Distributed Mean Estimation for Multi-Message Shuffled Privacy

Antonious M. Girgis[1] **Suhas Diggavi**[1]

Abstract

In this paper, we study distributed mean estimation (DME) under privacy and communication constraints in the multi-message shuffle model. We propose communication-efficient algorithms for privately estimating the mean of bound ℓ_2-norm and ℓ_∞-norm norm vectors. Our algorithms are designed by giving unequal privacy at different resolutions of the vector (through binary expansion) and appropriately combining it with co-ordinate sampling. We show that our proposed algorithms achieve order-optimal privacy-communication-performance trade-offs.

1. Introduction

We consider distributed mean estimation (DME) problem, where a set of clients are connected to a (untrusted) server to estimate the average of the clients' data. DME has wide applications including federated learning (FL), in which the central server estimates the mean of the local updates at each round of the FedAvg (McMahan et al., 2017). However, DME faces two major challenges in the real world. (i) *Privacy:* the clients' data might contain sensitive information, and hence, each client wants to preserve privacy of her own local data. (ii) *Communication:* the connection between the server and clients might be over wireless/band-limited networks, and hence, the communication becomes a bottleneck for estimation. We focus on the shuffle model of differential privacy (DP), where the clients are connected to the server through a secure shuffler that randomly permutes the clients' responses before passing them to the server (Bittau et al., 2017; Erlingsson et al., 2019; Cheu et al., 2019).

We propose mechanisms for DME of bounded ℓ_∞-norm and ℓ_2-norm vectors that matches the lower bound. Furthermore, our proposed schemes can be applied in the local DP model and achieves order-optimal privacy-communication-accuracy trade-offs. Communication efficiency is obtained by appropriately sampling the co-ordinates and using finite number of bits through a binary expansion. Our core idea for privacy is to allocate unequal privacy for different resolutions of the real vector, obtained through a binary expansion of it. We allocate increasing privacy with the order of bits, *i.e.,* lower privacy for most significant bits (MSBs); this gives better performance in terms of mean squared error (MSE), as MSBs are more important. This, combined with careful accounting for the composition using RDP, gives our privacy guarantees and performance.

1.1. Contributions and Literature Context

Distributed mean estimation (DME) in the local model of DP is well-studied with a characterization of the optimal privacy-communication-utility trade-off (see (Chen et al., 2020; Girgis et al., 2021c; Asi et al., 2022) and reference therein). Our scheme when applied to the local DP model also achieves the optimal privacy-communication-utility trade-offs for LDP framework (see Theorems 4.1 and 5.1). However, LDP mechanisms suffer from high MSE comparing to the central DP mechanisms. To improve the performance of the LDP mechanism without a need for a trusted server, the shuffle model has been proposed (Bittau et al., 2017; Erlingsson et al., 2019; Cheu et al., 2019), where a secure shuffler randomly permutes the private messages of the clients before sending them to the untrusted server. This was further enhanced by using multi-message shufflers (MMS) in (Balle et al., 2020c; Ghazi et al., 2020b).

In this work, we establish the fundamental privacy-communication-performance trade-offs for computing *vector sum* in the multi-message shuffle (MMS) model, see Theorems 4.2 and 5.2. . For bounded ℓ_2-norm vectors, our proposed scheme achieves MSE $\mathcal{O}\left(\frac{d}{n^2\epsilon^2}\right)$ that requires only $\mathcal{O}\left(\min\{n\epsilon^2, d\log\left(\frac{n\epsilon^2}{d}\right)\}\right)$bits per client.

To put this in context, our result is order-wise better than the private vector summation result in (Cheu et al., 2022), which had communication $\mathcal{O}(d\sqrt{n})$ per client, where d is the vector dimension. The works of (Balle et al., 2019; 2020c; Ghazi et al., 2020b) focused on the *scalar* private real summation problem. For single-message shuffle model, Balle *et. al.* presented lower and matching upper bounds for scalar private real summation, showing that the MSE is order $\Theta\left(n^{1/3}\right)$. In (Cheu et al., 2019), the MMS mechanism is

[1] University of California, Los Angeles, CA, USA. Correspondence to: Antonious M. Girgis <amgirgis@ucla.edu>.

Workshop of Federated Learning and Analytics in Practice, colocated with 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. Copyright 2023 by the author(s).

optimal in MSE but needs $\mathcal{O}(\sqrt{n})$-bits per client. In (Balle et al., 2020c; Ghazi et al., 2020b), a MMS mechanism based on IKOS scheme (Ishai et al., 2006) was proposed for real summation in which each client needs to send only $\mathcal{O}(1)$ messages to the shuffler, each of size $\mathcal{O}(\log(n))$ bits. Our mechanism, when applied to the scalar case, achieves the optimal MSE with $\mathcal{O}\left(\log\left(n\epsilon^2\right)\right)$-bits per client that improves the communication cost in the high privacy regime (see Theorem 4.2).

