LEARNING WITH USER-LEVEL LOCAL DIFFERENTIAL PRIVACY

Anonymous authors
Paper under double-blind review

ABSTRACT

User-level privacy is important in distributed systems. Previous research primarily focuses on the central model, while the local models have received much less attention. Under the central model, user-level DP is strictly stronger than the item-level one. However, under the local model, the relationship between user-level and item-level LDP becomes more complex, thus the analysis is crucially different. In this paper, we first analyze the mean estimation problem and then apply it to stochastic optimization, classification, and regression. In particular, we propose adaptive strategies to achieve optimal performance at all privacy levels. Moreover, we also obtain information-theoretic lower bounds, which show that the proposed methods are minimax optimal up to logarithmic factors. Unlike the central DP model, where user-level DP always leads to slower convergence, our result shows that under the local model, the convergence rates are nearly the same between user-level and item-level cases for distributions with bounded support. For heavy-tailed distributions, the user-level rate is even faster than the item-level one.

1 INTRODUCTION

Differential privacy (DP) (Dwork et al., 2006) is one of the mainstream schemes for privacy protection. The traditional DP framework is item-level, which focuses on the privacy of each sample (Dwork et al., 2014). However, in many real-world scenarios such as federated learning (Liu et al., 2025; Bakas et al., 2026), each user provides multiple samples, which need to be treated as a whole for privacy protection. Therefore, in recent years, user-level differential privacy has emerged and has received widespread attention from researchers (Charles et al., 2025; Qiu et al., 2026).

Existing research on user-level DP mainly focuses on central models (Ghazi et al., 2023). Much less effort has been made on local models. (Acharya et al., 2023) analyzes the discrete distribution estimation problem under user-level local differential privacy (LDP). Another existing work is Bassily & Sun (2023), which analyzes the optimization problem under user-level ϵ-LDP with $\epsilon < 1$. However, user-level LDP has a much wider range of potential applications (Yang et al., 2020), which have been rarely explored before. Moreover, the case with $\epsilon > 1$ is also important. Due to limited sample size, for a better privacy-utility tradeoff, it is inevitable in many cases to user larger ϵ. Therefore, a more in-depth study of learning under user-level LDP is needed. A challenge under the local model is to make the method suitable for all privacy levels. Under the central model, we can just let the noise scales as $1/\epsilon$, while the algorithm structure remains the same for different ϵ. Nevertheless, under the local model, privatization takes place before aggregation. To achieve optimal performance, privacy mechanisms need to be tailored to each privacy level ϵ. Moreover, it is also not straightforward to derive the information-theoretic lower bound, since under the local model, user-level LDP is not necessarily stronger than the item-level one, thus the item-level lower bounds do not directly lead to user-level counterparts.

In this paper, we discuss a wide range of statistical tasks under user-level ϵ-LDP. We analyze the mean estimation problem first, including one-dimensional and multi-dimensional cases. We then apply the mean estimation methods to other tasks, including stochastic optimization, classification, and regression. For each task, we provide algorithms and analyze the theoretical convergence rates. Moreover, we derive the information-theoretic lower bounds, which shows that the newly proposed methods are minimax rate optimal up to a logarithm factor. The results are shown in Table 1, in which the non-private term is omitted for simplicity. Under central DP, user-level DP is strictly

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Table 1: Comparison of performance under user-level and item-level LDP.

Tasks	user-level n users, m samples per user	item-level nm samples
Mean, bounded	$\tilde{O}\Big(\frac{d}{nm(\epsilon^2\wedge\epsilon)}\Big)$ $(n(\epsilon^2\wedge 1)\gtrsim d\ln m,$ Theorem 5)	$O\Big(\frac{d}{nm(\epsilon^2\wedge\epsilon)}\Big)$ (Asi et al., 2022)
Mean, heavy-tail	$\tilde{O}\left(\frac{d\ln m}{mn(\epsilon^2\wedge\epsilon)}\vee\left(\frac{d}{m^2n(\epsilon^2\wedge\epsilon)}\right)^{1-\frac{1}{p}}\right)$ $(n(\epsilon^2\wedge 1)\gtrsim d\ln m,$ Remark 2)	$O\left(\left(\frac{d}{nm(\epsilon^2\wedge\epsilon)}\right)^{1-\frac{1}{p}}\right)$ (Duchi et al., 2018)[a]
Stochastic optimization	$\tilde{O}\left(\sqrt{\frac{d}{nm(\epsilon^2\wedge\epsilon)}}\right)$ $(n(\epsilon^2\wedge 1)\gtrsim d\ln n\ln m,$ Theorem 8)	$\tilde{O}\left(\sqrt{\frac{d}{nm(\epsilon^2\wedge\epsilon)}}\right)$ (Duchi et al., 2013)[b]
Classification	$\tilde{O}\Big((mn(\epsilon^2\wedge\epsilon))^{-\frac{\beta(1+\gamma)}{2(d+\beta)}}\Big)$ $(n(\epsilon^2\wedge 1)\gtrsim d\ln(mn),$ Theorem 9)	$O\Big((mn(\epsilon^2\wedge\epsilon))^{-\frac{\beta(1+\gamma)}{2(d+\beta)}}\Big)$ (Berrett & Butucea, 2019)
Regression	$\tilde{O}\Big((mn(\epsilon^2\wedge\epsilon))^{-\frac{\beta}{d+\beta}}\Big)$ $(n(\epsilon^2\wedge 1)\gtrsim d\ln(mn),$ Theorem 11)	$O\Big((mn(\epsilon^2\wedge\epsilon))^{-\frac{\beta}{d+\beta}}\Big)$ (Berrett et al., 2021)

[a] For heavy-tailed distributions, Duchi et al. (2018) analyzed the one-dimensional case. We generalize it to d dimensions.

[b] Duchi et al. (2013) analyzed the case with $\epsilon\le 1/4$. We generalize it to larger ϵ.

stronger than item-level one, and thus always leads to a slower convergence rate (Levy et al., 2021). On the contrary, under the local model, the same convergence rates are derived between user-level and item-level cases for distributions with bounded support. If the distribution is heavy-tailed, then perhaps surprisingly, the user-level rate is even faster, such as those shown in the second row in Table 1.

To achieve optimal performance at all privacy levels, we design an adaptive method tailored to each ϵ. For example, for the d dimensional mean estimation problem, with $\epsilon<1$, the dataset is divided into d groups to estimate each component. The number of groups then decreases with increasing ϵ. With very large ϵ, there is only one group, so the error becomes close to the non-private case. With the algorithm varying among different ϵ, it is natural to see phase transitions in the bounds in Table 1. For the establishment of minimax lower bounds, we revisit classical minimax theory (Tsybakov, 2009) to derive tight bounds on the distances between privatized samples.

The main contributions of this paper are summarized as follows.

- For the mean estimation problem, we use a two-stage approach for $d=1$. With higher dimensionality, for ℓ_∞ support, our method divides users into groups, and the strategy of such grouping is tailored to the privacy level ϵ. We then use Kashin's representation to obtain a tight result for ℓ_2 support.
- We apply the mean estimation to the stochastic optimization problem and derive a rate of $\tilde{O}(d/(nm(\epsilon^2\wedge\epsilon))$, matches the item-level bound in Duchi et al. (2013) under the same total sample size.
- For nonparametric classification and regression, we divide the support into grids and apply the Hadamard transform, which is shown to be optimal under user-level LDP.

