{SL}_d(\Sigma)$, denoted by $f_\Sigma^{\mathcal{SL}}$, is

$$f_\Sigma^{\mathcal{SL}}(x) = 2(2\pi)^{-\frac{d}{2}}|\Sigma|^{-\frac{1}{2}}\big(x^\top\Sigma^{-1}x/2\big)^{\frac{v}{2}}K_v\big\{\big(2x^\top\Sigma^{-1}x\big)^{1/2}\big\},$$

where $v = (2 - d)/2$ and $K_v(u)$ is the modified Bessel function of the third kind. The complexity of $f_{I_d}^{\mathcal{SL}}$ makes deriving a closed-form expression for $T_{d,c}(\alpha)$ challenging due to the involvement of the modified Bessel function. Simulation is a viable approach to attain values for $T_{d,c}$. It generates samples from the distributions P and Q in (3.1) and simulates $a(k)$ and $b(k)$, leading to the empirical trade-off curve $\{(a(k), b(k)) : k \in [0, \infty)\}$.

However, for $d = 1$, a tangible expression is available for $T_{1,c}$. As $\mathcal{SL}_1(I_1)$ is the Laplace distribution $L(\sqrt{1/2})$, it can be proved that $T_{1,c}(\alpha) = F_{\mathrm{Lap}}(F_{\mathrm{Lap}}^{-1}(1 - \alpha) - \sqrt{2}c)$, where F_{Lap} is the cumulative distribution function of the standard Laplace distribution $L(1)$. In this case,

according to Dwork et al. (2006b), $T_{1,c}(\alpha) \geq f_{\sqrt{2}c,0}(\alpha)$ for $\alpha \in [0,1]$, where $f_{\sqrt{2}c,0}(\alpha)$ is given in Proposition 2.1.

The monotonicity of $T_{d,c}$ and $T_{d,c,\delta}$ with respect to d leads us to consider the asymptotic trade-off functions $T_{d,c}(\alpha)$ and $T_{d,c,\delta}(\alpha)$ for $d \to \infty$, which turns out to be trade-off functions themselves, as shown below. Recall that a SL random vector $S = (S_1, \ldots, S_d)^\top \sim \mathcal{SL}_d(I_d)$ can be generated by $S = \sqrt{W}X$, where $W \sim \text{Exp}(1)$, $X = (X_1, \ldots, X_d)^\top \sim N(0, I_d)$ and W is independent of X. It is noted that the most powerful test for the hypotheses in (3.1) can be written as the conditional density ratio of S_1, given $S_2, \ldots, S_d$, and

$$(d-1)^{-1}\big(S_2^2 + \cdots + S_d^2\big) \xrightarrow{a.s.} W \quad \text{as } d \to \infty$$

according to the law of large numbers. These inspire us to consider testing the hypotheses

$$H_0 : P = (0 + \sqrt{W}X_1, W) \quad \text{vs.} \quad H_1 : Q = (c + \sqrt{W}X_1, W). \tag{3.4}$$

Let $F_c(x) = \int_0^\infty \Phi(c^{-1}w^{1/2}x + c(4w)^{-1/2})e^{-w}\,\mathrm{d}w$ be the cumulative distribution function of $cX_1W^{-1/2} - c^2(2W)^{-1}$. The following proposition derives the trade-off function for the hypotheses (3.4). Let $h_c(\alpha) = F_c^{-1}(1-\alpha)/c$.

PROPOSITION 3.2. *(i) For the hypotheses in* (3.4) *with* $c > 0$, *its trade-off function is*

$$\beta_c(\alpha) = \left[1 + \left\{\frac{\sqrt{2}}{h_c(\alpha) + \sqrt{2 + h_c^2(\alpha)}}\right\}^2\right]^{-1} \exp\left\{-\frac{c}{h_c(\alpha) + \sqrt{2 + h_c^2(\alpha)}}\right\},$$

where $\beta_c(0) = 1$, $\beta_c(1) = 0$ *and* $\beta_c(\alpha) > 0$ *for* $\alpha \in [0,1)$.

(ii) Furthermore, the function $\alpha \mapsto \beta_c(\alpha)$ *is symmetric about the* 45*-degree line, strictly decreasing, convex, continuous, and* $\beta_c(\alpha) \leq 1 - \alpha$ *for* $\alpha \in [0,1]$.

The proof is deferred to Section S2.4 in the SM. Proposition 3.2 shows that $\beta_c(\alpha)$ is the probability of type II error of the most powerful test for the hypotheses in (3.4) with α significance level. Let

$$\beta_{c,\delta}(\alpha) = \begin{cases} 0 & \text{if } \alpha > 1 - \delta, \\ (1-\delta)\beta_c\left(\dfrac{\alpha}{1-\delta}\right) & \text{if } 0 \leq \alpha \leq 1 - \delta. \end{cases}$$

As $\beta_c(\alpha)$ is a trade-off function, $\beta_{c,\delta}(\alpha)$ is also a trade-off function. The following theorem shows that $\beta_c(\alpha)$ and $\beta_{c,\delta}(\alpha)$ are the asymptotic limits of $T_{d,c}(\alpha)$ and $T_{d,c,\delta}(\alpha)$ as $d \to \infty$, and are the lower bounds for all $T_{d,c}(\alpha)$ and $T_{d,c,\delta}(\alpha)$, respectively.

THEOREM 3.2. *(i) For any* $c > 0$ *and* $\delta \in (0,1)$, $\lim_{d\to\infty} \sup_{\alpha\in[0,1]} |T_{d,c}(\alpha) - \beta_c(\alpha)| = 0$ *and* $\lim_{d\to\infty} \sup_{\alpha\in[0,1]} |T_{d,c,\delta}(\alpha) - \beta_{c,\delta}(\alpha)| = 0$.

(ii) For every $d \geq 1$, $c > 0$ *and* $\delta \in (0,1)$, $T_{d,c}(\alpha) \geq \beta_c(\alpha)$ *and* $T_{d,c,\delta}(\alpha) \geq \beta_{c,\delta}(\alpha)$ *for every* $\alpha \in [0,1]$.

The proof is deferred to Section S2.5 in the SM. Note that $\beta_{c,\delta}(\alpha)$ is obtained by substituting $T_{d,c}(\alpha)$ in (3.3) by its limit $\beta_c(\alpha)$. Since $\beta_c(\alpha)$ is an approximation of $T_{d,c}(\alpha)$, $\beta_{c,\delta}(\alpha)$ is an approximation of $T_{d,c,\delta}(\alpha)$ for the ZIL mechanism when d is large. Theorem 3.2(ii) suggests that we can use $\beta_{c,\delta}$ as a lower bound for $T_{d,c,\delta}$, while Theorem 3.2(i) ensures that the lower bound is tight when d is large, meaning that $\beta_{c,\delta}$ can be used to describe the privacy protection level of the ZIL mechanism. Figure 1(a) presents $T_{d,0.5,0.05}$ curves for selected values of d and $\beta_{0.5,0.05}$, which shows that $\beta_{0.5,0.05}$ can well approximate $T_{d,0.5,0.05}$ for d as small as 4.

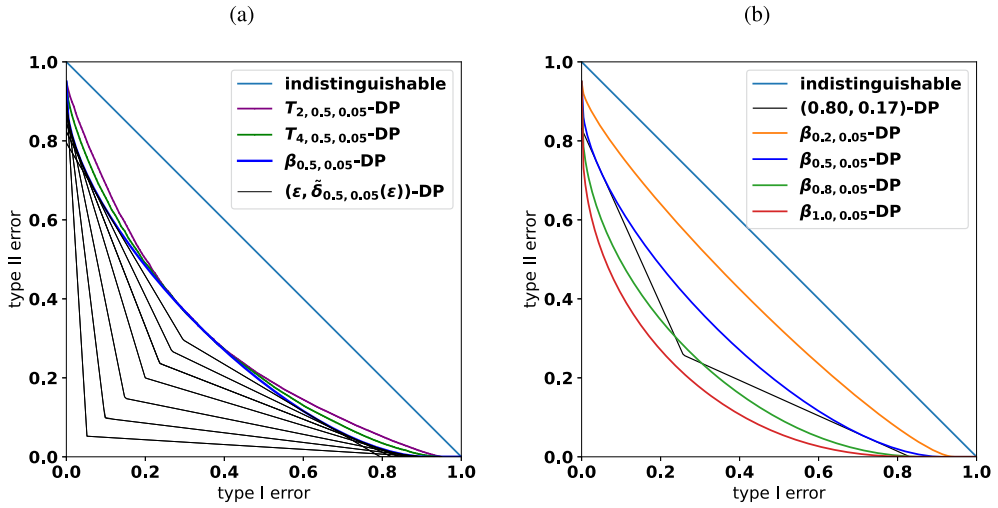

FIG. 1. *Asymptotic trade-off function* $\beta_{0.5,0.05}(\alpha)$ (*blue*) *along with a set of its* $(\varepsilon, \tilde{\delta}_{0.5,0.05}(\varepsilon))$*-DP envelop* (*black polylines*) *for* $\varepsilon = 0.5, 0.7, 0.9, 1.2, 1.6, 2.1, 2.8$ *and the trade-off functions* $T_{2,0.5,0.05}$ (*purple*) *and* $T_{4,0.5,0.05}$ (*green*) *evaluated by* 10^5 *simulations in Panel* (*a*). *The* $(0.8, 0.17)$*-DP polyline* (*black*) *along with four* $\beta_{c,0.05}(\alpha)$ *curves for* $c = 0.2$ (*orange*), 0.5 (*blue*), 0.8 (*green*), *and* 1 (*red*) *in Panel* (*b*). *The indistinguishable curve* $\{(\alpha, 1-\alpha) : \alpha \in [0,1]\}$ (*cyan*) *is also marked.*

With $T_{d,c}$, we can compare the levels of the attribute DP and (individual) DP that the ZIL mechanism satisfies. It is noted that if $c(d)$ as a function of d diverges such that $\lim_{d\to\infty} c(d) = \infty$, then $\lim_{d\to\infty} T_{d,c(d)}(\alpha) \le \lim_{d\to\infty} T_{1,c(d)}(\alpha) = 0$ for $\alpha \in (0,1)$, which implies that the differential privacy guarantee tends to disappear if $c(d) \to \infty$. In Theorem 3.1, $c_A = \max_{j\in\{1,\dots,d\}} \operatorname{diam}(\mathcal{X}_j)/\lambda$ for LADP. If the range of each component $\operatorname{diam}(\mathcal{X}_j)$ is the same, then as d increases, c_A remains unchanged, which prevents the degradation of the LADP for any fixed λ. In contrast, for the (individual) local differential privacy (LDP), $c_I = \operatorname{diam}(\mathcal{X})/\lambda$, which diverges to infinity as $d \to \infty$, implying the LDP cannot be protected as more variables are collected from each individual if the noise variance (related to λ) is unchanged with respect to the dimension.

The β_c-DP (β_c-ADP) and $\beta_{c,\delta}$-DP ($\beta_{c,\delta}$-ADP) are special forms of the f-DP (f-ADP) defined in Definition 2.3 (Definition 2.4). The following theorem provides the level of (ε, δ)-DP ((ε, δ)-ADP) achieved by the $\beta_{c,\delta}$-DP ($\beta_{c,\delta}$-ADP).

