Exact Optimality in Communication-Privacy-Utility Tradeoffs

Berivan Isik [1] **Wei-Ning Chen** [1] **Ayfer Ozgur** [1] **Tsachy Weissman** [1] **Albert No** [2]

Abstract

We study the mean estimation problem under communication and local differential privacy constraints. While previous work has proposed *order*-optimal algorithms for the same problem (i.e., asymptotically optimal as we spend more bits), *exact* optimality (in the non-asymptotic setting) still has not been achieved. We take a step towards characterizing the *exact*-optimal approach in the presence of shared randomness and identify several necessary conditions for *exact* optimality. We prove that one of the necessary conditions is to utilize a rotationally symmetric shared random codebook. Based on this, we propose a randomization mechanism where the codebook is a randomly rotated simplex – satisfying the necessary properties of the *exact*-optimal codebook. The proposed mechanism is based on a k-closest encoding which we prove to be *exact*-optimal for the randomly rotated simplex codebook. (See (Isik et al., 2023a) for the full version.)

1. Introduction

The distributed mean estimation problem has attracted attention from the machine learning community as it is a canonical statistical formulation for many stochastic optimization problems such as distributed SGD (Agarwal et al., 2018; Barnes et al., 2020; Suresh et al., 2017; 2022) and federated learning (Vargaftik et al., 2022; 2021). As these tasks require data collection from the users, the mean estimation problem has often been studied under privacy constraints to protect users' sensitive information. More specifically, several works (Asi et al., 2022; Bhowmick et al., 2018; Duchi & Rogers, 2019; Duchi et al., 2013; 2018; Nguyên et al., 2016; Wang et al., 2019) have analyzed and improved the tradeoff between the utility and ε-local differential privacy (ε-LDP) – the predominant paradigm in privacy mechanisms, which guarantees that an adversary cannot distinguish the user data from the outcome of the privacy mechanism (Dwork et al., 2006; Kasiviswanathan et al., 2011). Among them, (Bhowmick et al., 2018; Duchi et al., 2013; 2018) developed algorithms that are asymptotically optimal, achieving an optimal mean squared error (MSE) proportional to $\Theta\left(\frac{d}{n\min(\varepsilon,\varepsilon^2)}\right)$, where n is the number of users, and d is the input dimension. Later, (Duchi & Rogers, 2019) proved the corresponding lower bounds that hold for all privacy regimes. However, only `PrivUnit` (Bhowmick et al., 2018) enjoys *exact* optimality among a large family of mechanisms, as proved by (Asi et al., 2022), while others provide only *order* optimality and their performance in practice depends heavily on the constant factors.

Another important consideration in the applications of mean estimation (such as federated learning) is the communication cost during user data collection. This has motivated extensive research on mean estimation (Suresh et al., 2017; Vargaftik et al., 2021; Zhang et al., 2013) and distributed SGD (Agarwal et al., 2018; Barnes et al., 2020; Gandikota et al., 2021; Lin et al., 2018; Wen et al., 2017) under communication constraints; and communication-efficient federated learning (Isik et al., 2023b;c; 2022; Vargaftik et al., 2022).

In addition to the lines of work that studied these constraints (either privacy or communication) separately, recently, there has also been advancement in the joint problem of mean estimation under both privacy and communication constraints. (Chen et al., 2020) introduced an *order*-optimal mechanism `SQKR` requiring only $O(\varepsilon)$ bits by using shared randomness – a random variable shared between the server and the user (see Section 2 for the formal definition). Later, (Shah et al., 2022) demonstrated better MSE with another *order*-optimal algorithm, `MMRC`, by simulating `PrivUnit` using an importance sampling technique (Chatterjee & Diaconis, 2018; Havasi et al., 2019) – again with shared randomness. In the absence of shared randomness, the *order*-optimal mechanisms proposed by (Chen et al., 2020) do not achieve the best-known accuracy under this setting and are outperformed by the lossless compression approach in (Feldman & Talwar, 2021) that compresses `PrivUnit` using a pseudo-random generator (PRG). Due to not using shared randomness, these mechanisms require more bits than others (Chen et al., 2020; Shah et al., 2022) that use shared randomness in the scenarios where it is actually available.

[1]Stanford University [2]Hongik University. Correspondence to: Berivan Isik <berivan.isik@stanford.edu>.

ICML 2023 Workshop on Federated Learning and Analytics in Practice: Algorithms, Systems, Applications, and Opportunities. This workshop does not have official proceedings and this paper is non-archival.

To our knowledge, no existing mechanism achieves *exact* optimality under both privacy and communication constraints with shared randomness[1]. In this work, we address this gap by treating the joint problem as a lossy compression problem under ε-LDP constraints. Our first contribution is to demonstrate that the *exact* optimal scheme with shared randomness can be represented as random coding with a codebook-generating distribution. Specifically, under b bits of communication constraint, the server and the user generate a codebook consisting of $M = 2^b$ vectors (codewords) using shared randomness. The user then selects an index of a vector under a distribution that satisfies ε-LDP, and the server claims the corresponding vector upon receiving the index. We show that this approach, "random coding with a codebook", is the optimal way to use shared randomness.

Next, we prove that the *exact*-optimal codebook-generating distribution must be rotationally symmetric. Based on this insight, we propose Random Rotating Simplex Coding (`RRSC`), where the codebook-generating distribution is a uniformly randomly rotating simplex. This choice of codebook distribution is reasonable as it maximizes the separation between codewords, which efficiently covers the sphere. The corresponding encoding scheme is the k-closest encoding, where the top-k closest codewords to the input obtain high sampling probability, and the remaining ones are assigned low probabilities. We show that this scheme is *exact*-optimal for the random rotating simplex codebook. As the number of bits b for communication increases, we demonstrate that the proposed scheme approaches `PrivUnit`, which is the *exact*-optimal scheme without communication constraints.

Finally, through simulations, we demonstrate that `RRSC` outperforms the existing *order*-optimal methods such as `SQKR` (Chen et al., 2020) and `MMRC` (Shah et al., 2022); and observe that the performance of `RRSC` is remarkably close to that of `PrivUnit` when $b = \epsilon$.

2. Problem Setting and Preliminaries

Local Differential Privacy (LDP). A randomized algorithm $\mathcal{Q} : \mathcal{X} \to \mathcal{Y}$ is ε-LDP if $\forall x, x' \in \mathcal{X}, y \in \mathcal{Y}, \frac{\mathcal{Q}(y|x)}{\mathcal{Q}(y|x')} \leq e^{\varepsilon}$.

In this work, we assume that the encoder and the decoder have access to a shared source of randomness $U \in \mathcal{U}$, where the random encoder (randomizer) $\mathcal{Q}$ privatizes x with additional randomness U. Then, the corresponding ε-LDP constraint is $\forall x, x' \in \mathcal{X}, y \in \mathcal{Y}, \frac{\mathcal{Q}(y|x,u)}{\mathcal{Q}(y|x',u)} \leq e^{\varepsilon}$.

Notation. $\mathbb{S}^{d-1} = \{u \in \mathbb{R}^d : ||u||_2 = 1\}$ denotes the unit sphere, $e_i \in \mathbb{R}^d$ the standard basis vector, $\lfloor k \rfloor$ the greatest integer less than or equal to k, and $[M] = \{1, \ldots, M\}$.

[1]Note that we can also eliminate shared randomness with a private coin setting. See Appendix J for a discussion.

Problem Setting. We consider the ℓ_2 mean estimation problem with n users where each user i has a private unit vector $v_i \in \mathbb{S}^{d-1}$ for $1 \leq i \leq n$. The server wants to recover the mean $\frac{1}{n}\sum_{i=1}^n v_i$ after each user sends a message using up to b bits under ε-LDP constraint. We allow shared randomness between each user and the server. More concretely, the i-th user and the server both have access to a random variable $U_i \in \mathbb{R}^t$ (which is independent of the private local vector v_i) for some $t \geq 1$ and the i-th user has a random encoder (randomizer) $f_i : \mathbb{S}^{d-1} \times \mathbb{R}^t \to [M]$, where $M = 2^b$. We denote by $Q_{f_i}(m_i|v_i, u_i)$ the transition probability induced by the random encoder f_i. We require that the random encoder f_i satisfies ε-LDP, i.e., $\frac{Q_{f_i}(m_i|v_i,u_i)}{Q_{f_i}(m_i|v_i',u_i)} \leq e^{\varepsilon}$ for all $v_i, v_i' \in \mathbb{S}^{d-1}, m_i \in [M]$ and P_{U_i}-almost all $u_i \in \mathbb{R}^t$. The server receives $m_i = f_i(v_i, U_i)$ from all users and generates unbiased estimate of the mean $\mathcal{A}(m_1, \ldots, m_n, U_1, \ldots, U_n)$ that satisfies $\mathbb{E}\left[\mathcal{A}(m_1, \ldots, m_n, U_1, \ldots, U_n)\right] = \frac{1}{n}\sum_{i=1}^n v_i$. Then, the goal is to minimize the worst-case error $\mathrm{Err}_n(f, \mathcal{A}, P_{U^n})$

$$\sup_{v_1,\ldots,v_n \in \mathbb{S}^{d-1}} \mathbb{E}\left[\left\|\mathcal{A}(m_1, \ldots, m_n, U_1, \ldots, U_n) - \frac{1}{n}\sum_{i=1}^n v_i\right\|_2^2\right], \tag{1}$$

where f denotes the collection of all encoders $(f_1, \ldots, f_n)$. We note that the error is also a function of the distribution of shared randomness, which was not the case for `PrivUnit` (Asi et al., 2022; Bhowmick et al., 2018).