2. Problem Formulation

We consider a set of n clients. Each client has a vector $\mathbf{x}_i \in \mathcal{X}$ for $i \in [n]$, where $\mathcal{X} \subset \mathbb{R}^d$ denotes a bounded subset of all possible inputs. For example, $\mathcal{X} \triangleq \mathbb{B}_2^d(r_2)$ denotes the d dimensional ball with radius r_2, i.e., each vector $\mathbf{x}_i$ satisfies $\|\mathbf{x}_i\|_2 \leq r_2$ for $i \in [n]$. The clients are connected to an untrustworthy server that wants to estimate the mean $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$. In this paper, we consider two distributed privacy models.

Local DP Model In the local DP model, our goal is to design two mechanisms: (i) A client-side mechanism $\mathcal{R} : \mathcal{X} \to \mathcal{Y}$ that generates a randomized output $\mathbf{y}_i \in \mathcal{Y}$. The local mechanism $\mathcal{R}$ satisfies privacy and communication constraints as follows: The output $\mathbf{y}_i = \mathcal{R}\left(\mathbf{x}_i\right)$ can be represented using only b-bits. Furthermore, the mechanism $\mathcal{R}$ satisfies ϵ_0-LDP. Each client sends the output $\mathbf{y}_i$ directly to the server. (ii) Server aggregator $\mathcal{A} : \mathcal{Y}^n \to \mathbb{R}^d$ to estimate the mean $\hat{\mathbf{x}} = \mathcal{A}\left(\mathbf{y}_1, \ldots, \mathbf{y}_n\right)$ such that the estimated mean $\hat{\mathbf{x}}$ is an unbiased estimate of the true mean $\overline{\mathbf{x}}$.

Shuffle Model The shuffle model consists of three parameters $(\mathcal{R}, \mathcal{S}, \mathcal{A})$: (i)*Encode:* a set of L local mechanisms $\mathcal{R}^{(k)} : \mathcal{X} \to \mathcal{Y}, k \in [L]$, each similar to the local DP model. Each client sends the L outputs $\mathbf{y}_i^{(k)}, k \in [L]$, where $\mathbf{y}_i^{(k)} \in \mathcal{Y}$, to the secure shufflers. (ii) *Shuffle:* a single secure shuffler $\mathcal{S}_k : \mathcal{Y}^n \to \mathcal{Y}^n$ receives n outputs $\mathbf{y}_i^{(k)}, i \in [n]$ and generates a random permutation $\pi^{(k)}$ of the received messages. The multi-message shuffle is a parallel set of L single-message shufflers $\{\mathcal{S}_k\}$. (iii) *Analyze:* the server receives the L shufflers' outputs and applies the aggregator $\mathcal{A} : \mathcal{Y}^{nL} \to \mathbb{R}^d$ to estimate the mean $\hat{\mathbf{x}} = \mathcal{A}\left(\mathbf{y}_{\pi^{(k)}(1)}, \ldots, \mathbf{y}_{\pi^{(k)}(n)}, k \in [L]\right)$. The shuffle model is (ϵ, δ)-DP if the view of the output of the L shufflers satisfies (ϵ, δ)-DP.

In the two privacy models, the performance of the estimator $\hat{\mathbf{x}}$ is measured by the expected loss:

$$\mathsf{MSE} = \sup_{\{\mathbf{x}_i \in \mathcal{X}\}} \mathbb{E}\left[\|\hat{\mathbf{x}} - \overline{\mathbf{x}}\|_2^2\right], \tag{1}$$

where the expectation is taken over the randomness of the private mechanisms. Our goal is to design communication-efficient and private schemes to generate an unbiased estimate of $\overline{x}$ while minimizing the expected loss (1). We start by the DME of binary vectors, where $\mathcal{X} \triangleq \{0,1\}^d$. Then, we study the DME for bounded ℓ_∞-norm *i.e.,* $\|\mathbf{x}_i\|_\infty \leq r_\infty$ and bounded ℓ_2-norm vectors, where $\|\mathbf{x}_i\|_2 \leq r_2$.

Remark 2.1 (parallel shufflers vs single shuffler). Observe that we describe the multi-message shuffle model using L independent shufflers, where each shuffler receives a single message from each client. We can also represent the multi-message shuffle model with a single shuffler that receives the total nL messages from all clients by indexing the messages of each client with a slight increase of the communication cost, see (Balle et al., 2020c) for more details.

3. Binary vectors

In this section, we consider binary vectors: $\mathbf{b}_i \in \{0,1\}^d$. The server wants to estimate the mean $\overline{\mathbf{b}} = \frac{1}{n}\sum_{i=1}^n \mathbf{b}_i$. This problem is a generalization to the scalar binary summation problem (Cheu et al., 2019). A straightforward solution is to apply the scalar mechanism in (Cheu et al., 2019) per coordinate that requires d bits per client. Our private mechanisms require $\mathcal{O}\left(\min\{\epsilon_0, d\}\right)$ and $\mathcal{O}\left(\min\{n\epsilon^2, d\}\right)$ bits per client in the local DP and shuffle models, respectively.