THEOREM 3.3. (*i*) *A mechanism is* β_c*-DP if and only if it is* $(\varepsilon, \delta_c(\varepsilon))$*-DP for all* $\varepsilon \ge 0$, *where*

$$\delta_c(\varepsilon) = 1 - e^{\varepsilon}\big(1 - F_c(\varepsilon)\big) - \left[1 + \left\{\frac{\sqrt{2}}{(\varepsilon/c) + \sqrt{2 + (\varepsilon/c)^2}}\right\}^2\right]^{-1} \exp\left\{\frac{-c}{(\varepsilon/c) + \sqrt{2 + (\varepsilon/c)^2}}\right\}.$$

Furthermore, a mechanism is $\beta_{c,\delta}$*-DP if and only if it is* $(\varepsilon, \tilde{\delta}_{c,\delta}(\varepsilon))$*-DP for all* $\varepsilon \ge 0$, *where* $\tilde{\delta}_{c,\delta}(\varepsilon) = 1 - (1-\delta)\cdot(1 - \delta_c(\varepsilon))$.

(*ii*) *The same correspondence holds between* β_c*-ADP and* $(\varepsilon, \delta_c(\varepsilon))$*-ADP, and between* $\beta_{c,\delta}$*-ADP and* $(\varepsilon, \tilde{\delta}_{c,\delta}(\varepsilon))$*-ADP, respectively, and it also holds for LDP and LADP.*

The proof is deferred to Section S2.6 in the SM. Theorem 3.3 indicates that there are families of $f_{\varepsilon,\delta_c(\varepsilon)}$ and $f_{\varepsilon,\tilde{\delta}_{c,\delta}(\varepsilon)}$ envelopes for β_c and $\beta_{c,\delta}$, respectively, as shown in Figure 1(a). From Theorem 3.1 and Theorem 3.2, $\mathrm{Mech}_{\mathrm{ZIL}}(\cdot;\delta,\lambda)$ satisfies $\beta_{c_A,\delta}$-ADP and $\beta_{c_I,\delta}$-DP, where $c_A = \max_{j\in\{1,\dots,d\}} \operatorname{diam}(\mathcal{X}_j)/\lambda$ and $c_I = \operatorname{diam}(\mathcal{X})/\lambda$.

Theorem 3.3 offers a practical guidance for finding (δ, λ) such that $\mathrm{Mech}_{\mathrm{ZIL}}(\cdot; \delta, \lambda)$ achieves a specified (ε', δ')-ADP or (ε', δ')-DP. First of all, for given ε', δ', and δ, we solve c' such that $\tilde{\delta}_{c',\delta}(\varepsilon') = \delta'$. Then $f_{\varepsilon',\delta'}$ is dominated by $\beta_{c',\delta}$. Second, choose $\lambda = \max_{j\in\{1,\ldots,d\}} \mathrm{diam}(\mathcal{X}_j)/c'$ for (ε', δ')-ADP and $\lambda = \mathrm{diam}(\mathcal{X})/c'$ for (ε', δ')-DP. The resulting ZIL mechanism can guarantee the required privacy level. Note that the above choice of λ depends on the choice of δ, and the solution c' exists if and only if $\delta < \delta'$. Furthermore, since $\beta_c(\alpha)$ is decreasing as c increases, for any $c \le c'$, we have $\beta_{c,\delta}(\alpha) \ge \beta_{c',\delta}(\alpha) \ge f_{\varepsilon',\delta'}(\alpha)$. Figure 1(b) demonstrate the above procedure for $\varepsilon' = 0.8$, $\delta' = 0.17$, and $\delta = 0.05$, where $c' = 0.5$ solves $\tilde{\delta}_{c',\delta}(\varepsilon') = \delta'$. It is seen from Figure 1(b) that $\beta_{0.5,0.05}$ precisely takes $f_{0.8,0.17}$ as its envelope.

4. Versatile differentially private M-estimation. In this section we introduce the estimation procedure based on the noisy data from the proposed differentially private ZIL mechanism, demonstrate its versatility for a general class of M-estimation that does not require the smoothness of the loss function, and explain its advantages for statistical inference.

4.1. *Doubly random corrected loss.* We consider the M-estimation in a semiparametric framework for general statistical inference. Let $\ell(x, \theta)$ be a loss function specified by an analyst with $x \in \mathbb{R}^d$ and $\theta \in \Theta \subset \mathbb{R}^p$. Suppose that the original data $\{X_i\}_{i=1}^n$ are drawn from a distribution F with a compact support $\mathcal{X} \subset \mathbb{R}^d$. Imposing this boundedness assumption on the data is mainly due to Definition 2.1, as only when $\mathrm{diam}(\mathcal{X}) < \infty$ can ZIL provide a meaningful DP guarantee. Boundedness conditions on the data (Wasserman and Zhou (2010), Lei (2011), Wang and Xu (2019)), statistics, or gradients (Avella-Medina, Bradshaw and Loh (2023)) are widely used in the existing DP literature. One possible way to relax the boundedness condition is to apply the truncation technique (Amorino and Gloter (2025)), which we will discuss in Section 8. Let $\tilde{X}_i^{(1)} = X_i + Z_i$ be the noisy data from the ZIL mechanism in Definition 3.1, where $\{Z_i\}_{i=1}^n \overset{\mathrm{IID}}{\sim} \mathrm{ZIL}(\delta, \lambda^2 I_d)$. The task of the analyst is to estimate the true parameter $\theta_0 = \mathrm{argmin}_{\theta\in\Theta} \mathbb{E}\ell(X_i, \theta)$ based on the noisy data $\{\tilde{X}_i^{(1)}\}_{i=1}^n$. The ZIL mechanism allows the analyst to choose any form of the loss function of interest under minimum regularity conditions, as demonstrated in the following.

Estimation of θ is related to parameter estimation for data with measurement error. If the original data $\{X_i\}_{i=1}^n$ were observable, one could attain the oracle M-estimator. Replacing the original data with the noisy version $\{\tilde{X}_i^{(1)}\}_{i=1}^n$ instead yields a naive estimator. Both estimators can be expressed as

$$\hat{\theta}_n^{\mathrm{ORA}} = \arg\min_{\theta\in\Theta} \sum_{i=1}^n \ell(X_i, \theta) \quad \text{and} \quad \hat{\theta}_n^{\mathrm{NAI}} = \arg\min_{\theta\in\Theta} \sum_{i=1}^n \ell\big(\tilde{X}_i^{(1)}, \theta\big),$$

where the naive estimator may be inconsistent for θ_0 since there is no guarantee that $n^{-1}\sum_{i=1}^n \ell(\tilde{X}_i^{(1)}, \theta)$ is an unbiased approximation of the underlying risk $\mathbb{E}\ell(X_i, \theta)$. To obtain a consistent estimator, a corrected loss function needs to be constructed which requires $x \mapsto \ell(x, \theta)$ to be sufficiently smooth or necessitates a truncated Fourier transform by numerical integration.

The use of the ZIL noise can avoid these issues and bring a fresh method for consistent and private parameter estimation. We propose a doubly random DP mechanism that adds additional SL noises on the released data $\{\tilde{X}_i^{(1)}\}_{i=1}^n$. Let $\tilde{X}_i^{(2)} = \tilde{X}_i^{(1)} + S_i$ for $i = 1, \ldots, n$, where $\{S_i\}_{i=1}^n$ are i.i.d. from the distribution $\mathcal{SL}_d(\delta\lambda^2 I_d)$, as shown in Algorithm 1. Using the two sets of privacy-protected data $\{\tilde{X}_i^{(1)}\}_{i=1}^n$ and $\{\tilde{X}_i^{(2)}\}_{i=1}^n$, we define a doubly random corrected loss (DR corrected loss)

$$\ell^{\mathrm{DR}}\big(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}, \theta; \delta\big) = \big(1 - \delta^{-1}\big)\ell\big(\tilde{X}_i^{(2)}, \theta\big) + \delta^{-1}\ell\big(\tilde{X}_i^{(1)}, \theta\big).$$

Algorithm 1: Doubly random differentially private (DP) mechanism

Input: original dataset $\{X_i\}_{i=1}^n$, privacy parameters δ and λ of the ZIL mechanism.

Step 1: generate the ZIL noisy data $\{\tilde{X}_i^{(1)}\}_{i=1}^n = \mathrm{Mech}_{\mathrm{ZIL}}(\{X_i\}_{i=1}^n; \delta, \lambda)$;

Step 2: generate the doubly randomized data $\tilde{X}_i^{(2)} = \tilde{X}_i^{(1)} + S_i$ for $i = 1, \ldots, n$ where $\{S_i\}_{i=1}^n \overset{\mathrm{IID}}{\sim} \mathcal{SL}_d(\delta\lambda^2 I_d)$.

Output: $\{\tilde{X}_i^{(1)}\}_{i=1}^n$, $\{\tilde{X}_i^{(2)}\}_{i=1}^n$ and δ.

The term "doubly random" (DR) comes from the use of both $\{\tilde{X}_i^{(1)}\}$ and $\{\tilde{X}_i^{(2)}\}$ to implicitly correct the loss function. Indeed, Theorem 5.1 shows that $\mathbb{E}\ell(X_i, \theta) = \mathbb{E}\ell^{\mathrm{DR}}(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}, \theta; \delta)$, indicating the DR corrected loss is unbiased to the loss function $\ell(X_i, \theta)$ with the original data.

The differentially private doubly random (DR) estimator is defined as

$$\hat{\theta}_n^{\mathrm{DR}} = \operatorname*{argmin}_{\theta\in\Theta} \sum_{i=1}^n \ell^{\mathrm{DR}}\big(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}, \theta; \delta\big). \tag{4.1}$$

The advantage of the DR corrected loss is that it does not involve any differentiation of the loss $\ell(x, \theta)$ or numerical integration, which lowers the requirement on the loss function and reduces computation complexity. This is especially useful for complex or nonsmooth loss functions, such as those in neural networks where ReLU activation functions are involved.

In the following we provide an explanation for the rationale of constructing the DR corrected loss under the special case that $x \mapsto \ell(x, \theta)$ is twice-differentiable. As $S_i \sim \mathcal{SL}_d(\delta\lambda^2 I_d)$, Lemma S4.3 in the SM shows that

$$\mathbb{E}\ell\big(\tilde{X}_i^{(1)}, \theta\big) = \mathbb{E}\ell\big(\tilde{X}_i^{(2)}, \theta\big) - \frac{\delta\lambda^2}{2} \sum_{k=1}^d \mathbb{E}\frac{\partial^2}{\partial x_k^2}\ell\big(\tilde{X}_i^{(2)}, \theta\big). \tag{4.2}$$

As $Z_i \sim \mathrm{ZIL}(\delta, \lambda^2 I_d)$ and $Z_i + S_i \sim \mathcal{SL}_d(\lambda^2 I_d)$, Lemma S4.3 again implies that

$$\mathbb{E}\ell(X_i, \theta) = \mathbb{E}\ell\big(\tilde{X}_i^{(2)}, \theta\big) - \frac{\lambda^2}{2} \sum_{k=1}^d \mathbb{E}\frac{\partial^2}{\partial x_k^2}\ell\big(\tilde{X}_i^{(2)}, \theta\big). \tag{4.3}$$

Combining (4.2) and (4.3), it leads to $\mathbb{E}\ell^{\mathrm{DR}}(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}, \theta; \delta) = \mathbb{E}\ell(X_i, \theta)$, which shows the unbiasedness of the DR corrected loss. Although the above derivation is based on the existence of $\partial^2\ell(x, \theta)/\partial x_k^2$ for all $k = 1, \ldots, d$, $\ell^{\mathrm{DR}}(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}, \theta; \delta)$ does not involve any differentiation, and the same conclusion even holds for $\ell(x, \theta)$ with some discontinuity, as shown in Theorem 5.1.