3. Main Results

3.1. Canonical Protocols

Similar to (Asi et al., 2022), we first define the canonical protocol under both communication and privacy constraints. The canonical protocols are where the server recovers each user's vector and estimates the mean by averaging them. In other words, the server has a decoder $g_i : [M] \times \mathbb{R}^t \to \mathbb{S}^{d-1}$ for $1 \leq i \leq M$ dedicated to the i-th user's encoder f_i, where the mean estimation is a simple additive aggregation, i.e., $\mathcal{A}^+(m_1, \ldots, m_n, U_1, \ldots, U_n) = \frac{1}{n}\sum_{i=1}^n g_i(m_i, U_i)$. Our first result is that the *exact*-optimal mean estimation scheme should follow this canonical protocol.

Lemma 3.1. *For any n-user mean estimation protocol $(f, \mathcal{A}, P_{U^n})$ satisfying unbiasedness and ε-LDP, there exists an unbiased canonical protocol with decoders $g = (g_1, \ldots, g_n)$, ε-LDP and lower MSE: $\mathrm{Err}_n(f, \mathcal{A}, P_{U^n}) \geq \frac{1}{n^2}\sum_{i=1}^n \mathrm{Err}_1(f_i, g_i, P_{U_i})$, where $\mathrm{Err}_1(f, g, P_U)$ is the worst-case error for a single user with a decoder g. (We provide the proof in Appendix B.)*

Since the *exact*-optimal n-user mean estimation scheme is simply additively aggregating user-wise *exact*-optimal scheme, throughout the paper, we will focus on the

single-user case and drop the index i when it is clear from the context. In this simpler formulation, we want the server to have an unbiased estimate $\hat{v} = g(m, U)$, i.e., $v = \mathbb{E}_{P_U,f}[g(f(v,U),U)] = \mathbb{E}_{P_U}\left[\sum_{m=1}^{M} g(m,U)Q_f(m|v,U)\right]$ for all $v \in \mathbb{S}^{d-1}$. We assume a deterministic decoder $g : [M] \times \mathbb{R}^t \to \mathbb{R}^d$, since the randomized decoder does not improve the performance. Then, the corresponding error is $D(v,f,g,P_U) = \mathbb{E}_{P_U,f}\left[\|g(f(U,v),U) - v\|^2\right] = \mathbb{E}_{P_U}\left[\sum_{m=1}^{M} \|g(m,U) - v\|^2 Q_f(m|v,U)\right]$. Finally, we want to minimize the following worst-case error over all (f,g) pairs that satisfy the unbiasedness condition: $\mathrm{Err}_1(f,g,P_U) = \sup_{v\in\mathbb{S}^{d-1}} D(v,f,g,P_U)$.

3.2. Exact Optimality of the Codebook

We propose a special way of leveraging shared randomness, which we term as *random codebook*. First, we define a codebook $U^M = (U_1,\dots,U_M) \in (\mathbb{R}^d)^M$, consisting of M number of d-dimensional random vectors generated via shared randomness. We then define the corresponding simple selecting decoder $g^+ : [M] \times (\mathbb{R}^d)^M \to \mathbb{R}^d$, which simply picks the m-th vector of the codebook upon receiving the message m from the user: $g^+(m, U^M) = U_m$. Our first theorem shows that there exists a scheme with a random codebook and a simple selecting decoder that achieves the *exact*-optimal error. More precisely, instead of considering the general class of shared randomness (with general dimension t) and the decoder, it is enough to consider the random codebook $U^M \in (\mathbb{R}^d)^M$ as the shared randomness and the simple selector g^+ as the decoder.

Lemma 3.2. *For any f, g, P_U with $U \in \mathbb{R}^t$ that are unbiased and that satisfy ε-LDP, there exists a shared randomness $\tilde{U}^M \in (\mathbb{R}^d)^M$ and random encoder $f_0 : \mathbb{S}^{d-1} \times (\mathbb{R}^d)^M \to [M]$ such that $D(v,f,g,P_U) = D(v,f_0,g^+,P_{\tilde{U}^M})$ for all $v \in \mathbb{S}^{d-1}$, where $f_0, g^+, P_{\tilde{u}^M}$ also satisfy unbiasedness and ε-LDP. (The proof is given in Appendix C.)*

Thus, without loss of generality, we can assume $t = M \times d$ and the random codebook U^M is the new shared randomness, where the decoder is a simple selector. Since we fix the decoder, we drop g to simplify our notation. We say the random encoder f satisfies unbiasedness condition if $\mathbb{E}_{P_U}\left[\sum_{m=1}^{M} U_m Q_f(m|v,U^M)\right] = v$, and the worst-case error is $\mathrm{Err}(f,P_{U^M}) = \sup_{v\in\mathbb{S}^{d-1}} D(v,f,P_{U^M}) = \sup_{v\in\mathbb{S}^{d-1}} \mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^{M} \|U_m - v\|^2 Q_f(m|v,U)\right]$. Thus, the goal is now to find the *exact*-optimum codebook generating distribution P_{U^M}, and the random encoder f (or the probability assignment $Q_f(\cdot|v,U)$). We then argue that the *exact*-optimal codebook should be rotationally symmetric.

Definition 3.3. Random codebook $U^M \in (\mathbb{R}^d)^M$ is rotationally symmetric if $(U_1,\dots,U_M) \overset{(d)}{=} (A_0U_1,\dots,A_0U_M)$ for any $d \times d$ orthonormal A_0.

Lemma 3.4. *Let P_{U^M} be a codebook generating distribution, and suppose random encoder f satisfies unbiasedness and ε-LDP. Then, there exists an random encoder f_1 and rotationally symmetric random codebook $\bar{U}^M$ such that $\mathrm{Err}(f,P_{U^M}) \geq \mathrm{Err}(f_1,P_{\bar{U}^M})$, which also satisfies unbiasedness and ε-LDP.*

This is mainly because the goal is to minimize the worst-case error, and the codebook-generating distribution should be symmetric in all directions. The proof is provided in Appendix D. The next lemma shows that the *exact*-optimal scheme has constant error for all $v \in \mathbb{S}^{d-1}$.

Lemma 3.5. *For any rotationally symmetric codebook generating distribution P_{U^M} and an unbiased randomized encoder f that satisfies ε-LDP, there exists a random encoder f_2 such that $\mathrm{Err}(f,P_{U^M}) \geq \mathrm{Err}(f_2,P_{U^M})$, where $D(v,f_2,P_{U^M}) = D(v',f_2,P_{U^M})$, for all $v,v' \in \mathbb{S}^{d-1}$. (The formal proof is given in Appendix E.)*

3.3. Rotationally Symmetric Simplex Codebook

Now, we focus on a particular rotationally symmetric codebook. Notice that it is natural to consider the maximally separated codebook so that the M vectors $U_1,\dots,U_M$ cover the source space effectively. For $M < d$, the maximally separated M vectors on the unit sphere $\mathbb{S}^{d-1}$ is a simplex. More precisely, let $s_1,\dots,s_M \in \mathbb{R}^d$ form a simplex:

$$(s_i)_j = \begin{cases} \frac{M-1}{\sqrt{M(M-1)}} & \text{if } i = j \\ -\frac{1}{\sqrt{M(M-1)}} & \text{if } i \neq j \text{ and } j \leq M \\ 0 & \text{if } j > M \end{cases}. \quad (2)$$

Then, we can define the rotationally symmetric simplex codebook U^M as $(U_1,U_2,\dots,U_M) \overset{(d)}{=} (rAs_1, rAs_2,\dots,rAs_M)$, where A is uniformly drawn orthogonal matrix and $r > 0$ is a normalizing constant. We then need to find the corresponding encoder f that minimizes the error. Recall that the error is $\mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^{M} \|U_m - v\|^2 Q_f(m|v,U)\right]$, and it is natural to assign high probabilities to the message m with low distortion $\|U_m - v\|^2$ as long as ε-LDP constraint allows. More precisely, we call the following probability assignment "k-closest" encoding:

$$Q_f(m|v,U^M) = \begin{cases} \frac{e^\varepsilon}{ke^\varepsilon+(M-k)} & \text{if } \|v-U_m\|^2 \text{ is one of the } \lfloor k \rfloor \text{ smallest} \\ \frac{(k-\lfloor k \rfloor)(e^\varepsilon-1)+1}{ke^\varepsilon+(M-k)} & \text{if } \|v-U_m\|^2 \text{ is the } \lfloor k \rfloor + 1\text{-th smallest} \\ \frac{1}{ke^\varepsilon+(M-k)} & \text{otherwise} \end{cases}. \quad (3)$$

The choice of $r = r_k$ is described in Section 3.4. We call this approach Randomly Rotating Simplex Coding (RRSC)

and provide the pseudocode in Algorithm 1. We note that the codewords U_m's with smallest $\|v - U_m\|^2$ and codewords U_m's with largest $\langle v, U_m \rangle$ coincide for a codebook with fixed-norm codewords U_m's, which is the case for the rotationally symmetric simplex codebook. We now present our main theorem.