In general, the results show that the user-level LDP requirement is similar or sometimes even weaker than the item-level one, which is crucially different from the central model.

Item-level DP. We start with mean estimation, which is a basic but important statistical task since it serves as building blocks of stochastic optimization and deep learning (Abadi et al., 2016), which requires estimating the mean of gradients. (Vithana et al., 2025; Agarwal et al., 2025; Huang et al., 2021; Liu et al., 2021; Hopkins et al., 2022; Vargaftik et al., 2021) studied mean estimation under central DP. For the local model, (Duchi et al., 2018) introduces an order optimal mean estimation method, which is then improved in (Li et al., 2023; Feldman & Talwar, 2021). Moreover, (Chen et al., 2020) achieved optimal communication cost. (Bhowmick et al., 2018) proposed PrivUnit, which

is then shown to be optimal in constants (Asi et al., 2022). (Asi et al., 2024a) proposed ProjUnit, which reduces the communication complexity of PrivUnit. Mean estimation can be used in other problems. For example, in stochastic optimization, various methods have been proposed under central DP requirements (Dvorkin et al., 2025; Xu et al., 2025; Xie et al., 2026; Feldman et al., 2020; Asi et al., 2021; Kamath et al., 2022). Under local DP, (Duchi et al., 2013) proposed a stochastic gradient method, which calculates the noisy gradient from each sample and then update the model. For nonparametric statistics, (Duchi et al., 2013; 2018) shows that the nonparametric density estimation under LDP has a convergence rate of $O((n\epsilon^2)^{-\beta/(d+\beta)})$ for small ϵ, which is inevitably slower than the non-private rate $O(n^{-\beta/(d+2\beta)})$ (Tsybakov, 2009). (Berrett & Butucea, 2019) and (Berrett et al., 2021) extend the analysis to nonparametric classification and regression problems, respectively.

User-level DP. Under central model, (Geyer et al., 2017) proposes a simple clipping method. (Levy et al., 2021) designs a two-stage approach for one-dimensional mean estimation, and then extends to higher dimension using the Hadamard transform. This method is then used in stochastic optimization problems (Bassily & Sun, 2023; Liu & Asi, 2024). Additionally, some works are focusing on black-box conversion from item-level DP to user-level, such as (Ghazi et al., 2021; Bun et al., 2023; Ghazi et al., 2023). Under the local model, (Acharya et al., 2023) studies the discrete distribution estimation problem. Bassily & Sun (2023) studies the optimization problem with $\epsilon < 1$.

Compared with existing works, to the best of our knowledge, our work is the first attempt to analyze mean estimation and stochastic optimization problems under user-level LDP for general ϵ. Unlike the central model, in which a single algorithm structure is enough, we have to design adaptive privacy mechanisms that are tailored to every possible ϵ under the local model. Moreover, we also provide the first analysis on nonparametric classification and regression problems under user-level ϵ-LDP.

2 PRELIMINARIES

Suppose there are n users, and each user has m identical and independently distributed (i.i.d) samples, denoted as $\mathbf{X}_{ij} \in \mathcal{X}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. Let $\mathbf{X}_i = \{\mathbf{X}_{i1}, \ldots, \mathbf{X}_{im}\}$ be the set of all samples stored in user i. Due to privacy concerns, users are unwilling to upload $\mathbf{X}_i$ directly. Instead, there is a privacy mechanism that transforms $\mathbf{X}_1, \ldots, \mathbf{X}_n$ into n random variables $\mathbf{Z}_1, \ldots, \mathbf{Z}_n \in \mathcal{Z}$ with $\mathbf{Z}_i = M_i(\mathbf{X}_i, \mathbf{Z}_1, \ldots, \mathbf{Z}_{i-1})$, in which $M_i : \mathcal{X} \times \mathcal{Z}^{i-1} \to \mathcal{Z}$ is a function with random output. The user-level LDP is defined as follows.

Definition 1. *Given a privacy parameter $\epsilon \geq 0$, the privacy mechanism M_i is user-level ϵ-LDP if for all i, all values of $\mathbf{z}_1, \ldots, \mathbf{z}_{i-1}$, all $\mathbf{x}, \mathbf{x}' \in \mathcal{X}^m$ and all $S \subseteq \mathcal{Z}$,*

$$P(\mathbf{Z}_i \in S | \mathbf{X}_i = \mathbf{x}, \mathbf{Z}_{1:i-1} = \mathbf{z}_{1:i-1}) \leq e^{\epsilon} P(\mathbf{Z}_i \in S | \mathbf{X}_i = \mathbf{x}', \mathbf{Z}_{1:i-1} = \mathbf{z}_{1:i-1}), \tag{1}$$

in which $\mathbf{Z}_i = M_i(\mathbf{X}_i, \mathbf{Z}_1, \ldots, \mathbf{Z}_{i-1})$, $\mathbf{Z}_{1:i-1} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_{i-1})$, $\mathbf{z}_{1:i-1} = (\mathbf{z}_1, \ldots, \mathbf{z}_{i-1})$.

Definition 1 requires that the distributions of $\mathbf{Z}_i$ should not change much even if the whole local dataset $\mathbf{X}_i = \{\mathbf{X}_{i1}, \ldots, \mathbf{X}_{im}\}$ is altered. From equation 1, even if the adversary can observe $\mathbf{Z}_i$, it can not infer the value of $\mathbf{X}_i$ exactly. Smaller ϵ indicates stronger privacy protection since it is harder to distinguish $\mathbf{X}_i$. The difference between item-level and user-level LDP is illustrated in Figure 1. In the item-level case, each sample is transformed into a privatized one, while in the user-level case, all samples of a user are combined to generate a privatized sample. For both item-level and user-level cases, at the final step, all privatized samples are aggregated to generate the output. One natural question is how the difficulty of achieving user-level LDP compares with the item-level counterparts. Regarding this question, we have the following statements.

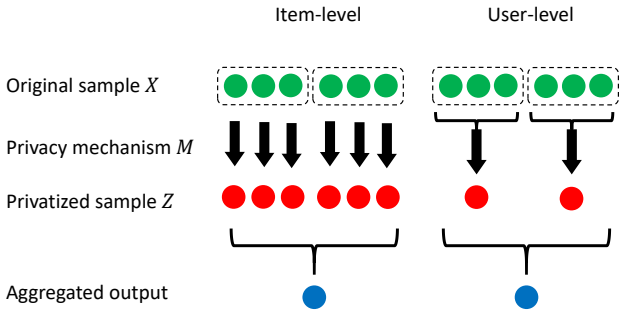


Figure 1: Comparison of item-level versus user-level LDP. Dashed rectangles represent users.

Proposition 1. *Based on Definition 1, for any statistical problems, there are two basic facts:*

(1) If item-level (ϵ/m)-LDP can be achieved with nm samples, then user-level ϵ-LDP can be achieved using n users with m samples per user;

(2) If item-level ϵ-LDP can be achieved with n samples, then user-level ϵ-LDP can be achieved using n users with m samples per user.