The following algorithm shows the procedure of the doubly random DP mechanism that facilitates the DR estimation.

Algorithm 1 is a differentially private mechanism. It takes a dataset $\mathbf{X} = \{X_i\}_{i=1}^n$ as input and returns the output of the ZIL mechanism $\{\tilde{X}_i^{(1)}\}_{i=1}^n$, along with its privacy parameter δ and a postprocessed product, the doubly randomized dataset $\{\tilde{X}_i^{(2)}\}_{i=1}^n$. Note that the DR corrected loss for any loss function can be computed using the outputs of the doubly random DP mechanism. According to the postprocessing property in Proposition 2.2, the privacy protection capability of the doubly random DP mechanism in Algorithm 1 is fully inherited from the privacy protection capability of the ZIL mechanism in Definition 3.1.

4.2. *Connection to measurement error literature.* Parameter estimation using privacy-protected data with added noise is well connected to the measurement error problem. The deconvolution method is an important method in the measurement error literature. Stefanski and Carroll (1990) proposed kernel deconvolution estimators for the underlying density. Wang, Stefanski and Zhu (2012) applied deconvolution to a smoothed check loss to correct for measurement error in quantile regression. Firpo, Galvao and Song (2017) also studied quantile regression with noisy covariates, estimating the density via deconvolution and plugging it into the estimating equation. Yang et al. (2020) designed a density estimator for noisy data by solving a linear system corresponding to the deconvolution problem. Kent and Ruppert (2024) provided the convergence rate of the estimator in Yang et al. (2020).

We now present a formulation of the deconvolution approach in the general context of the M-estimation that is designed to estimate the true parameter $\theta_0 = \operatorname{argmin}_{\theta\in\Theta} \mathbb{E}\ell(X_i, \theta)$ based on the noisy data $\{\tilde{X}_i\}_{i=1}^n$, where $\tilde{X}_i = X_i + \tilde{Z}_i$ and $\{\tilde{Z}_i\}_{i=1}^n$ are i.i.d. noises. Let $\varphi_{\ell,\theta}(t) = \int \ell(x, \theta)e^{it^\top x}\,\mathrm{d}x$ be the Fourier transform of the loss $\ell(x, \theta)$, and $\varphi_{\tilde{Z}}(t)$ be the characteristic function of $\tilde{Z}_i$. Suppose that, for every $\theta \in \Theta$, $x \mapsto \ell(x, \theta)$ is continuous and integrable, $t \mapsto \varphi_{\ell,\theta}(t)$ and $t \mapsto \varphi_{\ell,\theta}(t)/\varphi_{\tilde{Z}}(-t)$ is integrable, and $\mathbb{E}|\ell(X, \theta)| < \infty$. Then by applying Fourier inversion and Fubini theorem,

$$
\begin{aligned}
\mathbb{E}\ell(X, \theta) &= \int \ell(x, \theta)\,\mathrm{d}F_X(x) \\
&= \int_x \frac{1}{(2\pi)^d} \int_t e^{-it^\top x}\varphi_{\ell,\theta}(t)\,\mathrm{d}t\,\mathrm{d}F_X(x) \quad \text{(by Fourier inversion Lemma S3.1)} \\
&= \frac{1}{(2\pi)^d} \int_t \frac{\varphi_{\tilde{X}}(-t)}{\varphi_{\tilde{Z}}(-t)}\varphi_{\ell,\theta}(t)\,\mathrm{d}t \quad \text{(by Fubini and } \varphi_X(-t) = \varphi_{\tilde{X}}(-t)/\varphi_{\tilde{Z}}(-t)\text{)} \\
&= \frac{1}{(2\pi)^d} \int_t \frac{\int_x e^{-it^\top x}\,\mathrm{d}F_{\tilde{X}}(x)}{\varphi_{\tilde{Z}}(-t)}\varphi_{\ell,\theta}(t)\,\mathrm{d}t \\
&= \int\int \frac{1}{(2\pi)^d} e^{-it^\top x}\frac{\varphi_{\ell,\theta}(t)}{\varphi_{\tilde{Z}}(-t)}\,\mathrm{d}t\,\mathrm{d}F_{\tilde{X}}(x) \quad \text{(by Fubini's theorem),}
\end{aligned}
$$

which implies a corrected loss function

$$
\tilde{\ell}(x, \theta) = \frac{1}{(2\pi)^d} \int e^{-it^\top x}\frac{\varphi_{\ell,\theta}(t)}{\varphi_{\tilde{Z}}(-t)}\,\mathrm{d}t. \tag{4.4}
$$

If the integral exists, indicating that the underlying risk can be recovered by the corrected risk. However, the existence of the integral $\int e^{-it^\top x}\varphi_{\ell,\theta}(t)/\varphi_{\tilde{Z}}(-t)\,\mathrm{d}t$ requires restrictive conditions, which may not be satisfied for many loss functions and error distributions, for example, the L_2 loss with the Gaussian error. A remedy for this problem is to truncate the integral in (4.4) or multiply a rapidly decaying characteristic function in the numerator to counteract the divergence of $1/\varphi_{\tilde{Z}}(t)$ as $t \to \infty$.

Another approach based on (4.4) to construct the corrected loss function is based on differentiation of the loss $\ell(x, \theta)$, which requires sufficient smoothness of $x \mapsto \ell(x, \theta)$. Suppose the components of $\tilde{Z}_i$ are independent Laplace random variables, where $\tilde{Z}_i = (\tilde{Z}_{i1}, \ldots, \tilde{Z}_{id})^\top$ and $\tilde{Z}_{ij}$ follows the Laplace distribution with variance λ_j^2 for $j = 1, \ldots, d$. Applying the derivative theorem in Fourier transformation to (4.4), the corrected loss under componentwise independent Laplace noise is

$$
\tilde{\ell}^{\mathrm{L}}(x, \theta) = \left(1 - \frac{\lambda_1^2}{2}\frac{\partial^2}{\partial x_1^2}\right)\ldots\left(1 - \frac{\lambda_d^2}{2}\frac{\partial^2}{\partial x_d^2}\right)\ell(x, \theta).
$$

If $\tilde{Z}_i$ follows the SL distribution $\mathcal{SL}_d(\lambda^2 I_d)$, Lemma S4.3 in the SM provides the corresponding corrected loss $\tilde{\ell}^{\mathrm{SL}}(x,\theta) = \ell(x,\theta) - \frac{\lambda^2}{2}\sum_{k=1}^{d}\frac{\partial^2}{\partial x_k^2}\ell(x,\theta)$, which only depends on the second-order derivatives. Although the above forms of corrected functions avoid numerical integration, they require the loss function to be sufficiently smooth with respect to x, which is not satisfied by the quantile regression or the ReLU activation function. In contrast, the proposed method does not have such restrictions.

In comparison, the proposed DR corrected loss only requires mild conditions on the loss function $\ell(x,\theta)$ and is free from any condition on the Fourier transform $\varphi_{\ell,\theta}(t)$ in the frequency domain. Additionally, it does not require numerical integration or selection of hyperparameters for evaluating the corrected loss in (4.4), as in the case of the deconvoluted kernel density estimation where a smoothing parameter is needed. These make the proposed DR corrected loss more generally applicable and computationally more efficient.

5. Properties of doubly random corrected M-estimation. The study on the consistency and the asymptotic normality of the DR estimator requires the following conditions. Let $\mu^{(1)}$ and $\mu^{(2)}$ be two induced measures on $\mathbb{R}^d$ by $\tilde{X}_i^{(1)}$ and $\tilde{X}_i^{(2)}$, respectively.

CONDITION 1. *(i) For $\theta \in \Theta$, $\mathbb{E}|\ell(X_i,\theta)| < \infty$, $\mathbb{E}|\ell(\tilde{X}_i^{(1)},\theta)| < \infty$ and $\mathbb{E}|\ell(\tilde{X}_i^{(2)},\theta)| < \infty$. (ii) For $\theta \in \Theta$, $x \mapsto \ell(x,\theta)$ has a set of discontinuities denoted by D_θ. Assume that, for every $\theta \in \Theta$, $\mathbb{P}(X \in D_\theta) = 0$, and for any bounded set $B \in \mathbb{R}^d$, the intersection $D_\theta \cap B$ is finite. (iii) Assume that, for every $\theta \in \Theta$ and $M > 0$, $\sup_{x:\|x\|_2 \le M} |\ell(x,\theta)| < \infty$.*

CONDITION 2. *(i) The parameter space Θ is a compact set, and the true parameter θ_0 is an interior point of Θ. For any x, $\theta \to \ell(x,\theta)$ is continuous. (ii) Uniform law of large numbers: $\sup_{\theta\in\Theta} |n^{-1}\sum_{i=1}^n \ell(\tilde{X}_i^{(2)},\theta) - \mathbb{E}\ell(\tilde{X}_i^{(2)},\theta)| \xrightarrow{p} 0$ and $\sup_{\theta\in\Theta} |n^{-1}\sum_{i=1}^n \ell(\tilde{X}_i^{(1)},\theta) - \mathbb{E}\ell(\tilde{X}_i^{(1)},\theta)| \xrightarrow{p} 0$ as $n \to \infty$. (iii) Separability: for every $\varepsilon > 0$, $\inf_{\theta:d(\theta,\theta_0)\ge\varepsilon} \mathbb{E}\ell(X_i,\theta) > \mathbb{E}\ell(X_i,\theta_0)$.*

CONDITION 3. *(i) Assume $\ell(x,\theta)$ is differentiable at θ_0 with derivative $\nabla_\theta \ell(x,\theta_0)$ almost surely for $\mu^{(1)}$ and $\mu^{(2)}$, and $|\ell(x,\theta_1) - \ell(x,\theta_2)| \le \dot{\ell}(x)\|\theta_1 - \theta_2\|_2$ for every θ_1 and θ_2 in a neighborhood of θ_0 and a measurable function $\dot{\ell}$ with $\mathbb{E}\dot{\ell}(\tilde{X}_i^{(2)})^2 < \infty$ and $\mathbb{E}\dot{\ell}(\tilde{X}_i^{(1)})^2 < \infty$. (ii) Assume that the map $\theta \mapsto \mathbb{E}\ell(X_i,\theta)$ admits a second-order Taylor expansion at θ_0 with nonsingular symmetric second derivative matrix V_{θ_0}.*

Condition 1 is used to establish the unbiasedness of $\ell^{\mathrm{DR}}(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}, \theta)$. The integrability assumption in Condition 1(i) is fairly mild. Since both the ZIL and SL noises have finite second moments and the original observation X_i is bounded, this assumption is satisfied for the L_2 and check losses. Condition 1(ii) allows discontinuity points of $x \mapsto \ell(x,\theta)$, which is much weaker than the sufficient smoothness condition of $x \mapsto \ell(x,\theta)$ required by the deconvolution approach for loss adjustment. The loss functions in the contexts of linear regression, logistic regression, quantile regression, and neural networks all satisfy Condition 1(ii). To establish the asymptotic normality of the DR estimator, we require Conditions 2 and 3, which are standard assumptions for the M-estimation, as outlined in van der Vaart (1998). For example, the τ-quantile estimation task $\theta_0 = \arg\min_{\theta\in\Theta} \mathbb{E}(X_i - \theta)(\tau - \mathbf{1}(X_i - \theta < 0))$, along with the output of Algorithm 1, satisfies those three conditions provided that Θ is tight, $\mathbb{E}|X_i| < \infty$, $F'(\theta_0) > 0$ and $F''(\theta)$ exists in a neighborhood of θ_0 with $F''(\theta_0) \neq 0$. The example involving the ReLU function $\theta_0 = \arg\min_{\theta\in\Theta} \mathbb{E}(\theta - \mathrm{ReLU}(X_i))^2$ considered in the simulation study also satisfies the three conditions provided that Θ is tight and $\mathbb{E}X_i^2 < \infty$.