Theorem 3.6. *For a rotationally symmetric simplex codebook, there exists a k such that the "k-closest" encoding is the exact-optimum unbiased scheme that satisfies ε-LDP constraint. (The proof is given in Appendix F.)*

Algorithm 1 Randomly Rotating Simplex Coding $\texttt{RRSC}(k)$.

Inputs: $v \in \mathbb{S}^{d-1}$, k, $M = 2^b$.

Codebook Generation:
Generate the simplex $s_1, \ldots, s_M \in \mathbb{R}^d$ in (2).
Sample orthogonal matrix $A \in \mathbb{R}^{d \times d}$ uniformly using the shared random SEED.
Generate the codebook U^M: $(U_1, U_2, \ldots, U_M) \leftarrow (r_k A s_1, r_k A s_2, \ldots, r_k A s_M)$.
Encoding:
for $m \in [M]$ **do**
  if $\langle v, U_m \rangle$ is one of the k largest **then**
    $Q_f(m|v, U^M) \leftarrow \frac{e^\varepsilon}{ke^\varepsilon + (M-k)}$
  else
    $Q_f(m|v, U^M) \leftarrow \frac{1}{ke^\varepsilon + (M-k)}$
  end if
end for
Sample codeword index $m^* \leftarrow Q_f(\cdot|v, U^M)$.

Output: m^*, encoded in $b = \log M$ bits.

3.4. k-closest Encoding for General Rotationally Symmetric Codebook

In this section, we demonstrate that the k-closest encoding consistently yields an unbiased scheme for any rotationally symmetric codebook. To be more specific, for any given spherically symmetric codebook U^M, there exists a scalar r_k that ensures that the k-closest encoding with $r_k U^M = (r_k U_1, \ldots, r_k U_M)$ is unbiased. Let $T_k(v, U^M) = \{m : U_m \text{ is one of the } k\text{-closest}\}$, and without loss of generality, let us assume $v = e_1$. Then, $\mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^M Q_f(m|e_1, U^M) U_m\right] = \mathbb{E}_{P_{U^M}}\left[\frac{e^\varepsilon - 1}{ke^\varepsilon + (M-k)} \sum_{m \in T_k(e_1, U^M)} U_m\right]$, where we assume an integer k for the sake of simplicity. Since the codebook is rotationally symmetric and we pick k-closest vectors toward $v = e_1$, each codeword $U_m \in T_k(e_1, U^M)$ is symmetric in all directions other than $v = e_1$. Thus, in expectation, the decoded vector is aligned with e_1, and there exists r_k such that $r_k \times \mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^M Q_f(m|e_1, U^M) U_m\right] = e_1$. For a rotationally symmetric simplex codebook, where $U_m = A s_m$ for a uniform random orthogonal matrix A, we have an (almost) analytic formula.

Lemma 3.7. *Normalization constant r_k for $\texttt{RRSC}(k)$ is $r_k = \frac{ke^\varepsilon + M - k}{e^\varepsilon - 1}\sqrt{\frac{M-1}{M}\frac{1}{C_k}}$, where C_k is an expected sum of top-k coordinates of uniform random vector $a \in \mathbb{S}^{d-1}$. (The proof is in Appendix G.)*

4. Experiments

We empirically demonstrate the communication-privacy-utility tradeoffs of RRSC and compare it with *order*-optimal schemes under privacy and communication constraints, namely SQKR (Chen et al., 2020) and MMRC (Shah et al., 2022). We also show that RRSC performs comparably with PrivUnit (Bhowmick et al., 2018), which offers the *exact*-optimal privacy-utility tradeoffs without communication constraints (Asi et al., 2022). In our simulations, we use the "optimized" PrivUnit mechanism, called PrivUnitG, introduced in (Asi et al., 2022), which performs better than PrivUnit in practice since it provides an easy-to-analyze approximation of PrivUnit but with analytically better-optimized hyperparameters. Similar to (Chen et al., 2020; Shah et al., 2022), we generate data independently but non-identically to capture the distribution-free setting with $\mu \neq 0$. More precisely, for the first half of the users, we set $v_1, \ldots, v_{n/2} \overset{\text{i.i.d}}{\sim} N(1,1)^{\otimes d}$; and for the second half of the users, we set $v_{n/2+1}, \ldots, v_n \overset{\text{i.i.d}}{\sim} N(10,1)^{\otimes d}$. We further normalize each v_i to ensure that they lie on $\mathbb{S}^{d-1}$. We report the average ℓ_2 error over 10 rounds together with the confidence intervals. To find the optimal values for k and r_k, we compute the optimal r_k using the formula in (68) for $k = 1, \ldots, M$ and pick the k that gives the smallest r_k (which corresponds to the bias). To estimate the expectation C_k in (68), we run a Monte Carlo simulation with $1M$ trials. Additional experimental results are provided in Appendix H.

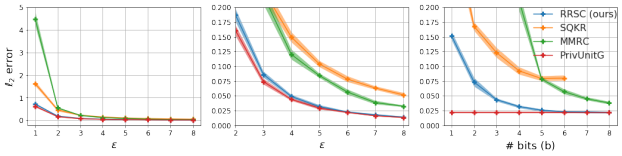

Figure 1: Comparison of RRSC with SQKR, MMRC, and PrivUnitG. **(left)** ℓ_2 error vs ε with $n = 5000$, $d = 500$. The number of bits is $b = \varepsilon$ for RRSC, SQKR, and MMRC. **(middle)** Same plot zoomed into higher ε, lower ℓ_2 error region. **(right)** ℓ_2 error vs number of bits b for $n = 5000$, $d = 500$, and $\varepsilon = 6$. For SQKR, we only report $b \leq \varepsilon = 6$ since it performs poorly when $b > \varepsilon$.

In Figure 1-(left, middle), we report ℓ_2 error for $\varepsilon = 1, \ldots, 8$, where for each method (except PrivUnitG), the number of bits is equal to $b = \varepsilon$. In Figure 1-(right), we

report ℓ_2 error by fixing $\epsilon = 6$ and sweeping the bitrate from $b = 1$ to $b = 8$ for RRSC and MMRC. For SQKR, we only sweep for $b \leq \varepsilon$ as it leads to poor performance for $b > \varepsilon$. In each figure, RRSC performs comparably to PrivUnitG even for small b and outperforms both SKQR and MMRC by large margins.

5. Discussion & Conclusion

In this work, we proved that using a rotationally symmetric codebook is a necessary condition for the *exact* optimality of mean estimation mechanisms under privacy and communication constraints. We then proposed Random Rotating Simplex Coding (RRSC) based on a k-closest encoding mechanism and proved that RRSC is *exact*-optimal for the random rotating simplex codebook. In Appendix J, we discuss some important features of RRSC such as convergence to PrivUnit and complexity; and finally provide conjectures for future work.

6. Acknowledgement

The authors would like to thank Peter Kairouz and Lucas Theis for the helpful discussions in the early stage of this work; and Hilal Asi, Vitaly Feldman, and Kunal Talwar for sharing their implementation for the PrivUnitG mechanism (Asi et al., 2022). This work was supported in part by a Stanford Graduate Fellowship, a Meta research grant, and an NSF Award #CCF-2213223.

References

Agarwal, N., Suresh, A. T., Yu, F. X. X., Kumar, S., and McMahan, B. cpsgd: Communication-efficient and differentially-private distributed sgd. *Advances in Neural Information Processing Systems*, 31, 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.

Barnes, L. P., Inan, H. A., Isik, B., and Özgür, A. rtop-k: A statistical estimation approach to distributed sgd. *IEEE Journal on Selected Areas in Information Theory*, 1(3): 897–907, 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.

Chatterjee, S. and Diaconis, P. The sample size required in importance sampling. *The Annals of Applied Probability*, 28(2):1099–1135, 2018.