The client-side mechanism is presented in Algorithm 1, where the parameter s determines the communication budget for each client and the parameter p determines the total privacy budget (see Theorem 3.1). For given $s \in \{1, \ldots, d\}$, each client splits the binary vector $\mathbf{b}_i$ into s sub-vectors, each with dimension $a = \lceil \frac{d}{s} \rceil$. Then, the client chooses uniformly at random one coordinate from each sub-vector and privatizes its bit using the binary randomized response (2RR) Algorithm 1 in the full version (Girgis & Diggavi, 2023). Observe that the output of Algorithm 1 can be represented as a sparse d-dimensional vector with only s non-zero bits.

When $s = d$, then each client applies the 2RR mechanism on each coordinate separately. On the other hand, when $s = 1$, the client chooses uniformly at random one coordinate and applies the 2RR mechanism. Thus, we get trade-offs between privacy-communication and accuracy. The server aggregator $\mathcal{A}^{\text{Bin}}$ is simply aggregating the received randomized bits. For completeness, we present the aggregator $\mathcal{A}^{\text{Bin}}$ in Algorithm 7 in the full version (Girgis & Diggavi, 2023)

Below, we state the bound on the MSE of the proposed mechanisms in the local DP and shuffle models. The proofs are deferred to the full version in (Girgis & Diggavi, 2023). Furthermore, we present RDP guarantees of our mechanisms for both local DP and shuffle models in the detailed proofs in the full version (Girgis & Diggavi, 2023).

Theorem 3.1 (Local DP model)**.** *The output of the local mechanism $\mathcal{R}_{p,s}^{Bin}$ can be represented us-*

Algorithm 1 : Local Randomizer $\mathcal{R}_{p,s}^{\text{Bin}}$

```
Public parameter: Privacy parameter $p$, and communication budget $s$.
Input: $\mathbf{b}_i \in \{0,1\}^d$.
If $\frac{d}{s}$ is not integer, add $(s\lceil \frac{d}{s} \rceil - d)$ dummy zeros to the binary vector $\mathbf{b}$. Let $a \leftarrow \frac{d}{s}$.
for $j \in [s]$ do
    Choose uniformly at random one coordinate $z_{ij} \leftarrow \mathsf{Unif}\left(\{(j-1)a, \ldots, ja\}\right)$.
    $y_{ij} \leftarrow a\mathcal{R}_p^{2RR}\left(\mathbf{b}_i[z_{ij}]\right)$
end for
Return: The client sends $s$ messages $\mathcal{Y}_i \leftarrow \{(a_{i1}, y_{i1}), \ldots, (a_{is}, y_{is})\}$.
```

ing $s\left(\log\left(\lceil d/s \rceil\right)+1\right)$*-bits. By choosing* $p = \frac{1}{2}\left(1-\sqrt{\frac{\epsilon_0^2/s^2}{\epsilon_0^2/s^2+4}}\right)$, *the mechanism* $\mathcal{R}_{p,s}^{Bin}$ *satisfies* ϵ_0*-LDP. Let* $\hat{\mathbf{b}}$ *be the output of the analyzer* $\mathcal{A}^{Bin}$. *The estimator* $\hat{\mathbf{b}}$ *is an unbiased estimate of* $\overline{\mathbf{b}}$ *with MSE:*

$$\mathsf{MSE}_{ldp}^{Bin} = \mathcal{O}\left(\frac{d^2}{n}\max\left\{\frac{1}{s}, \frac{s}{\epsilon_0^2}\right\}\right). \tag{2}$$

Theorem 3.1 shows that each client needs to send $s = \min\{\lceil \epsilon_0 \rceil, d\}$ communication bits to achieve MSE $\mathcal{O}\left(\frac{d^2}{n\min\{\epsilon_0, \epsilon_0^2\}}\right)$. Now, we move to the shuffle model, where we assume there exists s shufflers. The j-th shuffler randomly permutes the set of messages $\{(a_{ij}, y_{ij}) : i \in [n]\}$ from the n clients.