THEOREM 5.1. *Under Condition 1 with the noisy data* $\{\tilde{X}_i^{(1)}\}_{i=1}^n$ *and* $\{\tilde{X}_i^{(2)}\}_{i=1}^n$ *obtained in Algorithm 1, and the privacy parameter* δ*, it follows that, for every* $\theta \in \Theta$*,*

$$\mathbb{E}\ell(X_i, \theta) = \mathbb{E}\ell^{\mathrm{DR}}\big(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}, \theta; \delta\big).$$

The proof is deferred to Section S3.2 in the SM. Theorem 5.1 suggests that ℓ^{DR} acts as a surrogate for the underlying loss ℓ, which makes the differentially private DR estimator (4.1) consistent to the underlying θ. The reuse of the original data in constructing $\{\tilde{X}_i^{(1)}\}_{i=1}^n$ and $\{\tilde{X}_i^{(2)}\}_{i=1}^n$ does not introduce any bias because, conditional on the data, the dependence structure between them is properly accounted for. This is achieved as the dependence structure is completely determined by the independent ZIL and SL noise terms, whose distributions are known. Compared to the existing approaches to obtain a corrected loss in Section 4.2, which require the existence of higher-order derivatives of $x \mapsto \ell(x, \theta)$ or necessitate numerical integration, the DR corrected loss only requires that the $x \mapsto \ell(x, \theta)$ is continuous. In addition, the DR corrected loss only uses basic arithmetic operations when calculating the loss ℓ, making it computationally very efficient.

We say a differential privacy mechanism is versatile if it is: (i) applicable to general estimation tasks, (ii) applicable to any number of analyses, and (iii) adaptive to increasing data volume in online scenarios. First, as we prove in the Theorem 5.1, the DR estimator based on the ZIL mechanism can handle a wide range of loss functions, including nonsmooth ones. Second, once the ZIL mechanism is applied to a dataset $\mathbf{X}$, the output is fixed. That is, no matter how many analysts invoke the mechanism, the answer remains the same. Thus, no privacy degradation is incurred during these invocations. Third, in the online scenario, when new data come in, the mechanism only needs to publish the perturbed new data without changing the output corresponding to the old data. Consequently, the ZIL mechanism can adapt to increasing data volumes in online scenarios. Therefore, the ZIL mechanism is versatile.

The DR estimator is differentially private, as it is based on the output of the doubly random DP mechanism in Algorithm 1. According to Theorem 3.1 and the postprocessing property (Proposition 2.2), $\hat{\theta}_n^{\mathrm{DR}}$ satisfies $T_{d,c_A,\delta}$-LADP and $T_{d,c_I,\delta}$-LDP for $c_A = \max_{1\le j\le d} \operatorname{diam}(\mathcal{X}_j)/\lambda$ and $c_I = \operatorname{diam}(\mathcal{X})/\lambda$.

The following theorems present the consistency and the asymptotic normality of $\hat{\theta}_n^{\mathrm{DR}}$.

THEOREM 5.2. *Under Conditions 1 and 2,* $\hat{\theta}_n^{\mathrm{DR}} \xrightarrow{p} \theta_0$ *as* $n \to \infty$.

THEOREM 5.3. *Under Conditions 1, 2, and 3,* $\sqrt{n}(\hat{\theta}_n^{\mathrm{DR}} - \theta_0) \to^d N(0, V_{\theta_0}^{-1} A(\delta, \lambda) V_{\theta_0}^{-1})$*, as* $n \to \infty$ *where* $A(\delta, \lambda) = \mathbb{E}\{\nabla_\theta \ell^{\mathrm{DR}}(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}, \theta_0; \delta)\}\{\nabla_\theta \ell^{\mathrm{DR}}(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}, \theta_0; \delta)\}^\top$.

The proofs of both theorems are deferred to Section S3.3 in the SM. Recall that the M-estimator based on the original data, namely, the oracle estimator $\hat{\theta}_n^{\mathrm{ORA}}$, has the asymptotic variance $V_{\theta_0}^{-1} A^{\mathrm{ORA}} V_{\theta_0}^{-1}$ for $A^{\mathrm{ORA}} = \mathbb{E}\nabla_\theta \ell(X_i, \theta_0) \nabla_\theta \ell(X_i, \theta_0)^\top$. The difference in the asymptotic variances between the Oracle and the DR estimators lies in the difference between the middle terms of the sandwich form, namely, between A^{ORA} and $A(\delta, \lambda)$. The following corollary shows that under mild conditions $A(\delta, \lambda) \succeq A^{\mathrm{ORA}}$, indicating a cost in the estimation efficiency due to privacy protection.

COROLLARY 5.1. *Let* D_{θ_0} *be the set of discontinuities of* $x \mapsto \nabla_\theta \ell(x, \theta_0)$*. Assume that* $\mathbb{P}(X \in D_{\theta_0}) = 0$ *and for any bounded set* $B \subset \mathbb{R}^d$*, the cardinality of the intersection* $D_{\theta_0} \cap B$ *is finite. Additionally, suppose that, for every* $M > 0$, $\sup_{\|x\|_2 \le M} \|\nabla \ell(x, \theta_0)\|_2 < \infty$*. Then* $A(\delta, \lambda) \succeq A^{\mathrm{ORA}}$.

The proof of this corollary is deferred to Section S3.4 in the SM. It can be shown that, if $\delta = 1$ or $\lambda = 0$, $A(\delta, \lambda) = A^{\mathrm{ORA}}$, where in both cases, the dataset $\{\tilde{X}_i^{(1)}\}_{i=1}^n$ degenerates to the true dataset $\{X_i\}_{i=1}^n$. In the following we examine how the privacy parameters δ and λ influence $A(\delta, \lambda)$. Let $\nabla_{\theta x}\ell(X_1, \theta_0)$ denote the $p \times d$ matrix with the (j_1, j_2) entry given by $\partial^2\ell(X_1, \theta_0)/\partial\theta_{j_1}\partial x_{j_2}$, and $V_1 = \mathbb{E}\nabla_{\theta x}\ell(X_1, \theta_0)\nabla_{\theta x}\ell(X_1, \theta_0)^\top$. Furthermore, we define $V_2^\lambda = \mathbb{E}(\nabla_\theta\ell(X_1, \theta_0) - \nabla_\theta\ell(X_1 + S, \theta_0))(\nabla_\theta\ell(X_1, \theta_0) - \nabla_\theta\ell(X_1 + S, \theta_0))^\top$ and $V_3^\lambda = \mathbb{E}\nabla_{\theta x}\ell(X_1 + S, \theta_0)\nabla_{\theta x}\ell(X_1 + S, \theta_0)^\top$, where $S \sim \mathcal{SL}_d(\lambda^2 I_d)$ is independent of X_1. For a matrix A, let $\mathrm{Tr}(A)$ denote its trace.

PROPOSITION 5.1. *Assume that, for every $\theta \in \Theta$, $x \mapsto \ell(x, \theta)$ has a common compact support $\mathcal{X}$, and for each $i = 1, \ldots, p$, $x \mapsto \partial\ell(x, \theta_0)/\partial\theta_i$ is three times continuously differentiable. Then*: (*i*) *for $\delta > 0$ being fixed, $A(\delta, \lambda) = A^{\mathrm{ORA}} + (1/\delta - 1)V_1\lambda^2 + o(\lambda^2)$ as $\lambda \to 0$, and $A(\delta, \lambda) \to A^{\mathrm{ORA}}/\delta$ as $\lambda \to \infty$.*

(*ii*) *For $\lambda > 0$ being fixed, $A(\delta, \lambda) = A^{\mathrm{ORA}} + V_2^\lambda(1-\delta) + o(1-\delta)$ as $\delta \to 1$, and $A(\delta, \lambda) = \lambda^2 V_3^\lambda/\delta + o(1/\delta)$ as $\delta \to 0$. As $\lambda \to 0$, $V_2^\lambda \to 0$ and $V_3^\lambda \to V_1$. As $\lambda \to \infty$, $V_2^\lambda \to V_1$, and $\lambda^2\mathrm{Tr}(V_3^\lambda) = O(1)$ when $d \geq 3$, $O(\log\lambda)$ when $d = 2$, and $O(\lambda)$ when $d = 1$.*

The proof of this proposition is deferred to Section S3.5 in the SM. Proposition 5.1(i) demonstrates that the cost of the privacy $A(\delta, \lambda) - A^{\mathrm{ORA}}$ is of order λ^2 as $\lambda \to 0$, with the coefficient of the quadratic term being larger for smaller δ. The reason for the quadratic order rather than the linear order in λ is that the expectation of the score $\nabla_\theta\ell(X, \theta_0)$ is zero, making the linear term vanish. Proposition 5.1(i) also shows that, as $\lambda \to \infty$, $A(\delta, \lambda)$ tends to A^{ORA}/δ, indicating a privacy cost of $A^{\mathrm{ORA}}(1/\delta - 1)$ for the DR estimator, which nevertheless remains bounded. Proposition 5.1(ii) shows that, for λ being fixed, in contrast to the quadratic difference in the asymptotic variances as $\lambda \to 0$, the cost of privacy $A(\delta, \lambda) - A^{\mathrm{ORA}}$ is of order $1 - \delta$ as $\delta \to 1$. As $\delta \to 0$ for fixed λ, the leading asymptotic variance is $\lambda^2 V_3^\lambda/\delta$, which indicates substantially high privacy cost by making δ small. For δ being very small, the alternative SL and SDR estimators proposed in the next section could be used, as they have bounded asymptotic variances with respect to $\delta \in [0, 1]$.