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

Duchi, J. and Rogers, R. Lower bounds for locally private estimation via communication complexity. In *Conference on Learning Theory*, pp. 1161–1191. PMLR, 2019.

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.

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

Dwork, C., McSherry, F., Nissim, K., and Smith, A. 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.

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

Gandikota, V., Kane, D., Maity, R. K., and Mazumdar, A. vqsgd: Vector quantized stochastic gradient descent. In *International Conference on Artificial Intelligence and Statistics*, pp. 2197–2205. PMLR, 2021.

Ghadimi, S. and Lan, G. Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

Havasi, M., Peharz, R., and Hernández-Lobato, J. M. Minimal random code learning: Getting bits back from compressed model parameters. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=r1f0YiCctm.

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

Isik, B., Weissman, T., and No, A. An information-theoretic justification for model pruning. In *International Conference on Artificial Intelligence and Statistics*, pp. 3821–3846. PMLR, 2022.

Isik, B., Chen, W.-N., Ozgur, A., Weissman, T., and No, A. Exact optimality of communication-privacy-utility tradeoffs in distributed mean estimation. *arXiv preprint arXiv:2306.04924*, 2023a.

Isik, B., Pase, F., Gunduz, D., Koyejo, S., Weissman, T., and Zorzi, M. Communication-efficient federated learning through importance sampling. *arXiv preprint arXiv:2306.12625*, 2023b.

Isik, B., Pase, F., Gunduz, D., Weissman, T., and Michele, Z. Sparse random networks for communication-efficient federated learning. In *The Eleventh International Conference on Learning Representations*, 2023c. URL https://openreview.net/forum?id=k1FHgri5y3-.

Kashin, B. Section of some finite-dimensional sets and classes of smooth functions (in russian) izv. *Acad. Nauk. SSSR*, 41:334–351, 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.

Lin, Y., Han, S., Mao, H., Wang, Y., and Dally, B. Deep gradient compression: Reducing the communication bandwidth for distributed training. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=SkhQHMW0W.

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

Nguyên, T. T., Xiao, X., Yang, Y., Hui, S. C., Shin, H., and Shin, J. Collecting and analyzing data from smart device users with local differential privacy. *arXiv preprint arXiv:1606.05053*, 2016.

Shah, A., Chen, W.-N., Balle, J., Kairouz, P., and Theis, L. Optimal compression of locally differentially private mechanisms. In *International Conference on Artificial Intelligence and Statistics*, pp. 7680–7723. PMLR, 2022.

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.

Suresh, A. T., Sun, Z., Ro, J., and Yu, F. Correlated quantization for distributed mean estimation and optimization. In *International Conference on Machine Learning*, pp. 20856–20876. PMLR, 2022.

Vargaftik, S., Ben-Basat, R., Portnoy, A., Mendelson, G., Ben-Itzhak, Y., and Mitzenmacher, M. Drive: One-bit distributed mean estimation. *Advances in Neural Information Processing Systems*, 34:362–377, 2021.

Vargaftik, S., Basat, R. B., Portnoy, A., Mendelson, G., Itzhak, Y. B., and Mitzenmacher, M. Eden: Communication-efficient and robust distributed mean estimation for federated learning. In *International Conference on Machine Learning*, pp. 21984–22014. PMLR, 2022.

Wang, T., Zhao, J., Yang, X., and Ren, X. Locally differentially private data collection and analysis. *arXiv preprint arXiv:1906.01777*, 2019.

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

Wen, W., Xu, C., Yan, F., Wu, C., Wang, Y., Chen, Y., and Li, H. Terngrad: Ternary gradients to reduce communication in distributed deep learning. *Advances in neural information processing systems*, 30, 2017.

Zhang, Y., Wainwright, M. J., and Duchi, J. C. Communication-efficient algorithms for statistical optimization. *Advances in neural information processing systems*, 25, 2012.

Zhang, Y., Duchi, J., Jordan, M. I., and Wainwright, M. J. Information-theoretic lower bounds for distributed statistical estimation with communication constraints. *Advances in Neural Information Processing Systems*, 26, 2013.

A. Related Work

The ℓ_2 mean estimation problem is a canonical statistical formulation for many distributed stochastic optimization methods, such as communication (memory)-efficient SGD (Suresh et al., 2017; Vargaftik et al., 2021) or private SGD (Kasiviswanathan et al., 2011). For instance, as shown in (Ghadimi & Lan, 2013), as long as the final estimator of the mean is unbiased, the ℓ_2 estimation error (i.e., the variance) determines the convergence rate of the distributed SGD. As a result, there is a long thread of works that study the mean estimation problem under communication constraints (Barnes et al., 2020; Duchi et al., 2013; Gandikota et al., 2021; Suresh et al., 2017; Vargaftik et al., 2021; Zhang et al., 2012), privacy constraints (Asi et al., 2022; Huang et al., 2021), or a joint of both (Agarwal et al., 2018; Chen et al., 2020; Feldman & Talwar, 2021; Shah et al., 2022).

Among them, (Chen et al., 2020) shows that $\Theta(\varepsilon)$ bits are sufficient to achieve the *order*-optimal MSE $\Theta\left(\frac{d}{n\min(\varepsilon,\varepsilon^2)}\right)$ and proposes `SQKR`, an *order*-optimal mean estimation scheme under both privacy and communication constraints. Notice that the MSE of SQKR is *orderwise* optimal up to a constant factor. Later on, in (Feldman & Talwar, 2021), it is shown that the pre-constant factor in `SQKR` is indeed suboptimal, resulting in an unignorable gap in the MSE compared to `PrivUnit` – an optimal ℓ_2 mean estimation scheme under ε-LDP. In the original `PrivUnit`, the output space is a d-dimensional sphere $\mathbb{S}^{d-1}$ and hence requires $O(d)$ bits of communication, which is far from the optimal $O(\varepsilon)$ communication bound. However, (Feldman & Talwar, 2021) shows that one can (almost) losslessly compress `PrivUnit` via a pseudo-random generator (PRG). Under the assumption of an existing exponentially strong PRG, (Feldman & Talwar, 2021) proves that one can compress the output of `PrivUnit` using $\mathsf{polylog}(d)$ bits with negligible performance loss. Similarly, (Shah et al., 2022) shows that with the help of shared randomness, `PrivUnit` can be (nearly) losslessly compressed to $\Theta(\varepsilon)$ bits via a channel simulation technique, called `MMRC`. We remark that although the privacy-utility trade-offs in (Feldman & Talwar, 2021) and (Shah et al., 2022) are (nearly) *exactly* optimal, the communication efficiency is only *order*-optimal. That is, under an exact b-bit communication constraint, the MSEs of (Feldman & Talwar, 2021) (denoted as `FT21`) and `MMRC` (Shah et al., 2022) may be suboptimal. In this work, we aim to achieve the *exact*-optimal MSE under *both* communication and privacy constraints.

Furthermore, we show that `SQKR`, `FT21`, and `MMRC` can be viewed as special cases in our framework – i.e., (random) coding with their own codebook design. We elaborate on this in Section 5 and provide more details on prior work in Appendix I.

B. Proof of Lemma 3.1

Proof. For n-user mean estimation protocol $(f, \mathcal{A}, P_{U^M})$, following the notation and steps from (**?**)Proof of Lemma 3.1]asi2022optimal, we define the marginalized output

$$\tilde{g}_i(m_i, U_i; v^n) = \mathbb{E}_{\{m_j,U_j\}_{j\neq i}}\left[n\mathcal{A}(\{m_j,U_j\}_{j=1}^n) \;\middle|\; f_i(v_i,U_i)=m_i, U_i, v^{n\setminus i}\right]. \tag{4}$$

Then, we define the user-specific decoder by averaging $g_i(m_i, U_i; v^n)$ with respect to i.i.d. uniform P_{unif}:

$$g_i(m_i, U_i) = \mathbb{E}_{v^{n\setminus i}\sim P_{\text{unif}}}\left[\tilde{g}_i(m_i, U_i; v^n)\right] \tag{5}$$

where $v^{n\setminus i}$ indicates the v^n vector except v_i. Due to the symmetry of P_{unif}, it is clear that g_i is unbiased. We also define

$$\hat{\mathcal{R}}_{\leq i}(\{v_j,m_j,U_j\}_{j=1}^i) = \mathbb{E}_{v_j\sim P_{\text{unif}}, j>i}\left[n\mathcal{A}(\{m_j,U_j\}_{j=1}^n) - \sum_{j=1}^i v_j \;\middle|\; \{v_j,m_j,U_j\}_{j=1}^i\right] \tag{6}$$

Consider an average error where $v_1,\ldots,v_n$ are drawn i.i.d. uniformly on the sphere $\mathbb{S}^{d-1}$.