Theorem 3.2 (MMS model)**.** *The output of the local mechanism* $\mathcal{R}_{p,s}^{Bin}$ *can be represented using* $s\left(\log\left(\lceil d/s \rceil\right)+1\right)$ *bits. For every* $n \in \mathbb{N}$, $\epsilon \leq s$, *and* $\delta \in (0,1)$, *shuffling the outputs of* n *mechanisms* $\mathcal{R}_{p,s}^{Bin}$ *satisfies* (ϵ, δ)*-DP by choosing* $p = \frac{1}{2}\left(1-\sqrt{\frac{v^2}{v^2+4}}\right)$, *where* $v^2 = \frac{n\epsilon^2}{4s\log(1/\delta)}$. *Let* $\hat{\mathbf{b}}$ *be the output of the analyzer* $\mathcal{A}^{Bin}$. *The estimator* $\hat{\mathbf{b}}$ *is an unbiased estimate of* $\overline{\mathbf{b}}$ *with MSE:*

$$\mathsf{MSE}_{shuffle}^{Bin} = \mathcal{O}\left(\frac{d^2}{n^2}\max\left\{n\left(\frac{1}{s}-\frac{1}{d}\right), \frac{\log(1/\delta)}{\epsilon^2}\right\}\right). \tag{3}$$

Theorem 3.2 shows that each client requires to send $s = \mathcal{O}\left(\min\{n\epsilon^2, d\}\right)$ communication bits such that the error in the shuffle model is bounded by $\mathcal{O}\left(\frac{d^2}{n^2\epsilon^2}\right)$ that matches the MSE of central differential privacy mechanisms. For the scalar case when $d=1$, our results in Theorem 3.2 match the optimal MSE as in (Cheu et al., 2019).

4. Bounded ℓ_∞-norm vectors

In this section, we consider the bounded ℓ_∞-norm, where the ith client has a vector $\mathbf{x}_i$ such that $\|\mathbf{x}_i\|_\infty \leq r_\infty$ for $i \in [n]$. For ease of operation, we will scale each vector such that each coordinate becomes bounded in range $[0,1]$, and then re-scale it at the server-side. Let $\mathbf{z}_i = \frac{\mathbf{x}_i + r_\infty}{2r_\infty}$, where the operations are done coordinate-wise. Thus, we have that $\mathbf{z}_i[j] \in [0,1]$ for all $j \in [d]$ and $i \in [n]$, where $\mathbf{z}_i[j]$ denotes the jth coordinate of the vector $\mathbf{z}_i$. Observe that the vector $\mathbf{z}_i$ can be decomposed into a weighted summation of binary vectors $\mathbf{b}_i^{(k)} \in \{0,1\}^d, \forall k \geq 1$ as:

$$\mathbf{z}_i = \sum_{k=1}^{\infty} \mathbf{b}_i^{(k)} 2^{-k}, \tag{4}$$

where recursively, $\mathbf{b}_i^{(k)} = \lfloor 2^k\left(\mathbf{z}_i - \mathbf{z}_i^{(k-1)}\right)\rfloor, k \geq 1$ for $\mathbf{z}_i^{(0)} = \mathbf{0}$ and $\mathbf{z}_i^{(k)} = \sum_{l=1}^{k} \mathbf{b}_i^{(l)} 2^{-l}$.

To make our mechanism communication efficient, each client approximates the vector $\mathbf{z}_i$ by using the first m binary vectors $\{\mathbf{b}_i^{(k)} : 1 \leq k \leq m\}$. Note that the first m binary vectors give the best approximation to the real vector $\mathbf{z}_i$ with error $\|\mathbf{z}_i - \mathbf{z}_i^{(m)}\|_2^2 \leq d/4^m$. However, this mechanism creates a biased estimate of $\mathbf{z}_i$. Hence, to design an unbiased mechanism, the client approximates the vector $\mathbf{z}_i$ using the first $m-1$ binary vectors $\{\mathbf{b}_i^{(k)} : 1 \leq k \leq m-1\}$ of the binary representation above and the last binary vector $(\mathbf{u}_i)$ is reserved for unbiasness as follows:

$$\mathbf{u}_i[j] = \mathsf{Bern}\left(2^{m-1}(\mathbf{z}_i[j] - \mathbf{z}_i^{(m-1)}[j])\right), \tag{5}$$

where $\mathbf{z}_i^{(m-1)} = \sum_{k=1}^{m-1} \mathbf{b}_i^{(k)} 2^{-k}$ and $\mathsf{Bern}(p)$ denotes Bernoulli random variable with bias p. For completeness, we prove some properties of this quantization scheme in the full version (Girgis & Diggavi, 2023). Then, we estimate the mean of binary vectors $\{\mathbf{b}_i^{(k)} \in \{0,1\}^d : i \in [n]\}$ using Algorithm 1 with different privacy guarantees for each level $k \in [m]$, where we allocate lower privacy (higher privacy parameter ν_k) for the most significant bits (MSBs) (lower k) in order to get better performance in terms of MSE.

The private DME mechanism is given in Algorithm 2, where v controls the total privacy of the mechanism. There are two communication parameters: m controls the number of levels for quantization and s controls the number of dimensions used to represent each binary vector. In Theorems 4.1 and 4.2, we present how the privacy and communication parameters v, m, s affects the accuracy of the mechanism. The server aggregator $\mathcal{A}^{\ell_\infty}$ estimates the mean of each binary vectors $\{b_i^{(k)}\}$ and decodes the messages to generate an estimate to the true mean $\overline{\mathbf{z}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i$. Then, the server scales the vector $\overline{\mathbf{z}}$ to generate an unbiased estimate of the mean $\overline{\mathbf{x}}$. The server-side is presented in Algorithm 3 in the full version (Girgis & Diggavi, 2023). We prove the bound on the MSE of the proposed mechanisms in the local DP and MMS models in the following theorems. We defer the proofs to the full version (Girgis & Diggavi, 2023).