The assumption in Proposition 5.1 that $x \mapsto \ell(x, \theta)$ has a common compact support can be justified as follows. Since the data distribution F is assumed to have a compact support, the function $x \mapsto \ell(x, \theta)$ can be modified to have compact support without affecting the smoothness of $\ell(x, \theta)$ and the value of $\mathbb{E}[\ell(X, \theta)]$. Specifically, let $\eta(t) = \exp(-1/t)\mathbf{1}(t > 0)$ and $\Psi(t) = \eta(t)/(\eta(t) + \eta(1 - t))$. It can be verified that Ψ is infinitely differentiable on $\mathbb{R}$, monotonic on $(0, 1)$, and satisfies $\Psi(t) = 0$ for $t \leq 0$, and $\Psi(t) = 1$ for $t \geq 1$. Suppose that the compact support of F is contained within a ball $B(x_0, M)$ centered at a x_0 with radius M. Then we can modify the loss function to $\overline{\ell}(x, \theta) = \ell(x, \theta) \cdot \Psi(M^2 - \|x - x_0\|_2^2 + 1)$, which coincides with $\ell(x, \theta)$ for $\|x - x_0\|_2 \leq M$, and equals to 0 for $\|x - x_0\|_2 > \sqrt{M^2 + 1}$.

It is noted that the Noisy-GD estimator (Cai, Wang and Zhang (2021), Avella-Medina, Bradshaw and Loh (2023)) has no efficiency loss relative to the corresponding oracle estimator. However, the Noisy-GD satisfies only the central DP, as it needs to compute the average gradient over the entire data. In contrast, the ZIL mechanism satisfies the Local DP. As will be noted in the following, the Local DP generally induces larger efficiency loss than estimators of the Central DP mechanisms.

Duchi and Ruan (2024)'s analysis implies that, under the ε-LDP ($f_{\varepsilon,0}$-LDP) constraint, the noisy-SGD estimator for the M-estimation encounters a nonnegligible efficiency loss in a way that the asymptotic variance scales as $1/\varepsilon^2$. For one-dimensional parametric models under ε-LDP, Steinberger (2024) derived the optimal asymptotic variance among all regular estimators and designed a "two-stage" estimator that attains this optimal asymptotic variance, which scales as $1/\varepsilon^2$ under the Bernoulli model. In contrast, although Corollary 5.1

shows that the DR estimator incurs some efficiency loss, its asymptotic variance remains bounded with respect to ε for any fixed $\delta > 0$ under the $T_{d,\varepsilon,\delta}$-LDP constraint by setting $\varepsilon = \operatorname{diam}(\mathcal{X})/\lambda$. Since the asymptotic variance under the standard ε-LDP typically (Duchi and Ruan (2024)) grows at a rate of $1/\varepsilon^2$, this demonstrates that the efficiency loss of the DR estimator is substantially reduced by leveraging a weaker DP constraint. This improvement originates from the positive value of δ, as it makes the $T_{d,\varepsilon,\delta}$-LDP less stringent than the ε-LDP.

6. Smoothed doubly random estimation. In addition to the DR corrected loss, which does not require the loss function to be smooth, we propose a differentially private M-estimation if the loss is known to be smooth up to the second order.

Since $\{S_i\}_{i=1}^n$ and $\{Z_i + S_i\}_{i=1}^n$ are both SL distributed but with different variances, applying Lemma S4.3 twice and recalling that $\tilde{X}_i^{(1)} = \tilde{X}_i + Z_i$ and $\tilde{X}_i^{(2)} = \tilde{X}_i^{(1)} + S_i$, we have

$$\mathbb{E}\ell(X_i, \theta) = \mathbb{E}\left\{\ell\big(\tilde{X}_i^{(2)}, \theta\big) - \frac{\lambda^2}{2}\sum_{k=1}^{d}\frac{\partial^2}{\partial x_k^2}\ell\big(\tilde{X}_i^{(2)}, \theta\big)\right\} \quad \text{and} \tag{6.1}$$

$$\mathbb{E}\ell\big(\tilde{X}_i^{(1)}, \theta\big) = \mathbb{E}\left\{\ell\big(\tilde{X}_i^{(2)}, \theta\big) - \frac{\delta\lambda^2}{2}\sum_{k=1}^{d}\frac{\partial^2}{\partial x_k^2}\ell\big(\tilde{X}_i^{(2)}, \theta\big)\right\}, \tag{6.2}$$

if $x \mapsto \ell(x, \theta)$ is twice differentiable. Substituting (6.2) to (6.1), we obtain a smoothed doubly random corrected loss (SDR corrected loss) that recovers the underlying expected loss,

$$\ell^{\mathrm{SDR}}\big(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}, \theta; \delta, \lambda\big) = \ell\big(\tilde{X}_i^{(1)}, \theta\big) - \frac{(1-\delta)\lambda^2}{2}\sum_{k=1}^{d}\frac{\partial^2}{\partial x_k^2}\ell\big(\tilde{X}_i^{(2)}, \theta\big).$$

This leads to the SDR estimator

$$\hat{\theta}_n^{\mathrm{SDR}} = \operatorname*{argmin}_{\theta\in\Theta}\sum_{i=1}^{n}\left\{\ell\big(\tilde{X}_i^{(1)}, \theta\big) - \frac{(1-\delta)\lambda^2}{2}\sum_{k=1}^{d}\frac{\partial^2}{\partial x_k^2}\ell\big(\tilde{X}_i^{(2)}, \theta\big)\right\}.$$

The SDR corrected loss may be seen as a version of the DR corrected loss for situations where the loss is smooth to the second order.

From (6.1), one may have another unbiased corrected loss specialized for SL noise

$$\ell^{\mathrm{SL}}\big(\tilde{X}_i^{(2)}, \theta; \lambda\big) = \ell\big(\tilde{X}_i^{(2)}, \theta\big) - \frac{\lambda^2}{2}\sum_{k=1}^{d}\frac{\partial^2}{\partial x_k^2}\ell\big(\tilde{X}_i^{(2)}, \theta\big)$$

by using only one dataset $\{\tilde{X}_i^{(2)}\}_{i=1}^n$, which leads to the SL estimator

$$\hat{\theta}_n^{\mathrm{SL}} = \operatorname*{argmin}_{\theta\in\Theta}\sum_{i=1}^{n}\left\{\ell\big(\tilde{X}_i^{(2)}, \theta\big) - \frac{\lambda^2}{2}\sum_{k=1}^{d}\frac{\partial^2}{\partial x_k^2}\ell\big(\tilde{X}_i^{(2)}, \theta\big)\right\}.$$

To establish the asymptotic normality of $\hat{\theta}_n^{\mathrm{SL}}$ and $\hat{\theta}_n^{\mathrm{SDR}}$, we assume the following.

CONDITION 4. *The tuple* (ℓ, X, Θ) *satisfies the following*: (*i*) *For every* $\theta \in \Theta$, $\mathbb{E}|\ell(X, \theta)| < \infty$. (*ii*) *For every* $\theta \in \Theta$ *and* $k = 1, \ldots, d$, $x \mapsto \frac{\partial^2\ell(x,\theta)}{\partial x_k^2}$ *is continuous*. (*iii*) *For every* $\theta \in \Theta$, *there are constants* c, C *and* $a \in [0, 1)$ *such that* $|\ell(x, \theta)| \le Ce^{c\|x\|_2^a}$ *and* $|\frac{\partial^2\ell(x,\theta)}{\partial x_k^2}| \le Ce^{c\|x\|_2^a}$ *for every* $x \in \mathbb{R}^d$.

CONDITION 5. *The tuple* $(\ell, X, \Theta, \theta)$ *satisfies the following*: (*i*) *Let* $\{X_i\}_{i=1}^n \overset{i.i.d.}{\sim} X$. *Then* $\sup_{\theta\in\Theta} |\frac{1}{n}\sum_{i=1}^n \ell(X_i,\theta) - \mathbb{E}\ell(X,\theta)| \xrightarrow{p} 0$, *as* $n \to \infty$. (*ii*) *Let* μ_X *be the induced measure by* X. *Assume* $\ell(x,\theta)$ *is differentiable at* θ_0 *with derivative* $\nabla_\theta \ell(x,\theta_0)$ *almost surely for* μ_X, *and* $|\ell(x,\theta_1) - \ell(x,\theta_2)| \le \dot{\ell}(x)\|\theta_1 - \theta_2\|_2$ *for every* θ_1 *and* θ_2 *in a neighborhood of* θ_0 *and a measurable function* $\dot{\ell}$ *with* $\mathbb{E}\dot{\ell}(X)^2 < \infty$.

Condition 4 is to guarantee the unbiasedness of the ℓ^{SL} and ℓ^{SDR}. Condition 5 collects the standard conditions in the M-estimation for consistency and asymptotic normality.

THEOREM 6.1. *Suppose Condition* 2(*i*), 2(*iii*) *and* 3(*ii*) *hold. Suppose that the tuple* (ℓ, X_1, Θ) *satisfies Condition* 4: (*i*) *If the tuple* $(\ell^{\mathrm{SDR}}, (\tilde{X}_1^{(1)}, \tilde{X}_1^{(2)}), \Theta, \theta_0)$ *satisfies Condition* 5, *then*

$$\sqrt{n}(\hat{\theta}_n^{\mathrm{SDR}} - \theta_0) \xrightarrow{d} N\big(0, V_{\theta_0}^{-1} A^{\mathrm{SDR}}(\delta,\lambda) V_{\theta_0}^{-1}\big),$$

where $A^{\mathrm{SDR}}(\delta,\lambda) = \mathbb{E}\{\nabla_\theta \ell^{\mathrm{SDR}}(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}, \theta_0; \delta, \lambda)\}\{\nabla_\theta \ell^{\mathrm{SDR}}(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}, \theta_0; \delta, \lambda)\}^\top$.

(*ii*) *If the tuple* $(\ell^{\mathrm{SL}}, \tilde{X}_1^{(2)}, \Theta, \theta_0)$ *satisfies Condition* 5, *then*

$$\sqrt{n}(\hat{\theta}_n^{\mathrm{SL}} - \theta_0) \xrightarrow{d} N\big(0, V_{\theta_0}^{-1} A^{\mathrm{SL}}(\lambda) V_{\theta_0}^{-1}\big),$$

where $A^{\mathrm{SL}}(\lambda) = \mathbb{E}\{\nabla_\theta \ell^{\mathrm{SL}}(\tilde{X}_i^{(2)}, \theta_0; \lambda)\}\{\nabla_\theta \ell^{\mathrm{SL}}(\tilde{X}_i^{(2)}, \theta_0; \lambda)\}^\top$.

The proof is deferred to Section S4.2 in the SM. The following proposition analyzes how the privacy parameter λ influences $A^{\mathrm{SL}}(\lambda)$, and how λ and δ influence $A^{\mathrm{SDR}}(\delta,\lambda)$. Let μ be the Lebesgue measure on $\mathbb{R}^d$. Let $\Delta_x = \sum_{k=1}^d \frac{\partial^2}{\partial x_k^2}$ denote the Laplace operator. For a matrix A, let $A_{i,j}$ denote its (i,j) element. For $i \in \{1,\dots,p\}$, let $\ell_i(x) = \partial\ell(x,\theta_0)/\partial\theta_i$.