$$\begin{aligned}
&\mathbb{E}_{\{v_j,m_j,U_j\}_{j=1}^n}\left[\left\|n\mathcal{A}(\{m_j,U_j\}_{j=1}^n) - \textstyle\sum_{j=1}^n v_j\right\|^2\right] \\
&= \mathbb{E}_{\{v_j,m_j,U_j\}_{j=1}^n}\left[\left\|\hat{\mathcal{R}}_{\leq n}(\{v_j,m_j,U_j\}_{j=1}^n)\right\|^2\right] && (7)\\
&= \mathbb{E}_{\{v_j,m_j,U_j\}_{j=1}^n}\left[\left\|\hat{\mathcal{R}}_{\leq n}(\{v_j,m_j,U_j\}_{j=1}^n) - \hat{\mathcal{R}}_{\leq n-1}(\{v_j,m_j,U_j\}_{j=1}^{n-1}) + \hat{\mathcal{R}}_{\leq n-1}(\{v_j,m_j,U_j\}_{j=1}^{n-1})\right\|^2\right] && (8)\\
&= \mathbb{E}_{\{v_j,m_j,U_j\}_{j=1}^n}\left[\left\|\hat{\mathcal{R}}_{\leq n}(\{v_j,m_j,U_j\}_{j=1}^n) - \hat{\mathcal{R}}_{\leq n-1}(\{v_j,m_j,U_j\}_{j=1}^{n-1})\right\|^2\right]
\end{aligned}$$

$$+\mathbb{E}_{\{v_j,m_j,U_j\}_{j=1}^{n-1}}\left[\left\|\hat{\mathcal{R}}_{\leq n-1}(\{v_j,m_j,U_j\}_{j=1}^{n-1})\right\|^2\right] \tag{9}$$

$$=\sum_{i=1}^{n}\mathbb{E}_{\{v_j,m_j,U_j\}_{j=1}^{i}}\left[\left\|\hat{\mathcal{R}}_{\leq i}(\{v_j,m_j,U_j\}_{i=1}^{n})-\hat{\mathcal{R}}_{\leq i-1}(\{v_j,m_j,U_j\}_{j=1}^{i-1})\right\|^2\right] \tag{10}$$

$$\geq\sum_{i=1}^{n}\mathbb{E}_{m_i,U_i}\left[\left\|\mathbb{E}_{\{v_j,m_j,U_j\}_{j=1}^{i-1}}\left[\hat{\mathcal{R}}_{\leq i}(\{v_j,m_j,U_j\}_{i=1}^{n})-\hat{\mathcal{R}}_{\leq i-1}(\{v_j,m_j,U_j\}_{j=1}^{i-1})\right]\right\|^2\right] \tag{11}$$

$$=\sum_{i=1}^{n}\mathbb{E}_{m_i,U_i}\left[\|g_i(m_i,U_i)-v_i\|^2\right]. \tag{12}$$

Then, we need to show the same inequality for the worst-case error.

$$\sup_{v_1,\ldots,v_n}\mathbb{E}_{\{m_j,U_j\}_{j=1}^{n}}\left[\left\|n\mathcal{A}(\{m_j,U_j\}_{j=1}^{n})-\sum_{j=1}^{n}v_j\right\|^2\right]$$

$$\geq\mathbb{E}_{\{v_j,m_j,U_j\}_{j=1}^{n}}\left[\left\|n\mathcal{A}(\{m_j,U_j\}_{j=1}^{n})-\sum_{j=1}^{n}v_j\right\|^2\right] \tag{13}$$

$$=\sum_{i=1}^{n}\mathbb{E}_{v_i,m_i,U_i}\left[\|g_i(m_i,U_i)-v_i\|^2\right] \tag{14}$$

$$=\sum_{i=1}^{n}\sup_{v_i}\mathbb{E}_{m_i,U_i}\left[\|g_i(m_i,U_i)-v_i\|^2\right] \tag{15}$$

where the last equality (15) is from Lemma 3.2, Lemma 3.4, and Lemma 3.5. Thus, the user-specific decoder achieves lower MSE:

$$\mathrm{Err}_n(f,\mathcal{A},P_{U^n})\geq\frac{1}{n}\sum_{i=1}^{n}\mathrm{Err}_1(f_i,g_i,P_{U_i}). \tag{16}$$

Since we keep random encoder f_i the same, the canonical protocol with g_i also satisfies ε-LDP constraint. This concludes the proof. □

C. Proof of Lemma 3.2

Proof. Let $\tilde{U}_m = g(m,U)$ for all $1\leq m\leq M$. Without loss of generality $g(\cdot,U)$ is one-to-one, i.e., $\{u:\tilde{u}_m = g(m,u)$ for all $m\}$ has at most one element (with probability 1), and $u=g^{-1}(\tilde{u}^M)$ is well-defined. Then, we define a randomizer $f_0(v,\tilde{U}^M)$ that satisfies

$$Q_{f_0}(m|v,\tilde{u}^M)=Q_f(m|v,g^{-1}(\tilde{u}^M)). \tag{17}$$

It is clear that f_0 satisfies ε-LDP constraint. Then,

$$D(v,f_0,g^+,P_{\tilde{U}^M})=\mathbb{E}_{f_0,P_{\tilde{U}^M}}\left[\|g^+(f_0(v,\tilde{U}^M),\tilde{U}^M)-v\|^2\right] \tag{18}$$

$$=\mathbb{E}_{P_{\tilde{U}^M}}\left[\textstyle\sum_{m=1}^{M}Q_{f_0}(m|v,\tilde{U}^M)\|\tilde{U}_m-v\|^2\right] \tag{19}$$

$$=\mathbb{E}_{f,P_U}\left[\textstyle\sum_{m=1}^{M}Q_f(m|v,U)\|g(m,U)-v\|^2\right] \tag{20}$$

$$=\mathbb{E}_{f,P_U}\left[\|g(f(v,U),U)-v\|^2\right] \tag{21}$$

$$=D(v,f,g,P_U). \tag{22}$$

We also need to show that the composition of the new randomizer f_0 and selector g^+ is unbiased.

$$\mathbb{E}_{P_{\tilde{U}^M}}\left[g^+(f_0(v,\tilde{U}^M),\tilde{U}^M)\right]=\mathbb{E}_{f_0,P_{\tilde{U}^M}}\left[\textstyle\sum_{m=1}^{M}Q_{f_0}(m|v,\tilde{U}^M)\tilde{U}_m\right] \tag{23}$$

$$=\mathbb{E}_{f,P_U}\left[\sum_{m=1}^{M} Q_f(m|v,U)g(m,U)\right] \tag{24}$$
$$=\mathbb{E}_{f,P_U}\left[g(f(v,U),U)\right] \tag{25}$$
$$=v. \tag{26}$$

Finally, $Q_{f_0}(m|v,\tilde{u}^M)$ is a valid transition probability, since

$$\sum_{m=1}^{M} Q_{f_0}(m|v,\tilde{u}^M) = \sum_{m=1}^{M} Q_f(m|v,g^{-1}(\tilde{u}^M)) = 1 \tag{27}$$

for all $\tilde{u}^M$. This concludes the proof. □

D. Proof of Lemma 3.4

Proof. Let A be a uniformly random orthogonal matrix and $\bar{U}^M = A^\intercal U^M$. We further let f_1 be a randomized encoder that satisfies

$$Q_{f_1}(m|v,\bar{U}^M) = \mathbb{E}_A\left[Q_f(m|Av,A\bar{U}^M)|\bar{U}^M\right]. \tag{28}$$

Then, Q_{f_1} is a valid probability since

$$\sum_{m=1}^{M} Q_{f_1}(m|v,\bar{U}^M) = \mathbb{E}_A\left[\sum_{m=1}^{M} Q_f(m|Av,A\bar{U}^M)|\bar{U}^M\right] = 1. \tag{29}$$

Also, we have

$$\frac{Q_{f_1}(m|v,\bar{U}^M)}{Q_{f_1}(m|v',\bar{U}^M)} = \frac{\mathbb{E}_A\left[Q_f(m|Av,A\bar{U}^M)|\bar{U}^M\right]}{\mathbb{E}_A\left[Q_f(m|Av',A\bar{U}^M)|\bar{U}^M\right]} \tag{30}$$
$$\leq \frac{\mathbb{E}_A\left[e^{\varepsilon}Q_f(m|Av',A\bar{U}^M)|\bar{U}^M\right]}{\mathbb{E}_A\left[Q_f(m|Av',A\bar{U}^M)|\bar{U}^M\right]} \tag{31}$$
$$=e^{\varepsilon}. \tag{32}$$

Finally, we need to check unbiasedness.

$$\mathbb{E}_{P_{\bar{U}^M}}\left[Q_{f_1}(m|v,\bar{U}^M)\bar{U}_m\right] = \mathbb{E}_{A,P_{U^M}}\left[\sum_{m=1}^{M} Q_f(m|Av,A\bar{U}^M)\bar{U}_m\right] \tag{33}$$
$$=\mathbb{E}_{A,P_{U^M}}\left[\sum_{m=1}^{M} Q_f(m|Av,U^M)A^\intercal U_m\right] \tag{34}$$
$$=\mathbb{E}_A\left[A^\intercal \mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^{M} Q_f(m|Av,U^M)U_m\right]\right] \tag{35}$$
$$=\mathbb{E}_A\left[A^\intercal Av\right] \tag{36}$$
$$=v. \tag{37}$$

The key step is that the original encoder f is unbiased, which implies

$$\mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^{M} Q_f(m|Av,U^M)U_m\right] = Av \tag{38}$$

for all A.