In the above statements, (1) holds due to the group privacy property. For (2), if a task can be solved using n samples under item-level ϵ-LDP, then just randomly picking a sample from each user satisfies user-level ϵ-LDP. These results also suggest two baseline methods that transform item-level methods to user-level. However, these simple conversions are far from optimal. For the first one, (ϵ/m)-LDP is too strong. For the second one, many samples are wasted.

One may wonder if user-level LDP is a stronger requirement than the item-level one. In other words, if item-level ϵ-LDP can be achieved with nm samples, then can we achieve user-level ϵ-LDP using n users with m samples per user? Under the central model, the answer is affirmative: user-level DP is stronger because the definition of user-level ϵ-DP ensures item-level ϵ-DP (Levy et al., 2021). Nevertheless, under the local model, things become more complex. On the one hand, user-level LDP imposes stronger privacy requirements, since the distribution of the output variables can only change to a limited extent even when the local dataset is replaced as a whole. On the other hand, user-level LDP enables the encoder to obtain complete information of local datasets, thus the difficulty is somewhat reduced in this aspect. From Table 1, for many problems, with the same total sample sizes, user-level and item-level LDP yield nearly the same error bounds. If the distribution has tails, then the user-level LDP is even easier to achieve, which is perhaps surprising.

Before discussing each task in detail, we clarify some notations that will be used in subsequent sections. Denote $a \wedge b = \min(a, b)$, $a \vee b = \max(a, b)$, and $a \lesssim b$ if there exists a constant C such that $a \leq Cb$. Conversely, $a \gtrsim b$ means $a \geq Cb$. $a \sim b$ means that $a \lesssim b$ and $a \gtrsim b$ both hold.

3 Mean Estimation

For one-dimensional problem, we introduce a two-stage method. Despite that similar idea has also been used in central user-level DP (Levy et al., 2021), details and theoretical analysis are different. We then extend the analysis to high-dimensional problems. To achieve optimal convergence rate for all privacy levels, our strategies are designed separately for each ϵ.

3.1 One Dimensional Case

We start with the case such that the distribution has bounded support $\mathcal{X} = [-D, D]$ for some D, and introduce a two-stage method. The first stage uses half of the users to identify an interval $[L, R]$, which is much smaller than $[-D, D]$ but contains $\mu := \mathbb{E}[X]$ with high probability. The purpose of this stage is to significantly reduce the strength of Laplacian noise needed to protect privacy, and thus reduce the negative effect on the estimation accuracy caused by privacy mechanisms. At the second stage, the algorithm truncates the values into $[L, R]$, and adds a Laplacian noise to ensure ϵ-LDP at user-level. Finally, μ can be estimated with a simple average over the other half of users. The details are provided in Algorithm 1.

Algorithm 1 MeanEst1d: One dimensional mean estimation under user-level ϵ-LDP

Input: Dataset containing n users with m samples per user, i.e. X_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$
Output: Estimated mean $\hat{\mu}$
Parameter: h, Δ, D, ϵ

1: Calculate $Y_i = (1/m) \sum_{j=1}^{m} X_{ij}$ for $i = 1, \ldots, n/2$;
2: Divide $[-D, D]$ into B bins of length h;
3: $Z_{ik} = \mathbf{1}(Y_i \in B_k) + W_{ik}$ for $i = 1, \ldots, n/2$, $k = 1, \ldots, B$, in which $W_{ik} \sim \mathrm{Lap}(2/\epsilon)$;
4: Calculate $s_k = \sum_{i=1}^{n/2} Z_{ik}$ for $k = 1, \ldots, B$;
5: Let $\hat{k}^* = \arg\max_k s_k$;
6: $L = -D + (\hat{k}^* - 2)h$;
7: $R = -D + (\hat{k}^* + 1)h$;
8: $Z_i = (Y_i \vee (L - \Delta)) \wedge (R + \Delta) + W_i$ for $i = n/2 + 1, \ldots, n$, in which $W_i \sim \mathrm{Lap}((3h + 2\Delta)/\epsilon)$;
9: Calculate $\hat{\mu} = (2/n) \sum_{i=n/2+1}^{n} Z_i$;
10: **Return** $\hat{\mu}$

The privacy guarantee and the estimation error of Algorithm 1 are both analyzed in Theorem 1. In Algorithm 1, $\mathrm{Lap}(\lambda)$ means Laplacian distribution with parameter λ.

Theorem 1. *Algorithm 1 is user-level ϵ-LDP. If $n(\epsilon^2 \wedge 1) \geq c_1 \ln m$ for a constant c_1, then with $h = 4D/\sqrt{m}$ and $\Delta = D\sqrt{\ln n/m}$, the mean squared error of one dimensional mean estimation under user-level ϵ-LDP satisfies*

$$\mathbb{E}[(\hat{\mu} - \mu)^2] \lesssim \frac{D^2}{nm}\left(1 + \frac{\ln n}{\epsilon^2}\right). \tag{2}$$

The proof of Theorem 1 is shown in Appendix A. In Appendix A.2, we show that $[L, R]$ contains μ with high probability. To begin with, $\hat{k}^* \in \{k^* - 1, k^*, k^* + 1\}$ holds with high probability, in which k^* is the index of the bin containing μ, i.e. $\mu \in B_{k^*}$. Let L be the left bound of the $(\hat{k}^* - 1)$-th bin, and R be the right bound of the $(\hat{k}^* + 1)$-th bin, then with high probability, $\mu \in [L, R]$. In Appendix A.3, we then bound the bias and variance separately. As shown in Proposition 1, there are two baseline methods to achieve user-level LDP from item-level LDP. The first one is to achieve item-level (ϵ/m)-LDP for all samples. This yields a bound $O(D^2 m/(n\epsilon^2) + D^2/(nm))$. The second one is to achieve item-level ϵ-LDP for n samples randomly selected from n users, which also only yields $O(D^2/(n(\epsilon^2 \wedge 1)))$, significantly worse than the right hand side of equation 2.

In Theorem 1, the requirement $n(\epsilon^2 \wedge 1) \geq c_1 \ln m$ is necessary since if n is fixed, then the mean squared error will never converge to zero with increasing m. From an information-theoretic perspective, a fixed number of privatized variables can only transmit limited information (Cuff & Yu, 2016; Wang et al., 2016). Therefore, it is necessary to let n grow with m, which is also discussed in (Levy et al., 2021) for user-level central DP. Theorem 2 shows the information-theoretic minimax lower bound.

Theorem 2. *Denote $\mathcal{P}_{\mathcal{X}}$ as the set of all distributions supported on $\mathcal{X} = [-D, D]$, $\mathcal{M}_\epsilon$ as all mechanisms satisfying ϵ-LDP, then*

$$\inf_{\hat{\mu}} \inf_{M \in \mathcal{M}_\epsilon} \sup_{p \in \mathcal{P}_{\mathcal{X}}} \mathbb{E}\left[(\hat{\mu} - \mu)^2\right] \gtrsim \frac{D^2}{nm(\epsilon^2 \wedge 1)}. \tag{3}$$

Moreover, with fixed n, the mean squared error will not converge to zero as m increases. To be more precise, $\mathbb{E}[(\hat{\mu} - \mu)^2] \geq (1/4)D^2 e^{-n\epsilon(e^\epsilon - 1)}$.