Algorithm 2 : Local Randomizer $\mathcal{R}_{v,m,s}^{\ell_\infty}$

```
**Public parameter:** Privacy budget $v$, communication levels $m$, and communication coordinates per level $s$.
**Input:** $\mathbf{x}_i \in \mathbb{B}_\infty^d(r_\infty)$.
$\mathbf{z}_i \leftarrow (\mathbf{x}_i + r_\infty)/2r_\infty$
$\mathbf{z}_i^{(0)} \leftarrow 0$
**for** $k = 1, \ldots, m-1$ **do**
    $\mathbf{b}_i^{(k)} \leftarrow \lfloor 2^k(\mathbf{z}_i - \mathbf{z}_i^{(k-1)}) \rfloor$
    $v_k \leftarrow \frac{4^{\frac{-k}{3}}}{\left(\sum_{l=1}^{m-1} 4^{\frac{-l}{3}} + 4^{\frac{-m+1}{3}}\right)} v$
    $p_k \leftarrow \frac{1}{2}\left(1 - \sqrt{\frac{v_k^2/s^2}{v_k^2/s^2+4}}\right)$
    $\mathcal{Y}_i^{(k)} \leftarrow \mathcal{R}_{p_k,s}^{\text{Bin}}(\mathbf{b}_i^{(k)})$
    $\mathbf{z}_i^{(k)} \leftarrow \mathbf{z}_i^{(k-1)} + \mathbf{b}_i^{(k)} 2^{-k}$
**end for**
Sample $\mathbf{u}_i \leftarrow \mathsf{Bern}\left(2^{m-1}\left(\mathbf{z}_i - \mathbf{z}_i^{(m-1)}\right)\right)$
$v_m \leftarrow \frac{4^{\frac{-m+1}{3}}}{\left(\sum_{l=1}^{m-1} 4^{\frac{-l}{3}} + 4^{\frac{-m+1}{3}}\right)} v$
$p_m \leftarrow \frac{1}{2}\left(1 - \sqrt{\frac{v_m^2/s^2}{v_m^2/s^2+4}}\right)$
$\mathcal{Y}_i^{(m)} \leftarrow \mathcal{R}_{p_m,s}^{\text{Bin}}(\mathbf{u}_i)$
**Return:** The client sends $\mathcal{Y}_i \leftarrow \left\{\mathcal{Y}_i^{(1)}, \ldots, \mathcal{Y}_i^{(m)}\right\}$.
```

Theorem 4.1 (Local DP model)**.** *The output of the local mechanism $\mathcal{R}_{v,m,s}^{\ell_\infty}$ can be represented using $ms\left(\log\left(\lceil d/s \rceil\right) + 1\right)$ bits. By choosing $v = \epsilon_0$, the mechanism $\mathcal{R}_{v,m,s}^{\ell_\infty}$ satisfies ϵ_0-LDP. Let $\hat{\mathbf{x}}$ be the output of the analyzer $\mathcal{A}^{\ell_\infty}$. The estimator $\hat{\mathbf{x}}$ is an unbiased estimate of $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ with bounded MSE:*

$$\mathsf{MSE}_{LDP}^{\ell_\infty} = \mathcal{O}\left(\frac{r_\infty^2 d^2}{n} \max\left\{\frac{1}{d4^m}, \frac{1}{s}, \frac{s}{\epsilon_0^2}\right\}\right). \tag{6}$$

Theorem 4.1 shows that each client needs to set $m = 1$ and $s = \lceil \epsilon_0 \rceil$ communication bits to achieve MSE $\mathcal{O}\left(\frac{d^2}{n\min\{\epsilon_0, \epsilon_0^2\}}\right)$ when $\epsilon_0 \leq d$.

Theorem 4.2 (MMS model)**.** *The output of the local mechanism $\mathcal{R}_{v,m,s}^{\ell_\infty}$ can be represented using $ms\left(\log\left(\lceil d/s \rceil\right) + 1\right)$ bits. For every $n \in \mathbb{N}$, $\epsilon \leq ms$, and $\delta \in (0,1)$, the shuffling the outputs of n mechanisms $\mathcal{R}_{v,m,s}^{\ell_\infty}$ satisfies (ϵ, δ)-DP by choosing $v^2 = \frac{sn\epsilon^2}{4\log(1/\delta)}$. Let $\hat{\mathbf{x}}$ be the output of the analyzer $\mathcal{A}^{\ell_\infty}$. The estimator $\hat{\mathbf{x}}$ is an unbiased estimate of $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ with bounded MSE:*

$$\mathsf{MSE}_{\textit{shuffle}}^{\ell_\infty} = \mathcal{O}\left(\frac{r_\infty^2 d^2}{n^2} \max\left\{\frac{n}{d4^m}, n\left(\frac{1}{s} - \frac{1}{d}\right), \frac{\log(1/\delta)}{\epsilon^2}\right\}\right). \tag{7}$$

In Theorem 4.2, by setting $m = \lceil \log\left(n\epsilon^2/d\right) \rceil$ and $s = \mathcal{O}\left(\min\{n\epsilon^2, d\}\right)$, MSE is bounded by $\mathcal{O}\left(\frac{d^2}{n^2\epsilon^2}\right)$, which matches MSE of central differential privacy mechanisms.