PROPOSITION 6.1. *Under the conditions of Proposition* 5.1, *the following statements hold*: (*i*) *For any fixed* $\delta \in [0,1]$, $A^{\mathrm{SL}}(\lambda) \to A^{\mathrm{ORA}}$ *and* $A^{\mathrm{SDR}}(\delta,\lambda) \to A^{\mathrm{ORA}}$ *as* $\lambda \to 0$; *for any fixed* $\lambda \ge 0$, $A^{\mathrm{SDR}}(\delta,\lambda) \to A^{\mathrm{ORA}}$ *as* $\delta \to 1$.

(*ii*) *As* $\lambda \to \infty$, $\mathrm{Tr}(A^{\mathrm{SL}}(\lambda)) = O(r_d(\lambda))$ *and* $\sup_{\delta\in[0,1]} \mathrm{Tr}(A^{\mathrm{SDR}}(\delta,\lambda)) = O(r_d(\lambda))$, *where* $r_d(\lambda) = \lambda^2$ *for* $d \ge 3$, $r_d(\lambda) = \lambda^2 \log\lambda$ *for* $d = 2$, *and* $r_d(\lambda) = \lambda^3$ *for* $d = 1$.

(*iii*) *For* $i \in \{1,\dots,p\}$, *if* $\mu(\{x \in \mathcal{X} : [\Delta_x \ell_i(x)]^2 > 0\}) > 0$, *then for any fixed* $\delta \in [0,1]$, $\frac{A^{\mathrm{SDR}}(\delta,\lambda)_{i,i}}{A^{\mathrm{SL}}(\lambda)_{i,i}} \to (1-\delta)^2$ *as* $\lambda \to \infty$.

(*iv*) *For any fixed* $\lambda \ge 0$, $A^{\mathrm{SDR}}(\delta,\lambda) \to A^{\mathrm{SL}}(\lambda)$ *as* $\delta \to 0$.

The proof is deferred to Section S4.3 in the SM. Proposition 6.1(i) shows that the middle matrices $A^{\mathrm{SDR}}(\delta,\lambda)$ and $A^{\mathrm{SL}}(\lambda)$ of the asymptotic variance of the SDR and the SL estimators converge to their Oracle counterpart A^{ORA} for either $\lambda \to 0$ or $\delta \to 1$. It is noted that $\lambda \to 0$ or $\delta \to 1$ corresponds to the case of reduced privacy protection. As noted before, if $\delta = 0$ or $\lambda = 0$, the original data are retained, so the SDR or the SL estimators are equivalent to the oracle M-estimator. The results in Part(ii) depend on dimension d. This is because the density of $\mathcal{SL}_d(I_d)$ has different forms at the origin for $d = 1$, $d = 2$ and $d \ge 3$. Consequently, the one-dimensional and two-dimensional cases require separate treatment in Part(ii). Proposition 6.1(iii) shows that, for any fixed $\delta \in [0,1]$, as $\lambda \to \infty$, the diagonal elements of $A^{\mathrm{SDR}}(\delta,\lambda)$ are approximately $(1-\delta)^2$ times those of $A^{\mathrm{SL}}(\lambda)$, revealing a benefits of the SDR method.

Proposition 6.1 (iv) shows that, as $\delta \to 0$, the asymptotic variance of the SDR estimator converges to that of the SL estimator. Consequently, the efficiency of the SDR estimator does

not degenerate significantly under $\delta \to 0$, preventing the kind of efficiency degradation for the DR estimator reported in Proposition 5.1(iii). This represents a benefit for the smoothed doubly random private estimators.

An important question is on the choice of the privacy level on the estimation efficiency for the DR, SDR and SL estimators. When strong privacy is required, one can either make λ large or δ small. If λ is large, the DR estimator is preferred because it has bounded asymptotic variance (A^{ORA}/δ), unlike the SDR and SL estimators whose asymptotic variances diverge at least at the rate of λ^2. If δ is small, the SL and SDR estimators are preferred, as they yield smaller asymptotic variances, provided the loss function is twice continuously differentiable in x. When the privacy requirements are not high, the asymptotic variances of all three estimators are moderate, but the DR estimator is the most convenient to use, since constructing the DR-corrected loss only involves basic arithmetic operations.

Steinberger (2024) shows that for one-dimensional parametric models, the optimal asymptotic variance bound $(\sup_{Q\in\mathcal{Q}_\varepsilon} I_\theta(QP))^{-1}$ cannot be attained by a fixed mechanism Q_0 uniformly for all $\theta \in \Theta$. This indicates an unavoidable efficiency loss when striving for versatility, even within a given statistical task. The mechanism proposed in Steinberger (2024) aims to attain the optimal asymptotic variance while sacrificing versatility. Regarding the asymptotic variances of the SL and SDR estimators, we may conduct a comparison under the Bernoulli model $\{X_i\}_{i=1}^n \overset{i.i.d.}{\sim} \mathrm{Bernoulli}(\theta)$. Under ε-LDP, the optimal asymptotic variance given in Steinberger (2024) is $(\sup_{Q\in\mathcal{Q}_\varepsilon} I_\theta(QP))^{-1} = \frac{e^\varepsilon}{(e^\varepsilon-1)^2} + \theta(1-\theta)$. Let $\{Z_i\}_{i=1}^n \overset{i.i.d.}{\sim} \mathrm{ZIL}(\delta, 2/\varepsilon^2)$ and $\{S_i\}_{i=1}^n \overset{i.i.d.}{\sim} \mathcal{SL}_d(\delta\cdot 2/\varepsilon^2)$. Note that $Z_i + S_i \sim L(1/\varepsilon)$. In this case, without truncation of the loss function $(x-\theta)^2$, the SL estimator takes the form $\hat\theta_n^{\mathrm{SL}} = \frac{1}{n}\sum_{i=1}^n (X_i + Z_i + S_i)$, attains the ε-LDP, and has asymptotic variance $\frac{2}{\varepsilon^2} + \theta(1-\theta)$. On the other hand, the SDR estimator takes the form $\hat\theta_n^{\mathrm{SDR}} = \frac{1}{n}\sum_{i=1}^n (X_i + Z_i)$, attains $f_{\varepsilon,\delta}$-LDP, and has asymptotic variance $(1-\delta)\frac{2}{\varepsilon^2} + \theta(1-\theta)$. Thus, although the asymptotic variances of the SL and SDR estimators have the same rate $1/\varepsilon^2$ as the optimal asymptotic variance as $\varepsilon \to 0$, their variances are strictly larger. For the SL estimator, its asymptotic variance is approximately twice that of the optimal one. For the SDR estimator, when δ is close to 0 (as is typically the case), its variance is approximately $2(1-\delta)$ times the optimal variance as $\varepsilon \to 0$.

7. Simulation. We report results from simulation experiments designed to confirm the theoretical findings of the proposed DR, SL, and SDR estimators for DP M-estimation in the earlier sections under a nonsmooth L_2 loss, logistic regression, and quantile regression.

7.1. *Nonsmooth loss.* To demonstrate the ability of the proposed DR estimator for M-estimation with nonsmooth loss functions, we considered three loss functions $\ell_1(x,\theta) = (\theta - \mathrm{ReLU}(x))^2$, $\ell_2(x,\theta) = (\theta - \mathbf{1}_{[0.5,1]}(x))^2$ and $\ell_3(x,\theta) = (\theta - |\sin 2\pi x|)^2$ which are nonsmooth with respect to x, where $\mathrm{ReLU}(x) = \max(0,x)$ denotes the ReLU function and $\mathbf{1}_{[0.5,1]}(x)$ denotes the indicator function on $[0.5, 1]$. The original data $\{X_i\}_{i=1}^n$ were generated from the uniform $U(0,1)$. Noisy data $\{\tilde X_i^{(1)}\}_{i=1}^n$ and doubly randomized data $\{\tilde X_i^{(2)}\}_{i=1}^n$ were generated using Algorithm 1 with two sets of privacy parameters: $(\delta,\lambda) = (0.1, 0.94)$ and $(0.05, 1.4)$. The true parameters $\theta_{0,j} = \mathrm{argmin}_\theta\, \mathbb{E}\ell_j(X_i,\theta)$ for $j = 1, 2$ and 3. As $X_i \sim U(0,1)$, $\theta_{0,1} = \mathbb{E}\,\mathrm{ReLU}(X_i) = 0.5$, $\theta_{0,2} = \mathbb{E}\mathbf{1}_{[0.5,1]}(X_i) = 0.5$, and $\theta_{0,3} = \mathbb{E}|\sin 2\pi X_i| = 2/\pi$. The sample sizes were $n = 500$, 1000, 2000, 3000, 4000, and 5000.

Figure 2 displays the box plots of the estimation errors of the three estimators based on 5000 simulations with their root mean square errors (RMSEs) shown in Table 1. We only report the results for $n = 500$ and 1000, while results for other sample sizes are in Table S1

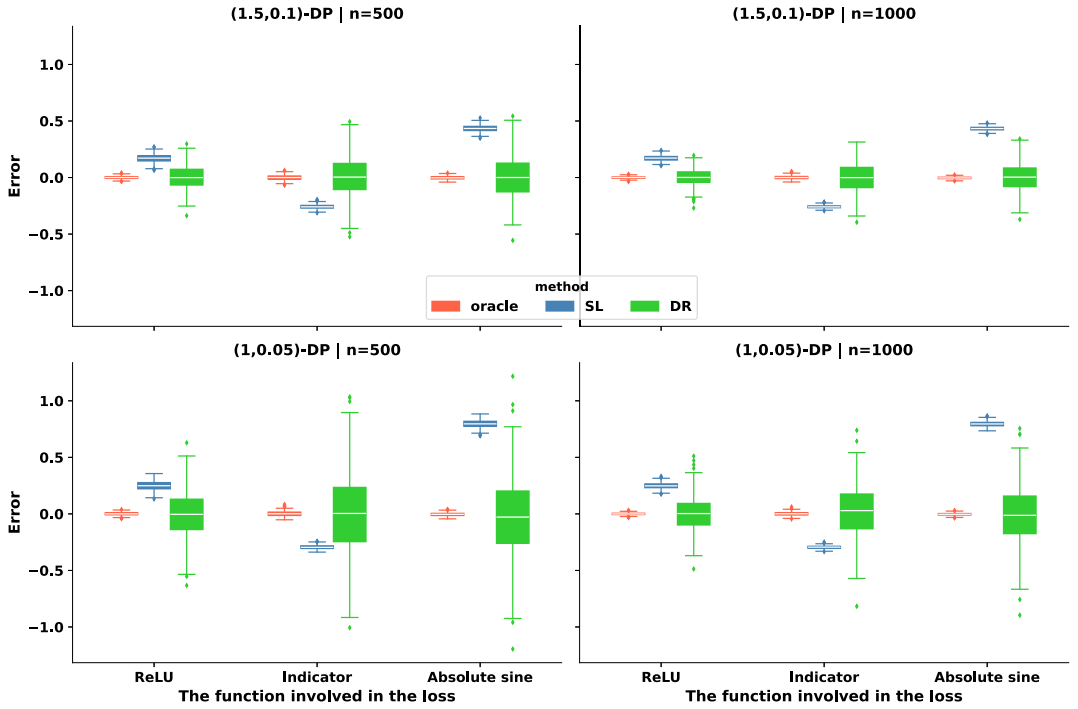

FIG. 2. *Box-plots of estimation errors of the oracle (left box), SL corrected-loss (middle box), and DR (right box) estimators for the ReLU, indicator and absolute sine functions, respectively. The first row corresponds to* $(\delta, \lambda) = (0.1, 0.94)$ *of the* $(1.5, 0.1)$*-DP, and the second row corresponds to* $(\delta, \lambda) = (0.05, 1.4)$ *of the* $(1, 0.05)$*-DP.*

of the SM. The figure and the tables show that the SL estimator was not a consistent estimator of the true parameter, as it is not suitable for nonsmooth loss. In contrast, the DR estimator was unbiased, and its RMSEs decreased with the increase of the sample size, which indicated its consistency and confirmed the finding in Theorem 5.1.