The proof of Theorem 2 is shown in Appendix A.4. The comparison between equation 2 and equation 3 shows that the upper and lower bounds match up to a logarithm factor, thus the two-stage method is nearly minimax optimal. Finally, we extend the method to the case with unbounded support. In this case, we replace step 1 in Algorithm 1 with $Y_i = -D \vee \left(\bar{X}_i \wedge D\right)$, in which $\bar{X}_i = (1/m)\sum_{j=1}^m X_{ij}$ is the i-th user-wise mean. Such clipping operation controls the sensitivity. Other steps are the same as Algorithm 1. The convergence rate is shown in Theorem 3.

Theorem 3. *Assume that $\mathbb{E}[|X|^p] \leq M_p < \infty$ for some finite constant M_p, with $p \geq 2$. If $n(\epsilon^2 \wedge 1) \geq c_1 \ln m$, then with Algorithm 1, except that step 1 is replaced by $Y_i = -D \vee \left(\bar{X}_i \wedge D\right)$, the mean squared error of $\hat{\mu}$ can be bounded by*

$$\mathbb{E}[(\hat{\mu} - \mu)^2] \lesssim M_p^{2/p}\left[\frac{\ln m}{mn\epsilon^2} \vee (m^2 n\epsilon^2)^{-\left(1-\frac{1}{p}\right)} + \frac{1}{mn}\right]. \tag{4}$$

The selection of D and the proof of Theorem 3 are shown in Appendix A.5. Here we provide an intuitive understanding of the phase transition in equation 4. As long as $p \geq 2$, from central limit theorem, with large m, similar to the case with bounded support, Y_i is nearly normally distributed, and the tail is like a Gaussian distribution. Therefore, the convergence rate of the mean squared error is still $O(\ln m/(mn\epsilon^2))$, the same as the case with bounded support. However, if m is small, the Gaussian approximation no longer holds. In this case, the tail of the distribution of Y_i is polynomial. As a result, there is a phase transition in equation 4. Mean estimation for heavy-tailed distributions is an example that user-level LDP is easier to achieve than the item-level one. With nm samples, mean squared error under item-level ϵ-LDP is $O((nm\epsilon^2)^{1-1/p})$ (Duchi et al., 2018), significantly worse than equation 4.

3.2 Multi-dimensional Case

This section discusses the mean estimation problem with $d \geq 1$. Depending on the shape of the support set, the problem can be crucially different. Here we discuss two cases, i.e. ℓ_2 support $\mathcal{X}_2 = \{\mathbf{u} | \|\mathbf{u}\|_2 \leq D\}$, and ℓ_∞ support $\mathcal{X}_\infty = \{\mathbf{u} | \|\mathbf{u}\|_\infty \leq D\}$. For small ϵ, the mean squared error under item-level ϵ-LDP is $O(d/(n(\epsilon^2 \wedge \epsilon)))$ for ℓ_2 support, and $O(d^2/(n(\epsilon^2 \wedge \epsilon)))$ for ℓ_∞ support (Duchi et al., 2018; Asi et al., 2022; Feldman & Talwar, 2021; Asi et al., 2024a). Similar to the one-dimensional case, direct transformation to user-level according to Proposition 1 yields a suboptimal bound.

Algorithm 2: MeanEst: Multi-dimensional mean estimation under user-level ϵ-LDP with ℓ_∞ support

Input: Dataset containing n users with m samples per user, i.e., $\mathbf{X}_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$
Output: Estimated mean $\hat{\mu}$
Parameter: h, Δ, D, ϵ

```
if ε < 1 then
    Divide users randomly into d groups S_1, ..., S_d;
    for k = 1, ..., d do
        Estimate μ̂_k with S_k using Algorithm 1 under ε-LDP;
    end for
else if 1 ≤ ε < d ln n then
    Divide users into ⌈d/ε⌉ groups S_1, ..., S_⌈d/ε⌉;
    for k = 1, ..., ⌈d/ε⌉ do
        for l = (k − 1)ε + 1, ..., kε ∧ d do
            Estimate μ̂_l with S_k using Algorithm 1 under 1-LDP;
        end for
    end for
else
    for k = 1, ..., d do
        Estimate μ̂_k with all users using Algorithm 1 under (ε/d)-LDP;
    end for
end if
return μ̂ = (μ̂_1, ..., μ̂_d)
```

ℓ_∞ **Support.** To begin with, we focus on this relatively simpler case. The method depends on the value of ϵ. Details are stated in Algorithm 2.

1) High privacy ($\epsilon < 1$). Users are assigned randomly into d groups, and the k-th group is used to estimate μ_k (the k-th component of $\mu := \mathbb{E}[\mathbf{X}]$) for $k = 1, \ldots, K$. Since the size of each group is n/d, from equation 2, we have $\mathbb{E}[(\hat{\mu}_k - \mu_k)^2] \lesssim \frac{D^2}{(n/d)m}\left(1 + \frac{\ln(n/d)}{\epsilon^2}\right) \lesssim \frac{D^2 d \ln n}{nm(\epsilon^2 \wedge 1)}$.

2) Medium privacy ($1 \leq \epsilon < d \ln n$). In this case, the privacy requirement is weaker than the case with $\epsilon < 1$. Therefore, a group of users can be used to estimate more components, with ϵ-LDP still satisfied. Without loss of generality, suppose that ϵ is an integer (otherwise one can just strengthen the requirement to $\lfloor \epsilon \rfloor$-LDP). In this case, users are randomly allocated to $\lceil d/\epsilon \rceil$ groups. Each group is used to estimate ϵ components, and each component is estimated under user-level 1-LDP. From basic composition theorem (Dwork et al., 2010), estimating ϵ components of μ satisfies user-level ϵ-LDP. Denote n_0 as the number of users in each group, then $\mathbb{E}[(\hat{\mu}_k - \mu_k)^2] \lesssim \frac{D^2}{n_0 m}(1 + \ln n_0) \sim \frac{D^2 \ln(n\epsilon/d)}{(n\epsilon/d)m} \lesssim \frac{D^2 d}{nm\epsilon} \ln n$. In the first step, we replace n and ϵ in equation 2 with n_0 and 1 respectively, since now we are using a group with n_0 users to achieve 1-LDP.

3) Low privacy ($\epsilon \geq d \ln n$). In this case, the privacy protection is much less important. We hope that the estimation error is as close to the non-private case as possible. Based on such intuition, we no longer divide users into groups. Instead, our method just estimates each component under user-level (ϵ/d)-LDP, then the whole algorithm is ϵ-LDP. In this case, the mean squared error of each component is bounded by $\mathbb{E}[(\hat{\mu}_k - \mu_k)^2] \lesssim \frac{D^2}{nm}\left(1 + \frac{d^2 \ln n}{\epsilon^2}\right) \lesssim \frac{D^2}{nm}$. Note that $\mathbb{E}[\|\hat{\mu} - \mu\|^2] \leq \sum_{k=1}^d \mathbb{E}[(\hat{\mu}_k - \mu_k)^2]$. A combination of the above privacy cases yields the following theorem.