5. Bounded ℓ_2-norm Vectors

For bounded ℓ_2-norm, i.e., $\mathbf{x}_i$ such that $\|\mathbf{x}_i\|_2 \leq r_2$ for $i \in [n]$, we first use the random rotation proposed in (Suresh et al., 2017) to bound the ℓ_∞-norm of the vector with radius $r_\infty = \mathcal{O}\left(\frac{r_2}{\sqrt{d}}\right)$. Then, we apply the bounded ℓ_∞-norm algorithm in Section 4. The complete algorithms are presented in the full version (Girgis & Diggavi, 2023).

Theorem 5.1 (Local DP model)**.** *The output of the local mechanism $\mathcal{R}_{v,m,s}^{\ell_2}$ can be represented using $ms\left(\log\left(\lceil d/s \rceil\right) + 1\right)$ bits. By choosing $v = \epsilon_0$, the mechanism $\mathcal{R}_{v,m,s}^{\ell_2}$ satisfies ϵ_0-LDP. Let $\hat{\mathbf{x}}$ be the output of the analyzer $\mathcal{A}^{\ell_2}$. With probability at least $1 - \beta$, the estimator $\hat{\mathbf{x}}$ is an unbiased estimate of $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ with MSE:*

$$\mathsf{MSE}_{LDP}^{\ell_2} = \tilde{\mathcal{O}}\left(\frac{r_2^2}{n} \max\left\{\frac{1}{4^m}, \frac{d}{s}, \frac{ds}{\epsilon_0^2}\right\}\right), \tag{8}$$

where $\tilde{\mathcal{O}}$ hides $\log(nd)$ factor.

Theorem 5.2 (MMS model)**.** *The output of the local mechanism $\mathcal{R}_{v,m,s}^{\ell_2}$ can be represented using $ms\left(\log\left(\lceil d/s \rceil\right) + 1\right)$ bits. For every $n \in \mathbb{N}$, $\epsilon \leq ms$, and $\delta \in (0,1)$, the shuffling the outputs of n mechanisms $\mathcal{R}_{v,m,s}^{\ell_2}$ satisfies (ϵ, δ)-DP by choosing $v^2 = \frac{n\epsilon^2}{s\log(1/\delta)}$. Let $\hat{\mathbf{x}}$ be the output of the analyzer $\mathcal{A}^{\ell_2}$. With probability at least $1 - \beta$, the estimator $\hat{\mathbf{x}}$ is an unbiased estimate of $\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ with MSE:*

$$\mathsf{MSE}_{\textit{MMS}}^{\ell_2} = \tilde{\mathcal{O}}\left(r_2^2 \max\left\{\frac{1}{n4^m}, \frac{1}{n}\left(\frac{d}{s} - 1\right), \frac{d\log(1/\delta)}{n^2\epsilon^2}\right\}\right), \tag{9}$$

where $\tilde{\mathcal{O}}$ hides $\log(nd)$ factor.

Remark 5.3 (Kashin's represention). Observe that the MSE in (9) has a factor of $(\log(nd))$ due to the random rotation matrix. We can remove this factor by using the Kashin's representation (Kashin, 1977) to transform the bounded ℓ_2-norm vector into a bounded ℓ_∞-norm vector (see e.g., (Lyubarskii & Vershynin, 2010; Caldas et al., 2018; Chen et al., 2020))

Next we present a lower bound for DME under privacy and communication constraints, which can be derived using results from (Chen et al., 2022) and (Bun et al., 2014).

