A Proof of Lemma 3.1

Proof. For n-user mean estimation protocol $(f, \mathcal{A}, P_{U^M})$, following the notation and steps from [2, Proof of Lemma 3.1], we define the marginalized output

$$\tilde{g}_i(m_i, U_i; v^n) = \mathbb{E}_{\{m_j, U_j\}_{j \neq i}} \left[n\mathcal{A}(\{m_j, U_j\}_{j=1}^n) \;\middle|\; f_i(v_i, U_i) = m_i, U_i, v^{n \setminus i} \right]. \tag{35}$$

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}:

$$g_i(m_i, U_i) = \mathbb{E}_{v^{n \setminus i} \sim\, P_{\text{unif}}} \left[\tilde{g}_i(m_i, U_i; v^n) \right] \tag{36}$$

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

$$\hat{\mathcal{R}}_{\leq i}(\{v_j, m_j, U_j\}_{j=1}^i) = \mathbb{E}_{v_j \sim P_{\text{unif}}, j > i} \left[n\mathcal{A}(\{m_j, U_j\}_{j=1}^n) - \sum_{j=1}^i v_j \;\middle|\; \{v_j, m_j, U_j\}_{j=1}^i \right] \tag{37}$$

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

$$\begin{aligned}
&\mathbb{E}_{\{v_j, m_j, U_j\}_{j=1}^n} \left[\left\| n\mathcal{A}(\{m_j, U_j\}_{j=1}^n) - \textstyle\sum_{j=1}^n v_j \right\|^2 \right] \\
&= \mathbb{E}_{\{v_j, m_j, U_j\}_{j=1}^n} \left[\left\| \hat{\mathcal{R}}_{\leq n}(\{v_j, m_j, U_j\}_{j=1}^n) \right\|^2 \right] && (38) \\
&= \mathbb{E}_{\{v_j, m_j, U_j\}_{j=1}^n} \left[\left\| \hat{\mathcal{R}}_{\leq n}(\{v_j, m_j, U_j\}_{j=1}^n) - \hat{\mathcal{R}}_{\leq n-1}(\{v_j, m_j, U_j\}_{j=1}^{n-1}) + \hat{\mathcal{R}}_{\leq n-1}(\{v_j, m_j, U_j\}_{j=1}^{n-1}) \right\|^2 \right] && (39) \\
&= \mathbb{E}_{\{v_j, m_j, U_j\}_{j=1}^n} \left[\left\| \hat{\mathcal{R}}_{\leq n}(\{v_j, m_j, U_j\}_{j=1}^n) - \hat{\mathcal{R}}_{\leq n-1}(\{v_j, m_j, U_j\}_{j=1}^{n-1}) \right\|^2 \right] \\
&\quad + \mathbb{E}_{\{v_j, m_j, U_j\}_{j=1}^{n-1}} \left[\left\| \hat{\mathcal{R}}_{\leq n-1}(\{v_j, m_j, U_j\}_{j=1}^{n-1}) \right\|^2 \right] && (40) \\
&= \sum_{i=1}^n \mathbb{E}_{\{v_j, m_j, U_j\}_{j=1}^i} \left[\left\| \hat{\mathcal{R}}_{\leq i}(\{v_j, m_j, U_j\}_{i=1}^n) - \hat{\mathcal{R}}_{\leq i-1}(\{v_j, m_j, U_j\}_{j=1}^{i-1}) \right\|^2 \right] && (41) \\
&\geq \sum_{i=1}^n \mathbb{E}_{m_i, U_i} \left[\left\| \mathbb{E}_{\{v_j, m_j, U_j\}_{j=1}^{i-1}} \left[\hat{\mathcal{R}}_{\leq i}(\{v_j, m_j, U_j\}_{i=1}^n) - \hat{\mathcal{R}}_{\leq i-1}(\{v_j, m_j, U_j\}_{j=1}^{i-1}) \right] \right\|^2 \right] && (42) \\
&= \sum_{i=1}^n \mathbb{E}_{m_i, U_i} \left[\| g_i(m_i, U_i) - v_i \|^2 \right]. && (43)
\end{aligned}$$

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

$$\begin{aligned}
&\sup_{v_1, \ldots, v_n} \mathbb{E}_{\{m_j, U_j\}_{j=1}^n} \left[\left\| n\mathcal{A}(\{m_j, U_j\}_{j=1}^n) - \sum_{j=1}^n v_j \right\|^2 \right] \\
&\geq \mathbb{E}_{\{v_j, m_j, U_j\}_{j=1}^n} \left[\left\| n\mathcal{A}(\{m_j, U_j\}_{j=1}^n) - \sum_{j=1}^n v_j \right\|^2 \right] && (44) \\
&= \sum_{i=1}^n \mathbb{E}_{v_i, m_i, U_i} \left[\| g_i(m_i, U_i) - v_i \|^2 \right] && (45) \\
&= \sum_{i=1}^n \sup_{v_i} \mathbb{E}_{m_i, U_i} \left[\| g_i(m_i, U_i) - v_i \|^2 \right] && (46)
\end{aligned}$$

where the last equality (46) is from Lemma 3.2, Lemma 3.4, and Lemma 3.5. Thus, the user-specific decoder achieves lower MSE:

$$\mathrm{Err}_n(f, \mathcal{A}, P_{U^n}) \geq \frac{1}{n} \sum_{i=1}^{n} \mathrm{Err}_1(f_i, g_i, P_{U_i}). \tag{47}$$