Theorem 4. *Under user-level ϵ-LDP, if $n(\epsilon^2 \wedge 1) \geq c_1 d \ln m$, in which c_1 is the constant in Theorem 1, then the mean squared error of multi-dimensional mean estimation in $\mathcal{X}_\infty$ with Algorithm 2 is bounded by $\mathbb{E}\left[\|\hat{\mu} - \mu\|_2^2\right] \lesssim \frac{D^2 d}{nm}\left(1 + \frac{d \ln n}{\epsilon^2 \wedge \epsilon}\right)$.*

We would like to remark that under central DP, the loss caused by privacy mechanisms and the non-private loss are two separate terms, and we only need to select the aggregator to minimize the latter one, which does not depend on ϵ. However, under the local model, privatization takes place before aggregation. Depending on ϵ, the optimal randomization can be crucially different. Therefore, it is necessary to discuss each ϵ separately. In Theorem 4, we give a complete picture of the estimation error caused by different privacy levels. In particular, with $\epsilon \to \infty$, the upper bound converges to $D^2 d/(nm)$, which is just the non-private rate.

ℓ_2 **Support.** Consider that ℓ_2 support is smaller than the ℓ_∞ support, we expect that the bound of mean squared error can be improved over Theorem 4. Directly applying Algorithm 2 does not make any improvement. Therefore, a more efficient approach is needed to achieve a better bound. Towards this goal, we use Kashin's representation (Lyubarskii & Vershynin, 2010), which has also been used in other problems related to stochastic estimation (Feldman et al., 2021; Chen et al., 2023; Asi et al., 2024b). To begin with, we rephrase Kashin's representation as follows.

Lemma 1. *(Kashin's representation, rephrased from Theorem 2.2 in (Lyubarskii & Vershynin, 2010)) There exists a matrix $\mathbf{U} \in \mathbb{R}^{2d\times d}$ and a constant K, such that $\mathbf{U}^T\mathbf{U} = \mathbf{I}_d$, in which I_d is the $d \times d$ identity matrix, and for all $\mathbf{x}$ with $\|\mathbf{x}\|_2 \le 1$, $\|\mathbf{U}\mathbf{x}\|_\infty \le K/\sqrt{d}$.*

Based on Lemma 1, our method constructs matrix $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_{2d})^T \in \mathbb{R}^{2d\times d}$. Then we can transform all samples. Let $\mathbf{X}'_{ij} = \mathbf{U}\mathbf{X}_{ij}$ for $i = 1, \dots, n$, $j = 1, \dots, m$. Correspondingly, denote $\theta = \mathbf{U}\mu$ as the mean vector after transformation. Then μ can be estimated by estimating θ first. Since $\mathbf{X}_{ij} \in \mathcal{X}_2$, $\|\mathbf{X}_{ij}\|_2 \le D$ holds. According to Lemma 1, $\|\mathbf{X}'_{ij}\|_\infty \le KD/\sqrt{d}$. Therefore, we have transformed the ℓ_2 support into ℓ_∞ support, thus θ can be estimated using Algorithm 2. The only difference is that now the supremum norm is reduced from D to $KD/\sqrt{d}$. After getting $\hat{\theta}$, we then transform it back to ℓ_2 support, i.e. $\hat{\mu} = \mathbf{U}^T\hat{\theta}$. Since $\hat{\theta}$ is user-level ϵ-LDP, it is guaranteed that $\hat{\mu}$ is also user-level ϵ-LDP. The following theorem bounds the mean squared error of $\hat{\mu}$.

Theorem 5. *Under user-level ϵ-LDP, if $n(\epsilon^2 \wedge 1) \ge c_1 d \ln m$, then the mean squared error of multi-dimensional mean estimation in $\mathcal{X}_2$ is bounded by $\mathbb{E}\left[\|\hat{\mu} - \mu\|_2^2\right] \lesssim \frac{D^2}{nm}\left(1 + \frac{d\ln n}{\epsilon^2\wedge\epsilon}\right)$.*

The proof of Theorem 5 is shown in Appendix B.1.

Remark 1. *If the support is ℓ_1, then we can also let $\mathbf{U} = \mathbf{H}_d/\sqrt{d}$, in which $\mathbf{H}_d$ is the $d \times d$ Hadamard matrix (Hedayat & Wallis, 1978). This can be used in the discrete distribution estimation problem. With alphabet size A, each sample $\mathbf{X}_{ij}$ can be viewed as a A dimensional vector, such that $\mathbf{X}_{ijk} = 1$ for some k and $\mathbf{X}_{ijl} = 0$ for $k \ne l$. Then the ℓ_2 estimation error is bounded by $O(A\ln n/(nm(\epsilon^2 \wedge \epsilon)))$, which matches (Acharya et al., 2023) up to logarithm factor.*

The corresponding minimax lower bounds are shown as follows.

Theorem 6. *Denote $\mathcal{P}_{\mathcal{X},p}$ as the set of all distributions supported on $\mathcal{X}_p = \{\mathbf{u} | \|\mathbf{u}\|_p \le D\}$, $\mathcal{M}_\epsilon$ as all mechanisms satisfying user-level ϵ-LDP. Then for $p \in [1, 2]$, with n users and m samples per user,*

$$\inf_{\hat{\mu}} \inf_{M\in\mathcal{M}_\epsilon} \sup_{p\in\mathcal{P}_{\mathcal{X},p}} \mathbb{E}\left[\|\hat{\mu} - \mu\|_2^2\right] \gtrsim \frac{D^2 d}{nm(\epsilon^2 \wedge \epsilon)}. \tag{5}$$

Theorem 7. *Denote $\mathcal{P}_{\mathcal{X},\infty}$ as the set of all distributions supported on $\mathcal{X}_\infty$, $\mathcal{M}_\epsilon$ as all mechanisms satisfying ϵ-LDP. Then with n users and m samples per user,*

$$\inf_{\hat{\mu}} \inf_{M\in\mathcal{M}_\epsilon} \sup_{p\in\mathcal{P}_{\mathcal{X},\infty}} \mathbb{E}\left[\|\hat{\mu} - \mu\|_2^2\right] \gtrsim \frac{D^2 d^2}{nm(\epsilon^2 \wedge \epsilon)}. \tag{6}$$

The proof of Theorem 6 and 7 are shown in Appendix B.2 and B.3, respectively. The upper bounds of MSE in Theorem 4 and Theorem 5 match the lower bounds equation 6 and equation 5. These results indicate that our methods for high dimensional mean estimation under user-level LDP are minimax optimal.

Remark 2. *Now we extend the analysis to unbounded support. If $\mathbb{E}[|X_k|^p] \le M_p$ for all $k = 1, \dots, d$, then with $n\epsilon^2 \ge c_1 d \ln m$ for some constant c_1,*

$$\mathbb{E}\left[\|\hat{\mu} - \mu\|_2^2\right] \lesssim M_p^{2/p}\left[\frac{d^2\ln m}{mn(\epsilon^2\wedge\epsilon)} \vee \left(\frac{d}{m^2 n(\epsilon^2\wedge\epsilon)}\right)^{1-1/p} + \frac{d}{mn}\right]. \tag{7}$$

Under a stronger condition $\mathbb{E}[\|\mathbf{X}\|_2^p] \le M_p < \infty$, the mean squared error can be bounded by

$$\mathbb{E}\left[\|\hat{\mu} - \mu\|_2^2\right] \lesssim M_p^{2/p}\left[\frac{d\ln m}{mn(\epsilon^2\wedge\epsilon)} \vee \left(\frac{d}{m^2 n(\epsilon^2\wedge\epsilon)}\right)^{1-1/p} + \frac{1}{mn}\right], \tag{8}$$

which is smaller than the rate under coordinate-wise p-th order bounded moment by a factor d. The detailed arguments can be found in Appendix B.4.