Theorem 5.4 (Lower Bound For central DP model)**.** *Let $n, d \in \mathbb{N}$, $\epsilon > 0$, $r_2 \geq 1$, and $\delta = o(\frac{1}{n})$. For any $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{B}_2^d(r_2)$, the MSE is bounded below by:*

$$\mathsf{MSE}_{\textit{central}}^{\ell_2} = \Omega\left(r_2^2 \max\left\{\frac{d}{n^2\epsilon^2}, \frac{1}{n4^{b/d}}\right\}\right) \tag{10}$$

for any unbiased algorithm $\mathcal{M}$ that is (ϵ, δ)-DP with $b > d$-bits of communication per client. Furthermore, when $b < d$

bits per client, the MSE is bounded below by:

$$\mathsf{MSE}_{central}^{\ell_2} = \Omega\left(r_2^2 d \max\left\{\frac{1}{n^2\epsilon^2}, \frac{1}{nb}\right\}\right). \qquad (11)$$

Remark 5.5. (Optimality of our mechanism) When the communication budget $b > d$, we can see that our MSE in Theorem 5.2 matches the lower bound in Theorem 5.4 (up to logarithmic factor) by choosing $s = d$ and $m = b/d$. Furthermore, when the communication budget $b < d$, our algorithm achieve the lower bound by choosing $s = b$ and $m = 1$. Thus, our algorithm for MMS is order optimal for all privacy-communication regimes.

Acknowledgements

This work was supported in part by NSF grants 2139304, 2007714 and by Meta faculty research award.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pp. 308–318, 2016.

Agarwal, N., Suresh, A. T., Yu, F. X. X., Kumar, S., and McMahan, B. cpsgd: Communication-efficient and differentially-private distributed sgd. In *Advances in Neural Information Processing Systems*, pp. 7564–7575, 2018.

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

Balle, B., Bell, J., Gascón, A., and Nissim, K. The privacy blanket of the shuffle model. In *Advances in Cryptology–CRYPTO 2019: 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18–22, 2019, Proceedings, Part II 39*, pp. 638–667. Springer, 2019.

Balle, B., Barthe, G., Gaboardi, M., Hsu, J., and Sato, T. Hypothesis testing interpretations and renyi differential privacy. In Chiappa, S. and Calandra, R. (eds.), *International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 108 of *Proceedings of Machine Learning Research*, pp. 2496–2506. PMLR, 2020a.

Balle, B., Bell, J., Gascón, A., and Nissim, K. Private summation in the multi-message shuffle model. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, CCS '20, pp. 657–676, 2020b.

Balle, B., Bell, J., Gascón, A., and Nissim, K. Private summation in the multi-message shuffle model. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, pp. 657–676, 2020c.

Bell, J. H., Bonawitz, K. A., Gascón, A., Lepoint, T., and Raykova, M. Secure single-server aggregation with (poly)logarithmic overhead. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp. 1253–1269, 2020.

Bhowmick, A., Duchi, J., Freudiger, J., Kapoor, G., and Rogers, R. Protection against reconstruction and its applications in private federated learning. *arXiv preprint arXiv:1812.00984*, 2018.

Bittau, A., Erlingsson, Ú., Maniatis, P., Mironov, I., Raghunathan, A., Lie, D., Rudominer, M., Kode, U., Tinnes, J., and Seefeld, B. Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th symposium on operating systems principles*, pp. 441–459, 2017.

Bun, M., Ullman, J., and Vadhan, S. Fingerprinting codes and the price of approximate differential privacy. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pp. 1–10, 2014.

Caldas, S., Konečny, J., McMahan, H. B., and Talwalkar, A. Expanding the reach of federated learning by reducing client resource requirements. *arXiv preprint arXiv:1812.07210*, 2018.

Canonne, C. L., Kamath, G., and Steinke, T. The discrete gaussian for differential privacy. In *Advances in Neural Information Processing Systems NeurIPS*, 2020.

Chaudhuri, K., Guo, C., and Rabbat, M. Privacy-aware compression for federated data analysis. In *Uncertainty in Artificial Intelligence*, pp. 296–306. PMLR, 2022.

Chen, W.-N., Kairouz, P., and Ozgur, A. Breaking the communication-privacy-accuracy trilemma. *Advances in Neural Information Processing Systems*, 33:3312–3324, 2020.

Chen, W.-N., Choo, C. A. C., Kairouz, P., and Suresh, A. T. The fundamental price of secure aggregation in differentially private federated learning. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 3056–3089, 17–23 Jul 2022.

Cheu, A., Smith, A., Ullman, J., Zeber, D., and Zhilyaev, M. Distributed differential privacy via shuffling. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 375–403. Springer, 2019.

Cheu, A., Joseph, M., Mao, J., and Peng, B. Shuffle private stochastic convex optimization. In *International Conference on Learning Representations (ICLR)*, 2022.

Ding, B., Kulkarni, J., and Yekhanin, S. Collecting telemetry data privately. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 3574–3583, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pp. 429–438. IEEE, 2013.

Dwork, C. and Roth, A. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. D. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference (TCC)*, pp. 265–284, 2006.

Erlingsson, Ú., Pihur, V., and Korolova, A. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pp. 1054–1067, 2014.

Erlingsson, Ú., Feldman, V., Mironov, I., Raghunathan, A., Talwar, K., and Thakurta, A. Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2468–2479. SIAM, 2019.

Erlingsson, Ú., Feldman, V., Mironov, I., Raghunathan, A., Song, S., Talwar, K., and Thakurta, A. Encode, shuffle, analyze privacy revisited: Formalizations and empirical evaluation. *CoRR*, abs/2001.03618, 2020.