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

B 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., $\{u : \tilde{u}_m = g(m, u) \text{ for all } m\}$ 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

$$Q_{f_0}(m|v, \tilde{u}^M) = Q_f(m|v, g^{-1}(\tilde{u}^M)). \tag{48}$$

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

$$\begin{aligned}
D(v, f_0, g^+, P_{\tilde{U}^M}) =& \mathbb{E}_{f_0, P_{\tilde{U}^M}} \left[\|g^+(f_0(v, \tilde{U}^M), \tilde{U}^M) - v\|^2 \right] && (49)\\
=& \mathbb{E}_{P_{\tilde{U}^M}} \left[\textstyle\sum_{m=1}^{M} Q_{f_0}(m|v, \tilde{U}^M) \|\tilde{U}_m - v\|^2 \right] && (50)\\
=& \mathbb{E}_{f, P_U} \left[\textstyle\sum_{m=1}^{M} Q_f(m|v, U) \|g(m, U) - v\|^2 \right] && (51)\\
=& \mathbb{E}_{f, P_U} \left[\|g(f(v, U), U) - v\|^2 \right] && (52)\\
=& D(v, f, g, P_U). && (53)
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}_{P_{\tilde{U}^M}} \left[g^+(f_0(v, \tilde{U}^M), \tilde{U}^M) \right] =& \mathbb{E}_{f_0, P_{\tilde{U}^M}} \left[\textstyle\sum_{m=1}^{M} Q_{f_0}(m|v, \tilde{U}^M) \tilde{U}_m \right] && (54)\\
=& \mathbb{E}_{f, P_U} \left[\textstyle\sum_{m=1}^{M} Q_f(m|v, U) g(m, U) \right] && (55)\\
=& \mathbb{E}_{f, P_U} \left[g(f(v, U), U) \right] && (56)\\
=& v. && (57)
\end{aligned}$$

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

$$\sum_{m=1}^{M} Q_{f_0}(m|v, \tilde{u}^M) = \sum_{m=1}^{M} Q_f(m|v, g^{-1}(\tilde{u}^M)) = 1 \tag{58}$$

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

C 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

$$Q_{f_1}(m|v, \bar{U}^M) = \mathbb{E}_A \left[Q_f(m|Av, A\bar{U}^M) | \bar{U}^M \right]. \tag{59}$$

Then, Q_{f_1} is a valid probability since

$$\sum_{m=1}^{M} Q_{f_1}(m|v, \bar{U}^M) = \mathbb{E}_A \left[\sum_{m=1}^{M} Q_f(m|Av, A\bar{U}^M) | \bar{U}^M \right] = 1. \tag{60}$$

Also, we have

$$\frac{Q_{f_1}(m|v, \bar{U}^M)}{Q_{f_1}(m|v', \bar{U}^M)} = \frac{\mathbb{E}_A \left[Q_f(m|Av, A\bar{U}^M) | \bar{U}^M \right]}{\mathbb{E}_A \left[Q_f(m|Av', A\bar{U}^M) | \bar{U}^M \right]} \tag{61}$$

$$\leq \frac{\mathbb{E}_A\left[e^{\varepsilon} Q_f(m|Av', A\bar{U}^M)|\bar{U}^M\right]}{\mathbb{E}_A\left[Q_f(m|Av', A\bar{U}^M)|\bar{U}^M\right]} \tag{62}$$

$$= e^{\varepsilon}. \tag{63}$$

Finally, we need to check unbiasedness.

$$\mathbb{E}_{P_{\bar{U}^M}}\left[Q_{f_1}(m|v, \bar{U}^M)\bar{U}_m\right] = \mathbb{E}_{A,P_{U^M}}\left[\sum_{m=1}^M Q_f(m|Av, A\bar{U}^M)\bar{U}_m\right] \tag{64}$$

$$= \mathbb{E}_{A,P_{U^M}}\left[\sum_{m=1}^M Q_f(m|Av, U^M)A^\intercal U_m\right] \tag{65}$$

$$= \mathbb{E}_A\left[A^\intercal \mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^M Q_f(m|Av, U^M)U_m\right]\right] \tag{66}$$

$$= \mathbb{E}_A\left[A^\intercal Av\right] \tag{67}$$

$$= v. \tag{68}$$

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

$$\mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^M Q_f(m|Av, U^M)U_m\right] = Av \tag{69}$$

for all A.