4 STOCHASTIC OPTIMIZATION

The goal is to solve the following stochastic optimization problem. Define the loss function as $L(\theta) := \mathbb{E}[l(\mathbf{X}, \theta)]$, in which $\mathbf{X}$ is a random variable following distribution p. Given $\mathbf{X}_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, our goal is to find the minimizer $\theta^* = \min_{\theta \in \Theta} L(\theta)$. The estimator is designed as follows. Users are divided randomly into t_0 groups. We plan to update θ in t_0 steps. In the t-th step, we use one group of users to get an estimate of $\nabla L(\theta_t) = \mathbb{E}[\nabla l(\mathbf{X}, \theta_t)]$ using Algorithm 2, which includes the privacy mechanism. The result is denoted as $\mathbf{g}_t$, and the update rule of θ is: $\theta_{t+1} = \theta_t - \eta \mathbf{g}_t$, in which η is the learning rate.

Since Algorithm 2 satisfies ϵ-LDP at user-level, and each user is only used once, the whole algorithm with t_0 steps also satisfies ϵ-LDP.

These steps are summarized in Algorithm 3. In step 5, the MeanEst function refers to the multi-dimensional mean estimation method shown in Algorithm 2. Samples are privatized in this step. Therefore, Algorithm 3 satisfies user-level ϵ-LDP.

Algorithm 3 Stochastic optimization under user-level ϵ-LDP

Input: Dataset containing n users with m samples per user, i.e. $\mathbf{X}_{ij}, i = 1, \ldots, n, j = 1, \ldots, m$

Output: Estimated $\hat{\theta}$

1: Initialize θ_0;
2: Divide users into t_0 groups $S_0, \ldots, S_{T-1}$;
3: **for** $t = 0, 1, \ldots, t_0 - 1$ **do**
4: Calculate $\nabla l(\mathbf{X}_{ij}, \theta_t)$ for $i \in S_t$, $j = 1, \ldots, m$;
5: $\mathbf{g}_t = MeanEst(\{\nabla l(\mathbf{X}_{ij}, \theta_t) \mid i \in S_t, j \in [m]\})$;
6: $\theta_{t+1} = \theta_t - \eta \mathbf{g}_t$;
7: **end for**
8: **Return** $\hat{\theta} = \theta_{t_0}$

Now we provide a theoretical analysis, which is based on the following assumptions.

Assumption 1. *(a) $l(\mathbf{X}, \theta)$ is G-smooth; (b) For any θ, the gradient of l has bounded ℓ_2 norm with probability 1, i.e. $\|\nabla l(\mathbf{X}, \theta)\|_2 \leq D$; (c) L is γ-strong convex.*

The theoretical bound is shown in the following theorem.

Theorem 8. *With $\eta \leq 1/G$, the ℓ_2 error at t-th step can be bounded by*

$$\mathbb{E}\left[\|\theta_t - \theta^*\|_2\right] \leq \left(1 - \frac{1}{2}\eta\gamma\right)^t \|\theta_0 - \theta^*\|_2 + \frac{2D}{\gamma}\sqrt{\frac{Ct_0}{nm}\left(1 + \frac{d \ln n}{\epsilon^2 \wedge \epsilon}\right)}. \tag{9}$$

From equation 9, there exists two constants c_T and C_T, if $c_T \ln n \leq t_0 \leq C_T \ln n$, and $n(\epsilon^2 \wedge 1) \gtrsim d \ln n \ln m$, then the final estimate $\hat{\theta} = \theta_{t_0}$ satisfies $\mathbb{E}\left[\left\|\hat{\theta} - \theta^*\right\|_2\right] \lesssim D\sqrt{\frac{\ln n}{nm}\left(1 + \frac{d \ln n}{\epsilon^2 \wedge \epsilon}\right)}$.

The proof of Theorem 8 is provided in Appendix C. In (Duchi et al., 2013), it is shown that the bound for item-level case is $\tilde{O}(\sqrt{d/(n\epsilon^2)})$ for $\epsilon \leq 1/4$ with n samples. Therefore, with the same total number of samples, our bound matches the result in (Duchi et al., 2013).

5 NONPARAMETRIC CLASSIFICATION AND REGRESSION

Now we focus on nonparametric learning problems under user-level local DP. For nonparametric learning problems, the dataset contains n users with m samples per user, i.e. $\mathbf{X}_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$ and their label Y_{ij}. Following (Berrett & Butucea, 2019; Berrett et al., 2021), which focuses on item-level classification and regression problems, suppose that $\mathbf{X}$ is supported in $[0, 1]^d$, which is made for simplicity. It can be generalized to arbitrary bounded support. Denote $(\mathbf{X}, Y)$ as a test sample i.i.d to training samples, and the output of the classifier is $\hat{Y}$.

5.1 CLASSIFICATION

The risk is defined as $R = \mathrm{P}(\hat{Y} \neq Y)$. Define $\eta(\mathbf{x}) = \mathbb{E}[Y | \mathbf{X} = \mathbf{x}]$. Given the test sample at $\mathbf{x}$, the optimal classifier is $\hat{Y} = \mathrm{sign}(\eta(\mathbf{x}))$. The corresponding optimal risk, called Bayes risk, is $R^* = \mathrm{P}(\mathrm{sign}(\eta(\mathbf{X})) \neq Y) = \frac{1}{2}\mathbb{E}[1 - |\eta(\mathbf{X})|]$, η is unknown in practice. We have to learn η from the training data. Therefore, in reality, there is inevitably a gap between the risk of a practical classifier and the Bayes risk. Such gap is called excess risk $R - R^*$. To improve the efficiency, we propose a

method based on a transformation with Hadamard matrix (Hedayat & Wallis, 1978). We make some assumptions before stating our algorithm.

Assumption 2. *There exists constants C_a, C_b, f_L, such that: (a) For all $t > 0$, $P(|\eta(\mathbf{X})| < t) \leq C_a t^\gamma$; (b) For all $\mathbf{x}, \mathbf{x}' \in \mathcal{X} = [0,1]^d$, $|\eta(\mathbf{x}) - \eta(\mathbf{x}')| \leq C_b \|\mathbf{x} - \mathbf{x}'\|_2^\beta$; (c) $f(\mathbf{x}) \geq f_L$ for all $\mathbf{x} \in \mathcal{X}$.*

(a) is typically referred to as 'Tsybakov noise condition' (Chaudhuri & Dasgupta, 2014; Döring et al., 2017). (b) is the Hölder smoothness condition(Tsybakov, 2009). (c) is usually referred to as 'strong density assumption'(Döring et al., 2017; Gadat et al., 2016). The assumptions (a)-(c) are commonly used in nonparametric statistics and are the same as (Berrett & Butucea, 2019), except that we are now considering user-level LDP, while (Berrett & Butucea, 2019) is about item-level LDP.