Feldman, V., McMillan, A., and Talwar, K. Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling. In *2022 IEEE 62nd Annual Symposium on Foundations of Computer Science*. IEEE, 2022.

Feldman, V., McMillan, A., and Talwar, K. Stronger privacy amplification by shuffling for Renyi and approximate differential privacy. In *Proceedings of ACM-SIAM Symposium on Discrete Algorithms, SODA*, pp. 4966–4981. SIAM, 2023.

Ghazi, B., Golowich, N., Kumar, R., Manurangsi, P., Pagh, R., and Velingker, A. Pure differentially private summation from anonymous messages. In *1st Conference on Information-Theoretic Cryptography*, 2020a.

Ghazi, B., Kumar, R., Manurangsi, P., and Pagh, R. Private counting from anonymous messages: Near-optimal accuracy with vanishing communication overhead. In *International Conference on Machine Learning*, pp. 3505–3514. PMLR, 2020b.

Ghazi, B., Golowich, N., Kumar, R., Pagh, R., and Velingker, A. On the power of multiple anonymous messages: Frequency estimation and selection in the shuffle model of differential privacy. In *Advances in Cryptology - EUROCRYPT 2021 - Theory and Applications of Cryptographic Techniques*, volume 12698, pp. 463–488, 2021a.

Ghazi, B., Kumar, R., Manurangsi, P., Pagh, R., and Sinha, A. Differentially private aggregation in the shuffle model: Almost central accuracy in almost a single message. In *International Conference on Machine Learning*, pp. 3692–3701. PMLR, 2021b.

Girgis, A., Data, D., and Diggavi, S. Renyi differential privacy of the subsampled shuffle model in distributed learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 34:29181–29192, 2021a.

Girgis, A. M. and Diggavi, S. Multi-message shuffled privacy in federated learning. *arXiv preprint arXiv:2302.11152*, 2023.

Girgis, A. M., Data, D., and Diggavi, S. Differentially private federated learning with shuffling and client self-sampling. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pp. 338–343. IEEE, 2021b.

Girgis, A. M., Data, D., Diggavi, S., Kairouz, P., and Suresh, A. T. Shuffled model of differential privacy in federated learning. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 2521–2529. PMLR, 2021c.

Girgis, A. M., Data, D., Diggavi, S., Kairouz, P., and Suresh, A. T. Shuffled model of federated learning: Privacy, accuracy and communication trade-offs. *IEEE Journal on Selected Areas in Information Theory*, 2(1):464–478, 2021d.

Girgis, A. M., Data, D., Diggavi, S., Kairouz, P., and Suresh, A. T. Shuffled model of federated learning: Privacy, accuracy and communication trade-offs. *IEEE journal on selected areas in information theory*, 2(1):464–478, 2021e.

Girgis, A. M., Data, D., Diggavi, S., Suresh, A. T., and Kairouz, P. On the renyi differential privacy of the shuffle model. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pp. 2321–2341, 2021f.

Guo, C., Chaudhuri, K., Stock, P., and Rabbat, M. The interpolated mvu mechanism for communication-efficient private federated learning. *arXiv preprint arXiv:2211.03942*, 2022.

Ishai, Y., Kushilevitz, E., Ostrovsky, R., and Sahai, A. Cryptography from anonymity. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pp. 239–248. IEEE, 2006.

Kairouz, P., Bonawitz, K., and Ramage, D. Discrete distribution estimation under local privacy. In *International Conference on Machine Learning, ICML*, pp. 2436–2444, 2016.

Kairouz, P., Liu, Z., and Steinke, T. The distributed discrete gaussian mechanism for federated learning with secure aggregation. In *Proceedings International Conference on Machine Learning, ICML*, volume 139, pp. 5201–5212, 2021.

Kashin, B. S. Diameters of some finite-dimensional sets and classes of smooth functions. *Math. USSR, Izv*, 11(2): 317–333, 1977.

Kasiviswanathan, S. P., Lee, H. K., Nissim, K., Raskhodnikova, S., and Smith, A. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

Levy, D., Sun, Z., Amin, K., Kale, S., Kulesza, A., Mohri, M., and Suresh, A. T. Learning with user-level privacy. *Advances in Neural Information Processing Systems*, 34: 12466–12479, 2021.

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

McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pp. 1273–1282. PMLR, 2017.

Mironov, I. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, pp. 263–275. IEEE, 2017.

Suresh, A. T., Felix, X. Y., Kumar, S., and McMahan, H. B. Distributed mean estimation with limited communication. In *International conference on machine learning*, pp. 3329–3337. PMLR, 2017.

Ullman, J. Cs7880. rigorous approaches to data privacy. 2017. URL http://www.ccs.neu.edu/home/jullman/cs7880s17/HW1sol.pdf.

Warner, S. L. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965a.

Warner, S. L. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965b.