TABLE 1

The root mean square errors of the oracle, SL, and DR estimators for the parameters of the three nonsmooth loss functions based on 5000 *repetitions at* $(1.5, 0.1)$ *and* $(1, 0.05)$*-DP. The oracle estimator was calculated using the original noncorrupted data, which didn't protect data privacy*

n	DP	Method	ReLU	Indicator	Absolute sine
500	None	Oracle	0.012	0.023	0.015
	(1.5,0.1)-DP	SL	0.173	0.259	0.433
		DR	0.105	0.183	0.170
	None	Oracle	0.013	0.022	0.013
	(1,0.05)-DP	SL	0.252	0.294	0.796
		DR	0.184	0.326	0.358
1000	None	Oracle	0.009	0.016	0.011
	(1.5,0.1)-DP	SL	0.174	0.258	0.433
		DR	0.072	0.128	0.123
	None	Oracle	0.009	0.016	0.010
	(1,0.05)-DP	SL	0.250	0.294	0.796
		DR	0.131	0.230	0.257

Under the first DP parameter setting of $\delta = 0.1$ and $\lambda = 0.94$, both $\hat{\theta}_{n,j}^{\mathrm{DR}}$ and $\hat{\theta}_{n,j}^{\mathrm{SL}}$ achieved $(1.5, 0.1)$-DP. Under the second DP setting of $\delta = 0.05$ and $\lambda = 1.4$, they achieved $(1, 0.05)$-DP. The $(1, 0.05)$-DP provided higher privacy protection than the $(1.5, 0.1)$-DP as a larger λ or smaller δ indicated higher privacy protection. Note that the variance of the ZIL noise $\mathrm{ZIL}(\delta, \lambda^2 I_d)$ was $(1-\delta)\lambda^2 I_d$. In general, the added noises with higher variance resulted in stronger differential privacy guarantees. It is observed that, while the level of privacy protection increased, the performance of the SL and DR estimators deteriorates at a sample size. Meanwhile, the variance of the DR estimator was larger than that of the oracle estimator, representing a cost of privacy protection as indicated by Corollary 5.1.

7.2. *Logistic regression*. We considered the logistic model $\mathbb{P}(Y_i = 1|X_i, \beta) = \{1 + \exp(-X_i^\top \beta)\}^{-1}$, where $\beta = (\beta_1, \ldots, \beta_6)^\top$. The covariates $\{X_i\}_{i=1}^n$ were i.i.d. generated from a truncated multivariate normal distribution $\mathcal{N}(0_6, I_6)$ over a rectangle formed by a lower bound -1_6 and an upper bound 1_6, where 0_p and 1_p denote p-dimensional vectors of 0 and 1, respectively. The noisy covariates $\{\tilde{X}_i^{(1)}\}_{i=1}^n$ and the doubly randomized covariates $\{\tilde{X}_i^{(2)}\}_{i=1}^n$ were generated using Algorithm 1, with privacy parameters λ and δ set to $(0.5, 0.2)$ and $(1, 0.2)$, respectively. These settings provided $T_{6,4,0.2}$-ADP and $T_{6,2,0.2}$-ADP for the covariates, respectively. The true $\beta_* = 1_6$, and $n = 5000$, 7500 and 10,000.

We compared five M-estimators. The first one used the original data and minimized $\sum_{i=1}^n \{(1 - Y_i) \cdot X_i^\top \beta + \log(1 + \exp(-X_i^\top \beta))\}$ to obtain the oracle estimator $\hat{\beta}_n^{\mathrm{ORA}}$. The second one minimized the corrupted loss $\sum_{i=1}^n \{(1 - Y_i) \cdot \tilde{X}_i^{(1)\top} \beta + \log(1 + \exp(-\tilde{X}_i^{(1)\top} \beta))\}$ to obtain the naive estimator $\hat{\beta}_n^{\mathrm{NAI}}$. The others were the SL, DR, and SDR estimators.

The RMSEs of the estimated parameters based on 5000 replications are presented in Table 2. It shows that the naive estimator, which did not take any measure to counter the added noises, had the worst RMSEs in all cases, which was expected. For all three DP estimators (SL, SDR, and DR), the RMSE decreased with larger sample sizes but increased as the privacy level was strengthened from $T_{6,4,0.2}$-ADP to $T_{6,2,0.2}$-ADP. Notably, the RMSEs of the SL, SDR, and DR estimators were always higher than those of the oracle estimator, reflecting the cost of ensuring privacy protection. Meanwhile, the SDR estimator consistently yielded a smaller RMSE compared to the SL estimator, which is consistent with the result in Proposition 6.1(iii). As λ increases, the ratio of the RMSEs of SDR and SL to that of DR increases, which aligns with the second part of Proposition 5.1(i) and the result in Proposition 6.1(ii), since both results indicate that the RMSEs of the SL and SDR estimators divided by that of the DR estimator are expected to increase as λ increases.

7.3. *Quantile regression*. We considered the median regression $\mathrm{Median}(Y_i|X_i) = \beta_0 + X_i^\top \beta$ as in Pan et al. (2022), where $\beta = (\beta_1, \ldots, \beta_6)^\top$. The covariates $\{X_i = (X_{i1}, \ldots, X_{i6})^\top\}_{i=1}^n$ were i.i.d. generated from a multivariate truncated normal distribution $\mathcal{N}(0_6, I_6)$ with the lower and upper truncation bounds -1_6 and 1_6, respectively. Let $\{\varepsilon_i\}_{i=1}^n$ be i.i.d. $N(0, \sigma^2)$ with $\sigma = 1$ and the true parameters $\beta_{*,0} = 1$ and $\beta_* = 1_6$ so that $1 + \sum_{j=1}^6 X_{ij}$ is the conditional median of Y_i for $i = 1, \ldots, n$. The noisy covariates $\{\tilde{X}_i^{(1)}\}_{i=1}^n$ and the doubly randomized covariates $\{\tilde{X}_i^{(2)}\}_{i=1}^n$ were generated using Algorithm 1 with the privacy parameters λ and δ set to $(2, 0.2)$ and $(2.5, 0.2)$, respectively. These settings provided $T_{6,2,0.2}$-ADP and $T_{6,0.8,0.2}$-ADP for the covariates. The sample sizes were $n = 2500$, 5000, and 7500.

We compared four methods for estimation. The first method used original data to minimize $\sum_{i=1}^n |Y_i - X_i^\top \beta|$ and obtain the oracle estimator $\hat{\beta}_n^{\mathrm{ORA}}$. The second method directly minimized $\sum_{i=1}^n |Y_i - \tilde{X}_i^{(1)\top} \beta|$ to obtain the naive estimator $\hat{\beta}_n^{\mathrm{NAI}}$. The third method, proposed by Wang, Stefanski and Zhu (2012), minimized a smoothed corrected loss by kernel smoothing the absolute value function, which is denoted as sCL. The fourth one was the proposed

TABLE 2

The root mean square errors of the oracle, naive, SL, SDR, and DR estimators for the logistic regression coefficients based on 5000 *simulations at* $T_{6,4,0.2}$*-ADP and* $T_{6,2,0.2}$*-ADP. The oracle estimator was calculated using the original noncorrupted data, which didn't protect data privacy*

n	Attribute-level DP	Method	β_1	β_2	β_3	β_4	β_5	β_6
5000	None	Oracle	0.105	0.107	0.106	0.107	0.103	0.105
	$T_{6,4,0.2}$	Naive	0.728	0.730	0.731	0.728	0.730	0.731
		SL	0.270	0.265	0.262	0.267	0.270	0.271
		SDR	0.244	0.239	0.234	0.238	0.242	0.242
		DR	0.495	0.498	0.495	0.489	0.494	0.495
	None	Oracle	0.104	0.103	0.107	0.111	0.098	0.102
	$T_{6,2,0.2}$	Naive	0.905	0.901	0.913	0.907	0.907	0.911
		SL	0.610	0.618	0.586	0.600	0.609	0.622
		SDR	0.536	0.542	0.517	0.535	0.551	0.557
		DR	0.769	0.751	0.749	0.752	0.782	0.766
7500	None	Oracle	0.086	0.087	0.085	0.086	0.088	0.086
	$T_{6,4,0.2}$	Naive	0.727	0.729	0.728	0.728	0.730	0.730
		SL	0.217	0.218	0.215	0.216	0.218	0.217
		SDR	0.195	0.197	0.191	0.193	0.197	0.193
		DR	0.409	0.407	0.402	0.407	0.411	0.408
	None	Oracle	0.086	0.087	0.085	0.086	0.088	0.086
	$T_{6,2,0.2}$	Naive	0.910	0.912	0.911	0.910	0.912	0.913
		SL	0.522	0.528	0.518	0.518	0.518	0.516
		SDR	0.445	0.455	0.437	0.441	0.447	0.438
		DR	0.706	0.705	0.707	0.713	0.713	0.713
10,000	None	Oracle	0.074	0.075	0.075	0.073	0.075	0.076
	$T_{6,4,0.2}$	Naive	0.728	0.729	0.728	0.729	0.729	0.728
		SL	0.190	0.189	0.184	0.187	0.187	0.186
		SDR	0.170	0.168	0.165	0.168	0.169	0.168
		DR	0.355	0.348	0.351	0.353	0.356	0.360
	None	Oracle	0.074	0.075	0.075	0.073	0.075	0.076
	$T_{6,2,0.2}$	Naive	0.913	0.914	0.913	0.914	0.913	0.913
		SL	0.460	0.461	0.452	0.459	0.461	0.458
		SDR	0.390	0.387	0.380	0.386	0.388	0.388
		DR	0.672	0.660	0.669	0.665	0.664	0.671

DR method. We did not consider the SL and SDR estimators, as the check loss is not twice differentiable everywhere. The average RMSEs of the parameter estimates based on 5000 replications are presented in Table 3.

Table 3 shows that the DR estimator had a smaller RMSE than sCL. Its RMSE decreased with larger sample sizes but increased as the DP level increased (from $T_{6,2,0.2}$-ADP to $T_{6,0.8,0.2}$-ADP). The RMSE of the DR estimator was always larger than that of the oracle estimator, as indicated by Corollary 5.1.