Theorem 9. *Under Assumption 2, if $n(\epsilon^2 \wedge 1) \geq c_2(\ln m + \ln n)$ for some constant c_2, then there exists a classifier (the algorithm is shown in Appendix D.1), such that*

$$R - R^* \lesssim (mn(\epsilon^2 \wedge \epsilon))^{-\frac{\beta(1+\gamma)}{2(d+\beta)}} \ln^{1+\gamma} n + \left(\frac{nm}{\ln n}\right)^{-\frac{\beta(1+\gamma)}{2\beta+d}}. \tag{10}$$

The proof of Theorem 9 is shown in Appendix D.2. With large ϵ, equation 10 reduces to $(mn/\ln n)^{-2\beta/(2\beta+d)}$, which matches the non-private rate up to logarithm factor (Tsybakov, 2009). The minimax bound is shown in the following theorem.

Theorem 10. *Denote $\mathcal{P}_{cls}$ as the set of all distributions p of $\mathbf{X}$ and regression function η that satisfy Assumption 2, $\mathcal{M}_\epsilon$ as all mechanisms satisfying ϵ-LDP, then for small ϵ,*

$$\inf_{\hat{Y}} \inf_{M \in \mathcal{M}_\epsilon} \sup_{(p,\eta) \in \mathcal{P}_{cls}} (R - R^*) \gtrsim (nm\epsilon^2)^{-\frac{\beta(1+\gamma)}{2(d+\beta)}} + (mn)^{-\frac{\beta(1+\gamma)}{2\beta+d}}. \tag{11}$$

The proof of Theorem 10 is shown in Appendix D.3. The comparison of Theorem 9 and Theorem 10 show that for small ϵ, the upper bound and lower bound match up to a logarithmic factor. Moreover, recall (Berrett & Butucea, 2019), the minimax lower bound under item-level DP is $(N\epsilon^2)^{-\beta(1+\gamma)/(2(d+\beta))}$. If $N = nm$, this bound also matches equation 10, indicating that the user-level case is nearly as hard as the item-level one in asymptotic sense up to a logarithmic factor.

5.2 Regression

For regression problem, we use the ℓ_2 loss as the metric, i.e. $R = \mathbb{E}\left[(\hat{\eta}(\mathbf{X}) - \eta(\mathbf{X}))^2\right]$. The support is divided similarly to classification. The bounds on the convergence rate of nonparametric regression and the corresponding minimax rate are shown in the following two theorems, respectively.

Theorem 11. *Under Assumption 2(b) and (c), and assume that the noise is bounded, such that with probability 1, $|Y| < T$ for some T, if $n(\epsilon^2 \wedge 1) \geq 2c_2(\ln m + \ln n)$, in which c_2 is the same constant in Theorem 9, then there exists an algorithm (described in Appendix E.1), such that the risk of nonparametric regression is bounded by $R \lesssim \left(\frac{mn(\epsilon^2 \wedge \epsilon)}{\ln^2 n}\right)^{-\frac{\beta}{d+\beta}} + \left(\frac{mn}{\ln n}\right)^{-\frac{2\beta}{2\beta+d}}$.*

Theorem 12. *Denote $\mathcal{P}_{reg}$ as the set of all distributions p of $\mathbf{X}$ and regression function η that satisfy the same assumption as Theorem 11, $\mathcal{Q}_\epsilon$ as all mechanisms satisfying ϵ-LDP, then for small ϵ,*

$$\inf_{\hat{\eta}} \inf_{Q \in \mathcal{Q}_\epsilon} \sup_{(p,\eta) \in \mathcal{P}_{reg}} R \gtrsim (nm\epsilon^2)^{-\frac{\beta}{d+\beta}} + (mn)^{-\frac{2\beta}{2\beta+d}}. \tag{12}$$

The proof of Theorem 11 and 12 are shown in Appendix E.2 and E.3, respectively. Similar to the classification, it can be found that the upper and lower bounds match up to logarithm factors.

6 Conclusion

In this paper, we have conducted a theoretical study of various statistical problems under user-level local differential privacy, including mean estimation, stochastic optimization, nonparametric classification, and regression. For each problem, we have proposed algorithms and provided information-theoretic minimax lower bounds. The results show that for many statistical problems, with the same total sample sizes, the errors under user-level and item-level ϵ-LDP are nearly of the same order.

REFERENCES

Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 308–318, 2016.

Jayadev Acharya, Yuhan Liu, and Ziteng Sun. Discrete distribution estimation under user-level local differential privacy. In *International Conference on Artificial Intelligence and Statistics*, pp. 8561–8585. PMLR, 2023.

Sushant Agarwal, Gautam Kamath, Mahbod Majid, Argyris Mouzakis, Rose Silver, and Jonathan Ullman. Private mean estimation with person-level differential privacy. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2819–2880. SIAM, 2025.

Hilal Asi, Vitaly Feldman, Tomer Koren, and Kunal Talwar. Private stochastic convex optimization: Optimal rates in l1 geometry. In *International Conference on Machine Learning*, pp. 393–403. PMLR, 2021.

Hilal Asi, Vitaly Feldman, and Kunal Talwar. Optimal algorithms for mean estimation under local differential privacy. In *International Conference on Machine Learning*, pp. 1046–1056. PMLR, 2022.

Hilal Asi, Vitaly Feldman, Jelani Nelson, Huy Nguyen, and Kunal Talwar. Fast optimal locally private mean estimation via random projections. *Advances in Neural Information Processing Systems*, 36, 2024a.

Hilal Asi, Vitaly Feldman, Jelani Nelson, Huy L Nguyen, Samson Zhou, and Kunal Talwar. Private vector mean estimation in the shuffle model: Optimal rates require many messages. *arXiv preprint arXiv:2404.10201*, 2024b.

Spyridon Bakas, Xiaoxiao Li, Prashant Shah, and Holger R Roth. Federated learning in healthcare: From research to real-world deployment. *Annual Review of Biomedical Engineering*, 28, 2026.

Milad Bakhshizadeh, Arian Maleki, and Victor H de la Pena. Sharp concentration results for heavy-tailed distributions. *Information and Inference: A Journal of the IMA*, 12(3):iaad011, 2023.

Raef Bassily and Ziteng Sun. User-level private stochastic convex optimization with optimal rates. In *International Conference on Machine Learning*, pp. 1838–1851. PMLR, 2023.

Thomas Berrett and Cristina Butucea. Classification under local differential privacy. *arXiv preprint arXiv:1912.04629*, 2019.

Thomas B Berrett, László Györfi, and Harro Walk. Strongly universally consistent nonparametric regression and classification with privatised data. *Electronic Journal of Statistics*, 15:2430–2453, 2021.

Abhishek Bhowmick, John Duchi, Julien Freudiger, Gaurav Kapoor, and Ryan Rogers. Protection against reconstruction and its applications in private federated learning. *arXiv preprint arXiv:1812.00984*, 2018.

Mark Bun, Marco Gaboardi, Max Hopkins, Russell Impagliazzo, Rex Lei, Toniann Pitassi, Satchit Sivakumar, and Jessica Sorrell. Stability is stable: Connections between replicability, privacy, and adaptive generalization. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pp. 520–527, 2023.