Now, we are ready to prove the main inequality.

$$\mathrm{Err}(f,P_{U^M}) = \sup_v D(v,f,P_{U^M}) \tag{39}$$
$$\geq \mathbb{E}_A\left[D(Av,f,P_{U^M})\right] \tag{40}$$
$$=\mathbb{E}_A\left[\mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^{M} Q_f(m|Av,U^M)\|U_m - Av\|^2\right]\right] \tag{41}$$

$$=\mathbb{E}_{P_{U^M},A}\left[\sum_{m=1}^{M} Q_f(m|Av, A\bar{U}^M)\|\bar{U}_m - v\|^2\right] \tag{42}$$

$$=\mathbb{E}_{P_{\bar{U}^M}}\left[\sum_{m=1}^{M} \mathbb{E}_A\left[Q_f(m|Av, A\bar{U}^M)|\bar{U}^M\right]\|\bar{U}_m - v\|^2\right] \tag{43}$$

$$=\mathbb{E}_{P_{\bar{U}^M}}\left[\sum_{m=1}^{M} Q_{f_1}(m|v, \bar{U}^M)\|\bar{U}_m - v\|^2\right] \tag{44}$$

$$=D(v, f_1, P_{\bar{U}^M}). \tag{45}$$

for all v. This concludes the proof. □

E. Proof of Lemma 3.5

Proof. For $v, v' \in \mathbb{S}^{d-1}$, let A_0 be an orthonormal matrix such that $v' = A_0 v$. Let f_2 be a randomized encoder such that

$$f_2(v, U^M) = f(Av, AU^M) \tag{46}$$

for uniform random orthonormal matrix. Then,

$$Q_{f_2}(m|v, U^M) = \mathbb{E}_A\left[Q_f(m|Av, AU^M)\right]. \tag{47}$$

Similar to the previous proofs, Q_{f_2} is a well-defined probability distribution, and f_2 is unbiased as well as ε-LDP. Since P_{U^M} is rotationally symmetric and f_2 is also randomized via the uniform random orthogonal matrix, we have

$$D(v', f_2, P_{U^M}) = D(A_0 v, f_2, P_{U^M}) = D(v, f_2, P_{U^M}). \tag{48}$$

Compared to a given randomizer f, we have

$$\mathrm{Err}(f, P_{U^M}) \geq \mathbb{E}_A\left[D(Av, f, P_{U^M})\right] \tag{49}$$

$$=\mathbb{E}_{A,P_{U^M}}\left[\sum_{m=1}^{M} Q_f(m|Av, U^M)\|Av - U^M\|^2\right] \tag{50}$$

$$=\mathbb{E}_{A,P_{U^M}}\left[\sum_{m=1}^{M} Q_f(m|Av, U^M)\|v - A^\intercal U^M\|^2\right] \tag{51}$$

$$=\mathbb{E}_{A,P_{U^M}}\left[\sum_{m=1}^{M} Q_f(m|Av, AU^M)\|v - U^M\|^2\right] \tag{52}$$

$$=\mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^{M} \mathbb{E}_A\left[Q_f(m|Av, AU^M)\right]\|v - U^M\|^2\right] \tag{53}$$

$$=D(v, f_2, P_{U^M}) \tag{54}$$

for all $v \in \mathbb{S}^{d-1}$. This concludes the proof. □

F. Proof of Theorem 3.6

Proof. The rotationally symmetric simplex codebook with normalization constant r is $(rAs_1, \ldots, rAs_M)$. Let f be the unbiased encoder satisfying ε-LDP. Let $Q_{\max} = \max Q_f(m|v, rAs^M)$ and $Q_{\min} = \min Q_f(m|v, rAs^M)$, our objective is to demonstrate that $Q_{\max}$ is less than or equal to $e^\varepsilon Q_{\min}$. We will employ a proof by contradiction to establish this. Suppose $Q_f(m_1|v_1, rA_1 s^M) > e^\varepsilon Q_f(m_2|v_2, rA_2 s^M)$ for some m_1, v_1, A_1, m_2, v_2, and A_2. Let $\tilde{A}$ be the row switching matrix where $r\tilde{A}A_1 s_{m_1} = rA_1 s_{m_2}$ and $r\tilde{A}A_1 s_{m_2} = rA_1 s_{m_1}$, then we have

$$Q_f(m_1|v_1, rA_1 s^M) = Q_f(m_2|\tilde{A}v_1, r\tilde{A}A_1 s^M). \tag{55}$$

We further let A' be an orthogonal matrix such that $A'\tilde{A}A_1 = A_2$, then

$$Q_f(m_2|\tilde{A}v_1, r\tilde{A}A_1 s^M) = Q_f(m_2|A'\tilde{A}v_1, rA'\tilde{A}A_1 s^M) \tag{56}$$

$$= Q_f(m_2|A'\tilde{A}v_1, rA_2 s^M) \tag{57}$$

If we let $v_1' = A'\tilde{A}v_1$, then

$$Q_f(m_2|v_1', rA_2 s^M) = Q_f(m_1|v_1, rA_1 s^M) \tag{58}$$

$$>e^{\varepsilon} Q_f(m_2|v_2, rA_2 s^M), \tag{59}$$

which contradicts the ε-LDP constraint.

For an unbiased encoder, the error is

$$\mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^{M} \|U_m - v\|^2 Q_f(m|v, U^M)\right] = \mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^{M} \|U_m\|^2 Q_f(m|v, U^M)\right] - 1 \tag{60}$$

$$= r^2 - 1. \tag{61}$$

Thus, we need to find r that minimizes the error.

On the other hand, the encoder needs to satisfy unbiasedness. Without loss of generality, we assume $v = e_1$, then we need

$$\mathbb{E}_A\left[\sum_{m=1}^{M} rAs_m Q_f(m|e_1, rAs^M)\right] = e_1, \tag{62}$$

where the expectation is with respect to the random orthonormal matrix A. If we focus on the first index of the vector, then

$$r \times \mathbb{E}_a\left[\sum_{m=1}^{M} a^{\intercal} s_m Q_f(m|e_1, rAs^M)\right] = 1, \tag{63}$$

where $a^{\intercal} = (a_1, \ldots, a_d)$ is the first row of A and has uniform distribution on the sphere $\mathbb{S}^{d-1}$. Thus, it is clear that assigning higher probability (close to $Q_{\max}$) to the larger $a^{\intercal} s_m$.

If $Q_{\max}$ is strictly smaller than $e^{\varepsilon} Q_{\min}$, then we can always scale up the larger probabilities and scale down the lower probabilities to keep the probability sum to one (while decreasing the error). Hence, we can assume that $Q_{\min} = q_0$ and $Q_{\max} = e^{\varepsilon} q_0$ for some $1 > q_0 > 0$.

Now, let k be such that

$$(M - \lfloor k \rfloor - 1) q_0 + q_i + \lfloor k \rfloor e^{\varepsilon} q_0 = 1, \tag{64}$$

where q_i is an intermediate value such that $q_i \in [q_0, e^{\varepsilon} q_0]$. Then, the optimal strategy is clear: (i) assign $e^{\varepsilon} q_0$ to $\lfloor k \rfloor$-th closest codewords s_m's, (ii) assign q_i to the $(\lfloor k \rfloor + 1)$-th closest codeword, and (iii) assign q_0 to the remaining codewords. This implies that the k-closest coding is optimal. □

G. Proof of Lemma 3.7

Proof. Due to symmetry, encoding e_1 with randomly rotating simplex codebook As^M is equivalent to encode uniform random $a \in \mathbb{S}^{d-1}$ with deterministic simplex codebook s^M. Then, we have

$$r_k \times \mathbb{E}\left[\frac{e^{\varepsilon} - 1}{ke^{\varepsilon} + (M - k)} \sum_{m \in \text{top-}k} a^{\intercal} s_m\right]$$

$$= r_k \times \mathbb{E}\left[\frac{e^{\varepsilon} - 1}{ke^{\varepsilon} + (M - k)} \sum_{m \in \text{top-}k} \sqrt{\frac{M}{M-1}} a_m - \sum_{i=1}^{M} a^{\intercal} s_m\right] \tag{65}$$

$$= r_k \times \mathbb{E}\left[\frac{e^{\varepsilon} - 1}{ke^{\varepsilon} + (M - k)} \sum_{m \in \text{top-}k} \sqrt{\frac{M}{M-1}} a_m\right]. \tag{66}$$

If we focus on the first coordinate of (66), we have

$$r_k \times \frac{e^{\varepsilon} - 1}{ke^{\varepsilon} + (M - k)} \sqrt{\frac{M}{M-1}} C_k = 1, \tag{67}$$

which concludes the proof. □

H. Additional Experimental Results

In Figure 2, we provide additional empirical results by sweeping the number of users n from $2,000$ to $10,000$ on the left and sweeping the dimension d from 200 to $1,000$ on the right.