8. Discussion. We have studied the ZIL mechanism and DP M-estimators (DR, SL, and SDR), characterizing their privacy guarantees via the trade-off function and establishing broad applicability with asymptotic properties and variance analysis. Overall, ZIL combined with the DR, SL, and SDR provides a versatile private learning framework. This versatility covers general estimation tasks, allows repeated analyses and growing data streams in online settings, which are not fully achieved by existing methods, as detailed in Section S5 in the SM.

TABLE 3

The root mean square errors of the oracle, naive, sCL (Wang, Stefanski and Zhu (2012)), and the proposed DR estimators for the median regression coefficients based on 5000 *simulations at* $T_{6,2,0.2}$*-ADP and* $T_{6,0.8,0.2}$*-ADP. The oracle estimator was calculated using the original noncorrupted data, which didn't protect data privacy*

n	Attribute-level DP	Method	β_0	β_1	β_2	β_3	β_4	β_5	β_6
2500	None	oracle	0.020	0.037	0.038	0.038	0.037	0.037	0.037
	$T_{6,2,0.2}$	naive	0.037	0.912	0.912	0.911	0.911	0.912	0.912
		sCL	0.111	0.632	0.630	0.637	0.631	0.634	0.635
		DR	0.094	0.443	0.438	0.444	0.438	0.446	0.439
	None	oracle	0.020	0.037	0.038	0.038	0.037	0.037	0.037
	$T_{6,0.8,0.2}$	naive	0.037	0.943	0.942	0.942	0.942	0.943	0.942
		sCL	0.143	0.738	0.731	0.734	0.733	0.737	0.735
		DR	0.100	0.499	0.503	0.503	0.512	0.507	0.504
5000	None	oracle	0.014	0.026	0.026	0.027	0.026	0.027	0.026
	$T_{6,2,0.2}$	naive	0.027	0.911	0.911	0.911	0.912	0.911	0.911
		sCL	0.074	0.484	0.488	0.487	0.487	0.491	0.488
		DR	0.061	0.302	0.296	0.299	0.296	0.300	0.297
	None	oracle	0.014	0.026	0.026	0.027	0.026	0.027	0.026
	$T_{6,0.8,0.2}$	naive	0.028	0.942	0.942	0.942	0.942	0.942	0.942
		sCL	0.101	0.636	0.638	0.636	0.640	0.638	0.637
		DR	0.065	0.375	0.376	0.377	0.374	0.380	0.375
7500	None	oracle	0.011	0.021	0.022	0.021	0.022	0.021	0.022
	$T_{6,2,0.2}$	naive	0.021	0.911	0.912	0.911	0.912	0.912	0.912
		sCL	0.058	0.392	0.395	0.391	0.392	0.392	0.394
		DR	0.049	0.246	0.245	0.244	0.244	0.242	0.240
	None	oracle	0.011	0.021	0.022	0.021	0.022	0.021	0.022
	$T_{6,0.8,0.2}$	naive	0.022	0.942	0.942	0.942	0.942	0.942	0.942
		sCL	0.081	0.556	0.563	0.558	0.559	0.558	0.560
		DR	0.052	0.301	0.304	0.300	0.300	0.303	0.296

Addressing the bounded data requirement of differential privacy remains a future work. To extend to the case where the data distribution does not have a compact support, we may apply the truncation technique to truncate the data to a bounded area, then add noise to each data point, and finally publish the noisy perturbed truncated data. Meanwhile, both the size of the bounded region and the variance of the noise should increase with the sample size, in order to eliminate truncation bias while still protecting privacy.

Acknowledgments. The authors thank the referees, an Associate Editor, and the Editor for constructive comments which have improved the content and presentation of the paper.

Funding. The authors were partially supported by National Natural Science Foundation of China Grants 12292980 and 92358303.

SUPPLEMENTARY MATERIAL

Supplement to "Versatile differentially private learning for general loss functions" (DOI: 10.1214/25-AOS2583SUPP; .pdf). There we present technical details, proofs of theorems and additional numerical results.

REFERENCES

AMORINO, C. and GLOTER, A. (2025). Minimax rate for multivariate data under componentwise local differential privacy constraints. *Ann. Statist.* **53** 1176–1202. MR4925120 https://doi.org/10.1214/25-aos2497

APPLE DIFFERENTIAL PRIVACY TEAM (2017). Learning with privacy at scale.

AVELLA-MEDINA, M. (2021). Privacy-preserving parametric inference: A case for robust statistics. *J. Amer. Statist. Assoc.* **116** 969–983. MR4270037 https://doi.org/10.1080/01621459.2019.1700130

AVELLA-MEDINA, M., BRADSHAW, C. and LOH, P.-L. (2023). Differentially private inference via noisy optimization. *Ann. Statist.* **51** 2067–2092. MR4678796 https://doi.org/10.1214/23-aos2321

BASSILY, R., SMITH, A. and THAKURTA, A. (2014). Private empirical risk minimization: Efficient algorithms and tight error bounds. In 55*th Annual IEEE Symposium on Foundations of Computer Science—FOCS* 2014 464–473. IEEE Comput. Soc., Los Alamitos, CA. MR3344896 https://doi.org/10.1109/FOCS.2014.56

BEIMEL, A., NISSIM, K. and STEMMER, U. (2016). Private learning and sanitization: Pure vs. approximate differential privacy. *Theory Comput.* **12** Paper No. 1, 61. MR3518176 https://doi.org/10.4086/toc.2016.v012a001

BIE, A., KAMATH, G. and SINGHAL, V. (2022). Private estimation with public data. In *Advances in Neural Information Processing Systems* **35** 18653–18666. Curran Associates, Red Hook.

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

CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186. MR0997599

CHANG, J., HU, Q., KOLACZYK, E. D., YAO, Q. and YI, F. (2024). Edge differentially private estimation in the β-model via jittering and method of moments. *Ann. Statist.* **52** 708–728. MR4744193 https://doi.org/10.1214/24-aos2365

CHAUDHURI, K., MONTELEONI, C. and SARWATE, A. D. (2011). Differentially private empirical risk minimization. *J. Mach. Learn. Res.* **12** 1069–1109. MR2786918

DING, B., KULKARNI, J. and YEKHANIN, S. (2017). Collecting telemetry data privately. In *Advances in Neural Information Processing Systems*, *Vol.* 30. Curran Associates, Red Hook.

DONG, J., ROTH, A. and SU, W. J. (2022). Gaussian differential privacy. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 3–54. With discussions and a reply by the authors. MR4400389

DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Local privacy and minimax bounds: Sharp rates for probability estimation. In *Proceedings of the* 27*th International Conference on Neural Information Processing Systems*, *Vol.* 1 1529–1537. Curran Associates, Red Hook.

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

DUCHI, J. C. and RUAN, F. (2024). The right complexity measure in locally private estimation: It is not the Fisher information. *Ann. Statist.* **52** 1–51. MR4718406 https://doi.org/10.1214/22-aos2227

DWORK, C., KENTHAPADI, K., MCSHERRY, F., MIRONOV, I. and NAOR, M. (2006a). Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology—EUROCRYPT* 2006. *Lecture Notes in Computer Science* **4004** 486–503. Springer, Berlin. MR2423560 https://doi.org/10.1007/11761679_29

DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006b). 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

DWORK, C. and ROTH, A. (2013). The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.* **9** 211–487. MR3254020 https://doi.org/10.1561/0400000042

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.

FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324 https://doi.org/10.1214/aos/1176348248

FIRPO, S., GALVAO, A. F. and SONG, S. (2017). Measurement errors in quantile regression models. *J. Econometrics* **198** 146–164. MR3628103 https://doi.org/10.1016/j.jeconom.2017.02.002

KASIVISWANATHAN, S. P., LEE, H. K., NISSIM, K., RASKHODNIKOVA, S. and SMITH, A. (2011). What can we learn privately? *SIAM J. Comput.* **40** 793–826. MR2823508 https://doi.org/10.1137/090756090

KENT, D. and RUPPERT, D. (2024). Smoothness-penalized deconvolution (SPeD) of a density estimate. *J. Amer. Statist. Assoc.* **119** 2407–2417. MR4797950 https://doi.org/10.1080/01621459.2023.2259028

KIFER, D. and MACHANAVAJJHALA, A. (2014). Pufferfish: A framework for mathematical privacy definitions. *ACM Trans. Database Syst.* **39** Art. 3, 36. MR3238192 https://doi.org/10.1145/2514689

KOTZ, S., KOZUBOWSKI, T. J. and PODGÓRSKI, K. (2001). *The Laplace Distribution and Generalizations*: *A Revisit with Applications to Communications*, *Economics*, *Engineering*, *and Finance*. Birkhäuser, Boston, MA. MR1935481 https://doi.org/10.1007/978-1-4612-0173-1

LEI, J. (2011). Differentially private M-estimators. In *Advances in Neural Information Processing Systems*, *Vol*. 24. Curran Associates, Red Hook.

LU, Q., CHEN, S. X. and QIU, Y. (2026). Supplement to "Versatile differentially private learning for general loss functions." https://doi.org/10.1214/25-AOS2583SUPP

NISSIM, K., RASKHODNIKOVA, S. and SMITH, A. (2007). Smooth sensitivity and sampling in private data analysis. In *STOC'*07*—Proceedings of the* 39*th Annual ACM Symposium on Theory of Computing* 75–84. ACM, New York. MR2402430 https://doi.org/10.1145/1250790.1250803

PAN, R., REN, T., GUO, B., LI, F., LI, G. and WANG, H. (2022). A note on distributed quantile regression by pilot sampling and one-step updating. *J. Bus. Econom. Statist.* **40** 1691–1700. MR4492062 https://doi.org/10.1080/07350015.2021.1961789

RAJKUMAR, A. and AGARWAL, S. (2012). A differentially private stochastic gradient descent algorithm for multiparty classification. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*. *Proceedings of Machine Learning Research* **22** 933–941. PMLR.

STEFANSKI, L. and CARROLL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184. MR1054861 https://doi.org/10.1080/02331889008802238

STEFANSKI, L. A. (1989). Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Comm. Statist. Theory Methods* **18** 4335–4358. MR1046712 https://doi.org/10.1080/03610928908830159

STEINBERGER, L. (2024). Efficiency in local differential privacy. *Ann. Statist.* **52** 2139–2166. MR4829483 https://doi.org/10.1214/24-aos2425

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

WANG, D. and XU, J. (2019). On sparse linear regression in the local differential privacy model. In *International Conference on Machine Learning*, *Vol*. 97 6628–6637. PMLR.

WANG, H. J., STEFANSKI, L. A. and ZHU, Z. (2012). Corrected-loss estimation for quantile regression with covariate measurement errors. *Biometrika* **99** 405–421. MR2931262 https://doi.org/10.1093/biomet/ass005

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

YANG, R., APLEY, D. W., STAUM, J. and RUPPERT, D. (2020). Density deconvolution with additive measurement errors using quadratic programming. *J. Comput. Graph. Statist.* **29** 580–591. MR4153183 https://doi.org/10.1080/10618600.2019.1704294

ZHANG, Y., XU, Q., TANG, N. and QU, A. (2024). Differentially private data release for mixed-type data via latent factor models. *J. Mach. Learn. Res.* **25** Paper No. [116], 37. MR4749768