Zachary Charles, Arun Ganesh, Ryan McKenna, H Brendan McMahan, Nicole Mitchell, Krishna Pillutla, and Keith Rush. Learning with user-level differential privacy under fixed compute budgets. In *2025 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pp. 901–920. IEEE, 2025.

Kamalika Chaudhuri and Sanjoy Dasgupta. Rates of convergence for nearest neighbor classification. *Advances in Neural Information Processing Systems*, 27, 2014.

Wei-Ning Chen, Peter Kairouz, and Ayfer Ozgur. Breaking the communication-privacy-accuracy trilemma. *Advances in Neural Information Processing Systems*, 33:3312–3324, 2020.

Wei-Ning Chen, Dan Song, Ayfer Ozgur, and Peter Kairouz. Privacy amplification via compression: Achieving the optimal privacy-accuracy-communication trade-off in distributed mean estimation. *Advances in Neural Information Processing Systems*, 36, 2023.

Paul Cuff and Lanqing Yu. Differential privacy as a mutual information constraint. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 43–54, 2016.

Maik Döring, László Györfi, and Harro Walk. Rate of convergence of k-nearest-neighbor classification rule. *The Journal of Machine Learning Research*, 18(1):8485–8500, 2017.

John C Duchi, Michael I Jordan, and Martin J Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th annual symposium on foundations of computer science*, pp. 429–438. IEEE, 2013.

John C Duchi, Michael I Jordan, and Martin J Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.

Vladimir Dvorkin, Ferdinando Fioretto, Pascal Van Hentenryck, Pierre Pinson, and Jalal Kazempour. Privacy-preserving convex optimization: When differential privacy meets stochastic programming. In *2025 IEEE 64th Conference on Decision and Control (CDC)*, pp. 8149–8156. IEEE, 2025.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pp. 265–284. Springer, 2006.

Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pp. 51–60. IEEE, 2010.

Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

Vitaly Feldman and Kunal Talwar. Lossless compression of efficient private local randomizers. In *International Conference on Machine Learning*, pp. 3208–3219. PMLR, 2021.

Vitaly Feldman, Tomer Koren, and Kunal Talwar. Private stochastic convex optimization: optimal rates in linear time. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 439–449, 2020.

Vitaly Feldman, Cristobal Guzman, and Santosh Vempala. Statistical query algorithms for mean vector estimation and stochastic convex optimization. *Mathematics of Operations Research*, 46(3): 912–945, 2021.

Sébastien Gadat, Thierry Klein, and Clément Marteau. Classification in general finite dimensional spaces with the k-nearest neighbor rule. *Annals of Statistics*, 2016.

Robin C Geyer, Tassilo Klein, and Moin Nabi. Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557*, 2017.

Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. User-level differentially private learning via correlated sampling. *Advances in Neural Information Processing Systems*, 34:20172–20184, 2021.

Badih Ghazi, Pritish Kamath, Ravi Kumar, Pasin Manurangsi, Raghu Meka, and Chiyuan Zhang. User-level differential privacy with few examples per user. *Advances in Neural Information Processing Systems*, 36, 2023.

A Hedayat and Walter Dennis Wallis. Hadamard matrices and their applications. *The annals of statistics*, pp. 1184–1238, 1978.

Samuel B Hopkins, Gautam Kamath, and Mahbod Majid. Efficient mean estimation with pure differential privacy via a sum-of-squares exponential mechanism. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1406–1417, 2022.

Ziyue Huang, Yuting Liang, and Ke Yi. Instance-optimal mean estimation under differential privacy. *Advances in Neural Information Processing Systems*, 34:25993–26004, 2021.

Gautam Kamath, Xingtu Liu, and Huanyu Zhang. Improved rates for differentially private stochastic convex optimization with heavy-tailed data. In *International Conference on Machine Learning*, pp. 10633–10660. PMLR, 2022.

Daniel Levy, Ziteng Sun, Kareem Amin, Satyen Kale, Alex Kulesza, Mehryar Mohri, and Ananda Theertha Suresh. Learning with user-level privacy. *Advances in Neural Information Processing Systems*, 34:12466–12479, 2021.

Mengchu Li, Thomas B Berrett, and Yi Yu. On robustness and local differential privacy. *The Annals of Statistics*, 51(2):717–737, 2023.

Daogao Liu and Hilal Asi. User-level differentially private stochastic convex optimization: Efficient algorithms with optimal rates. In *International Conference on Artificial Intelligence and Statistics*, pp. 4240–4248. PMLR, 2024.

Xiao-Yang Liu, Rongyi Zhu, Daochen Zha, Jiechao Gao, Shan Zhong, Matt White, and Meikang Qiu. Differentially private low-rank adaptation of large language model using federated learning. *ACM Transactions on Management Information Systems*, 16(2):1–24, 2025.

Xiyang Liu, Weihao Kong, Sham Kakade, and Sewoong Oh. Robust and differentially private mean estimation. *Advances in neural information processing systems*, 34:3887–3901, 2021.

Yurii Lyubarskii and Roman Vershynin. Uncertainty principles and vector quantization. *IEEE Transactions on Information Theory*, 56(7):3491–3501, 2010.

Yuan Qiu, Xiaokui Xiao, and Yin Yang. Dp-s4s: Accurate and scalable select-join-aggregate query processing with user-level differential privacy. *Proceedings of the ACM on Management of Data*, 4 (3 (SIGMOD):1–24, 2026.

Alexandre B Tsybakov. *Introduction to Nonparametric Estimation*. 2009.

Shay Vargaftik, Ran Ben-Basat, Amit Portnoy, Gal Mendelson, Yaniv Ben-Itzhak, and Michael Mitzenmacher. Drive: One-bit distributed mean estimation. *Advances in Neural Information Processing Systems*, 34:362–377, 2021.

Sajani Vithana, Viveck R Cadambe, Flavio P Calmon, and Haewon Jeong. Correlated privacy mechanisms for differentially private distributed mean estimation. In *2025 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pp. 590–614. IEEE, 2025.

Weina Wang, Lei Ying, and Junshan Zhang. On the relation between identifiability, differential privacy, and mutual-information privacy. *IEEE Transactions on Information Theory*, 62(9):5018–5029, 2016.

Antai Xie, Xiaoqiang Ren, Xinlei Yi, Tao Yang, and Xiaofan Wang. Compressed distributed stochastic nonconvex optimization with differential privacy. *arXiv preprint arXiv:2603.21640*, 2026.

Difei Xu, Meng Ding, Zihang Xiang, Jinhui Xu, and Di Wang. Beyond ordinary lipschitz constraints: Differentially private stochastic optimization with tsybakov noise condition. *arXiv preprint arXiv:2509.04668*, 2025.

Mengmeng Yang, Lingjuan Lyu, Jun Zhao, Tianqing Zhu, and Kwok-Yan Lam. Local differential privacy and its applications: A comprehensive survey. *arXiv preprint arXiv:2008.03686*, 2020.

Rao K Yarlagadda and John E Hershey. *Hadamard matrix analysis and synthesis: with applications to communications and signal/image processing*, volume 383. Springer Science & Business Media, 2012.