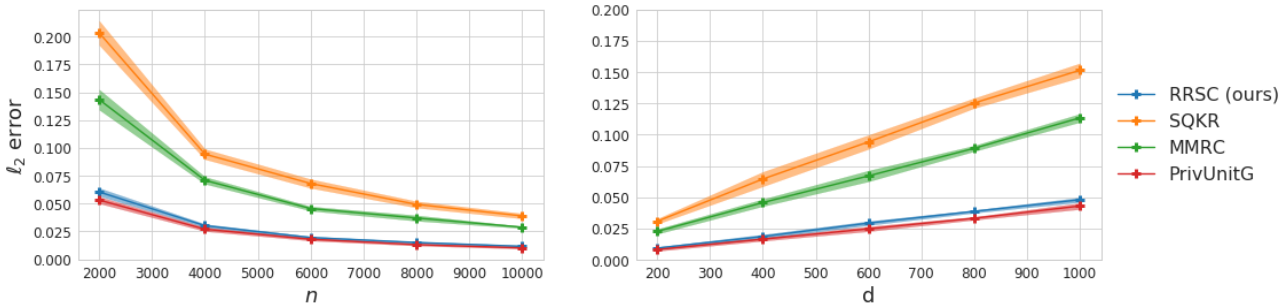

Figure 2: Comparison of `RRSC` with `SQKR` (Chen et al., 2020), `MMRC` (Shah et al., 2022), and `PrivUnitG` (Asi et al., 2022). **(left)** ℓ_2 error vs number of users n with $d = 500$, $\varepsilon = 6$, and the number of bits is $b = \varepsilon = 6$. $k = 1$ for each n. **(right)** ℓ_2 error vs dimension d for $n = 5000$, $\varepsilon = 6$, and the number of bits is $b = \varepsilon = 6$. $k = 1$ for for each d.

I. Additional Details on Prior LDP Schemes

For completeness, we provide additional details on prior LDP mean estimation schemes in this section, including `PrivUnit` (Bhowmick et al., 2018), `SQKR` (Chen et al., 2020), `FT21` (Feldman & Talwar, 2021), and `MMRC` (Shah et al., 2022).

I.1. `PrivUnit` (Bhowmick et al., 2018)

(Asi et al., 2022) considered the mean estimation problem under DP constraint (without communication constraint) when $\mathcal{X} = \mathbb{S}^{d-1} = \{v \in \mathbb{R}^d : \|v\|_1 = 1\}$. Since there is no communication constraint, they assumed canonical protocol where the random encoder is $f : \mathbb{S}^{d-1} \to \mathbb{R}^d$ and the decoder is a simple additive aggregator

$$g_n(f(v_1), \ldots, f(v_n)) = \frac{1}{n} \sum_{i=1}^{n} f(v_i).$$

The authors showed that PrivUnit is an exact optimal among the family of unbiased locally private procedures.

Recall that given an input vector $v \in \mathbb{S}^{d-1}$, the local randomized PrivUnit(p, q) has the following distribution up to normalization:

$$\text{PrivUnit}(p, q) \sim \begin{cases} Z | \langle Z, v \rangle \geq \gamma & \text{w.p. } p \\ Z | \langle Z, v \rangle < \gamma & \text{w.p. } 1 - p \end{cases}$$

where Z has a uniform distribution on $\mathbb{S}^{d-1}$. Let S_γ be the surface area of hypersphere cap $\{z \in \mathbb{S}^{d-1} | \langle z, v \rangle \geq \gamma\}$, with S_{-1} representing the surface area of the d dimensional hypersphere. We denoted $q = \Pr[Z_1 \leq \gamma] = (S_{-1} - S_\gamma)/S_{-1}$ (convention from (Bhowmick et al., 2018; Asi et al., 2022)). The normalization factor is required to obtain unbiasedness.

(Asi et al., 2022) also introduced `PrivUnitG`, which is a Gaussian approximation of `PrivUnit`. In this approach, Z is sampled from an i.i.d. $\mathcal{N}(0, 1/d)$ distribution. This simplifies the process of determining more accurate parameters p, q, and γ. Consequently, in practical applications, `PrivUnitG` surpasses `PrivUnit` in performance owing to superior parameter optimization.

I.2. `SQKR` (Chen et al., 2020)

Next, we outline the encoder and decoder of SQKR in this section. The encoding function mainly consists of three steps: (1) computing Kashin's representation, (2) quantization, and (3) sampling and privatization.

Compute Kashin's representation A tight frame is a set of vectors $\{u_j\}_{j=1}^N \in \mathbb{R}^d$ that satisfy Parseval's identity, i.e. $\|v\|_2^2 = \sum_{j=1}^N \langle u_j, v\rangle^2$ for all $v \in \mathbb{R}^d$. We say that the expansion $v = \sum_{j=1}^N a_j u_j$ is a Kashin's representation of x at level K if $\max_j |a_j| \leq \frac{K}{\sqrt{N}} \|v\|_2$ (Kashin, 1977). Lyubarskii & Vershynin (2010) shows that if $N > (1+\mu)\, d$ for some $\mu > 0$, then there exists a tight frame $\{u_j\}_{j=1}^N$ such that for any $x \in \mathbb{R}^d$, one can find a Kashin's representation at level $K = \Theta(1)$. This implies that we can represent the local vector v with coefficients $\{a_j\}_{j=1}^N \in [-c/\sqrt{d}, c/\sqrt{d}]^N$ for some constants c and $N = \Theta(d)$.

Quantization In the quantization step, each client quantizes each a_j into a 1-bit message $q_j \in \left\{-c/\sqrt{d}, c/\sqrt{d}\right\}$ with $\mathbb{E}[q_j] = a_j$. This yields an unbiased estimator of $\{a_j\}_{j=1}^N$, which can be described in $N = \Theta(d)$ bits. Moreover, due to the small range of each a_j, the variance of q_j is bounded by $O(1/d)$.

Sampling and privatization To further reduce $\{q_j\}$ to $k = \min(\lceil \varepsilon \rceil, b)$ bits, client i draws k independent samples from $\{q_j\}_{j=1}^N$ with the help of shared randomness, and privatizes its k bits message via 2^k-RR mechanism (Warner, 1965), yielding the final privatized report of k bits, which it sends to the server.

Upon receiving the report from client i, the server can construct unbiased estimators $\hat{a}_j$ for each $\{a_j\}_{j=1}^N$, and hence reconstruct $\hat{v} = \sum_{j=1}^N \hat{a}_j u_j$, which yields an unbiased estimator of v. In (Chen et al., 2020), it is shown that the variance of $\hat{v}$ can be controlled by $O\left(d/\min\left(\varepsilon^2, \varepsilon, b\right)\right)$.

I.3. `FT21` (Feldman & Talwar, 2021) and `MMRC` (Shah et al., 2022)

Both `FT21` and `MMRC` aim to simulate a given ε-LDP scheme. More concretely, consider an ε-LDP mechanism $q(\cdot|v)$ that we wish to compress, which in our case, `PrivUnit`. A number of candidates $u_1, \cdots, u_N$ are drawn from a fixed reference distribution $p(u)$ (known to both the client and the server), which in our case, uniform distribution on the sphere $\mathbb{S}^{d-1}$. Under `FT21` (Feldman & Talwar, 2021), these candidates are generated from an (exponentially strong) PRG, with seed length $\ell = \mathsf{polylog}(d)$. The client then performs rejection sampling and sends the seed of the sampled candidates to the server. See Algorithm 2 for an illustration.

Algorithm 2 Simulating LDP mechanisms via rejection sampling (Feldman & Talwar, 2021)

Inputs: ε-LDP mechanism $q(\cdot|v)$, ref. distribution $p(\cdot)$, seeded PRG $G : \{0,1\}^\ell \to \{0,1\}^t$, failure probability $\gamma > 0$.

$J = e^\varepsilon \ln(1/\gamma)$.
for $j \in \{1, \cdots, J\}$ **do**
  Sample a random seed $s \in \{0,1\}^\ell$.
  Draw $u \leftarrow p(\cdot)$ using the PRG G and the random seed s.
  Sample b from Bernoulli $\left(\frac{q(u|v)}{e^\varepsilon \cdot p(u)}\right)$.
  if $b = 1$ **then**
    BREAK
  end if
end for

Output: s

On the other hand, under `MRC` (Shah et al., 2022) the LDP mechanism is simulated via a minimal random coding technique (Havasi et al., 2019). Specifically, the candidates are generated via shared randomness, and the client performs an importance sampling and sends the index of the sampled one to the server, as illustrated in Algorithm 3. It can be shown that when the target mechanism is ε-LDP, the communication costs of both strategies are $\Theta(\varepsilon)$ bits. It is also worth noting that both strategies will incur some bias (though the bias can be made exponentially small as one increases the communication cost), and (Shah et al., 2022) provides a way to correct the bias when the target mechanism is `PrivUnit` (or general cap-based mechanisms).

Algorithm 3 Simulating LDP mechanisms via importance sampling (Shah et al., 2022)

Inputs: ε-LDP mechanism $q(\cdot|v)$, ref. distribution $p(\cdot)$, # of candidates M

Draw samples $u_1, \cdots, u_M$ from $p(u)$ using the shared source of randomness.
for $k \in \{1, \cdots, M\}$ **do**
  $w(k) \leftarrow q(u_k|v)/p(u_k)$.
end for
$\pi_{\texttt{MRC}}(\cdot) \leftarrow w(\cdot)/\sum_k w(k)$.
Draw $k^* \leftarrow \pi_{\texttt{MRC}}$.

Output: k^*

J. Additional Discussion

Convergence to PrivUnit As the communication constraint b increases, the *exact*-optimal scheme with communication constraint should coincide with the *exact*-optimal scheme *without* communication constraint, which is `PrivUnit`. Note that the rotationally symmetric simplex can be defined only when $M = 2^b < d$, due to its simplex structure. However, we have a natural extension where the codebook is a collection of M (nearly) maximally separated vectors on the sphere of radius r, where we can assume that M codewords are uniformly distributed on the sphere of radius r_k if M is large enough. Consider the case where $q = \frac{k}{M}$ is fixed and $M = 2^b$ is large. Since the k-closest encoding yields an unbiased scheme with error $\mathrm{Err}(f, P_{U^M}) = r_k^2 - 1$, where r_k is normalizing constant, for uniformly distributed M codewords on the sphere, the constant r_k should satisfy

$$r_k \times \frac{e^\varepsilon - 1}{ke^\varepsilon + (M-k)} \mathbb{E}\left[\sum_{m \in \text{top-}k} U_{m,1}\right] = 1 \tag{68}$$

where $U_{m,1}$ is the first coordinate of uniformly drawn U_m from the unit sphere $\mathbb{S}^{d-1}$. Then, as M increases, $U_{m,1}$ being one of the top-k becomes equivalent to $U_{m,1} > \gamma$, where γ is the threshold such that $\Pr[U_{m,1} > \gamma] = q$. Hence, assigning higher probabilities to the top-k closest codewords becomes equivalent to assigning high probabilities to the codewords with $\langle U_m, e_1 \rangle > \gamma$ where $v = e_1$. This is essentially how `PrivUnit` operates.

Complexity of `RRSC` Each user has $d \times d$ orthonormal matrix A and needs to find k smallest $\langle v, As_m \rangle$ for $1 \le m \le M$. Since $\langle v, As_m \rangle = \langle A^\intercal v, s_m \rangle$, it requires $O(d^2)$ to compute $A^\intercal v$ and additional $O(Md)$ to compute all inner products for $1 \le m \le M$. However, if $M \ll d$, we have a simpler equivalent scheme using

$$\langle A^\intercal v, s_m \rangle = \sqrt{\frac{M}{M-1}} a_m^\intercal v - \sum_{i=1}^{M} a_i^\intercal v \frac{1}{\sqrt{M(M-1)}}, \tag{69}$$

where $a_m^\intercal$ is the m-th row of the matrix A. Then, it only requires storing the first M rows of the matrix and $O(Md)$ to obtain all inner products in (69) by avoiding $O(d^2)$ to construct $A^\intercal v$.

On the other hand, the server computes As_m upon receiving a message m. The corresponding time complexity is $O(Md)$ (per user) since s_m has M non-zero values. We note that both `MMRC` (Shah et al., 2022) and `FT21` (Feldman & Talwar, 2021) require the same encoding complexity $O(Md)$ as `RRSC`, where they choose $M = O(\exp(\varepsilon))$.

Unified Framework It turns out that `SQKR` (Chen et al., 2020), `FT21` (Feldman & Talwar, 2021) and `MMRC` (Shah et al., 2022) can be viewed as special cases in our framework. Specifically, `SQKR` (Chen et al., 2020) uses Kashin's representation of $v = \sum_{j=1}^{N} a_j u_j$, where $\{a_j\}_{j=1}^N \in [-c/\sqrt{d}, c/\sqrt{d}]$ for some $(1+\mu)d$ with $\mu > 0$ and $c > 0$. Then the `SQKR` encoder quantizes each a_j into a 1-bit message q_i, and draws k samples with the help of shared randomness. This can be viewed as random coding with a codebook-generating distribution. More concretely, the corresponding codebook U^M consists of k non-zero values of $\pm c/\sqrt{d}$ where the randomness is from selecting k indices using shared randomness. On the other hand, since `MMRC` (Shah et al., 2022) is simulating the channel corresponding to a privacy mechanism, it can be viewed as pre-generating random codewords U^M according to the reference distribution, where the importance sampling is also a way of assigning probabilities to each codeword. As elaborated earlier, it is observed that with an increase in the communication constraint b, the suggested k-closest encoding gradually transforms into a threshold-based encoding, analogous to that of `MMRC`. The

codebook associated with `FT21` (Feldman & Talwar, 2021) depends on the PRG it uses. Let $\mathsf{PRG} : \{0,1\}^b \to \{0,1\}^{\Theta(d)}$ be a PRG that takes a b-bit seed and maps it into $\Theta(d)$ bits, where $b \ll d$, and let $g : \{0,1\}^{\Theta(d)} \to \mathbb{R}^d$. For example, if we represent each coordinate of $x \in \mathbb{R}^d$ as a 32-bit float, then $g(\cdot)$ maps the float representation of x (a 32-bit string) to x. With a PRG, `FT21` mimics `PrivUnit` by first generating a b-bit seed m, computing $g(\mathsf{PRG}(m))$, and then performing rejection sampling on the seed space. The above procedure can be treated as a special case in our framework, where the *deterministic* codebook consists of 2^b points on $\mathbb{R}^d$: $\mathcal{C}_{\mathsf{FT21}} := \left\{g(\mathsf{PRG}(m)) : m \in \{0,1\}^b\right\}$. The probabilities assigned to each codeword according to the rejection sampling are equivalent to a threshold-based assignment.

Shared randomness When $M \leq d+1$, additional randomization is required during codebook generation to achieve an unbiased scheme, as discussed in (Gandikota et al., 2021). Furthermore, both the encoder and decoder must possess this randomization information. In the proposed `RRSC` scheme, this randomization is achieved through the random rotation of the simplex code using shared randomness. However, it is possible to circumvent the need for shared randomness by having the server generate random rotation matrices using its private coin and communicate them to the users. This approach replaces shared randomness with downlink communication, which is typically more affordable than uplink communication. It should be noted that directly transmitting the rotation matrices would require $O(d^2)$ bits. Nonetheless, the server can generate them using a predetermined pseudo-random generator (PRG) and transmit only the seeds of it to the users. Drawing from a similar argument as in (Feldman & Talwar, 2021), assuming the existence of exponentially strong PRGs, seeds with $\mathsf{polylog}(d)$ bits are sufficient.

Future Work We showed the *exact*-optimality of k-closest encoding for the rotating simplex codebook. In general, it also achieves unbiasedness and the following error formulation $\mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^{M} Q_f(m|v, U^M)\|v - U_m\|^2\right]$ implies the *exact*-optimality of k-closest encoding for any rotationally symmetric codebook, which leads us to the following conjecture.

Conjecture J.1. *The proposed k-closest encoding is exact-optimal for any rotationally symmetric codebook.*

It also remains unclear whether k can depend on the realization of the codebook U^M in general, which we leave to future work. We also proved that the *exact*-optimal codebook must be rotationally symmetric. We conjecture that the maximally separated codebook (simplex codebook) is *exact*-optimal as it provides the most effective coverage of the space $\mathbb{S}^{d-1}$. This, too, is left as a topic for future work.

Conjecture J.2. *The rotationally symmetric simplex codebook is the exact-optimal codebook.*