Now, we are ready to prove the main inequality.

$$\mathrm{Err}(f, P_{U^M}) = \sup_v D(v, f, P_{U^M}) \tag{70}$$

$$\geq \mathbb{E}_A\left[D(Av, f, P_{U^M})\right] \tag{71}$$

$$= \mathbb{E}_A\left[\mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^M Q_f(m|Av, U^M)\|U_m - Av\|^2\right]\right] \tag{72}$$

$$= \mathbb{E}_{P_{U^M},A}\left[\sum_{m=1}^M Q_f(m|Av, A\bar{U}^M)\|\bar{U}_m - v\|^2\right] \tag{73}$$

$$= \mathbb{E}_{P_{\bar{U}^M}}\left[\sum_{m=1}^M \mathbb{E}_A\left[Q_f(m|Av, A\bar{U}^M)|\bar{U}^M\right]\|\bar{U}_m - v\|^2\right] \tag{74}$$

$$= \mathbb{E}_{P_{\bar{U}^M}}\left[\sum_{m=1}^M Q_{f_1}(m|v, \bar{U}^M)\|\bar{U}_m - v\|^2\right] \tag{75}$$

$$= D(v, f_1, P_{\bar{U}^M}). \tag{76}$$

for all v. This concludes the proof. □

D 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

$$f_2(v, U^M) = f(Av, AU^M) \tag{77}$$

for uniform random orthonormal matrix. Then,

$$Q_{f_2}(m|v, U^M) = \mathbb{E}_A\left[Q_f(m|Av, AU^M)\right]. \tag{78}$$

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

$$D(v', f_2, P_{U^M}) = D(A_0 v, f_2, P_{U^M}) = D(v, f_2, P_{U^M}). \tag{79}$$

Compared to a given randomizer f, we have

$$\mathrm{Err}(f, P_{U^M}) \geq \mathbb{E}_A\left[D(Av, f, P_{U^M})\right] \tag{80}$$

$$= \mathbb{E}_{A,P_{U^M}}\left[\sum_{m=1}^M Q_f(m|Av, U^M)\|Av - U^M\|^2\right] \tag{81}$$

$$= \mathbb{E}_{A,P_{U^M}}\left[\sum_{m=1}^M Q_f(m|Av, U^M)\|v - A^\intercal U^M\|^2\right] \tag{82}$$

$$= \mathbb{E}_{A,P_{U^M}}\left[\sum_{m=1}^M Q_f(m|Av, AU^M)\|v - U^M\|^2\right] \tag{83}$$

$$= \mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^M \mathbb{E}_A\left[Q_f(m|Av, AU^M)\right]\|v - U^M\|^2\right] \tag{84}$$

$$= D(v, f_2, P_{U^M}) \tag{85}$$

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

E 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

$$Q_f(m_1|v_1, rA_1 s^M) = Q_f(m_2|\tilde{A}v_1, r\tilde{A}A_1 s^M). \tag{86}$$

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

$$Q_f(m_2|\tilde{A}v_1, r\tilde{A}A_1 s^M) = Q_f(m_2|A'\tilde{A}v_1, rA'\tilde{A}A_1 s^M) \tag{87}$$

$$= Q_f(m_2|A'\tilde{A}v_1, rA_2 s^M) \tag{88}$$

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

$$Q_f(m_2|v_1', rA_2 s^M) = Q_f(m_1|v_1, rA_1 s^M) \tag{89}$$

$$> e^{\varepsilon} Q_f(m_2|v_2, rA_2 s^M), \tag{90}$$

which contradicts the ε-LDP constraint.

For an unbiased encoder, the error is

$$\mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^M \|U_m - v\|^2 Q_f(m|v, U^M)\right] = \mathbb{E}_{P_{U^M}}\left[\sum_{m=1}^M \|U_m\|^2 Q_f(m|v, U^M)\right] - 1 \tag{91}$$

$$= r^2 - 1. \tag{92}$$

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

$$\mathbb{E}_A\left[\sum_{m=1}^M rAs_m Q_f(m|e_1, rAs^M)\right] = e_1, \tag{93}$$

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

$$r \times \mathbb{E}_a\left[\sum_{m=1}^M a^{\intercal} s_m Q_f(m|e_1, rAs^M)\right] = 1, \tag{94}$$

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

$$(M - \lfloor k \rfloor - 1) q_0 + q_i + \lfloor k \rfloor e^{\varepsilon} q_0 = 1, \tag{95}$$

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. □

F Proof of Lemma 3.7

Proof. Following (28) with $U_m = As_m$ and $v = e_1$, we have

$$r_k \frac{e^{\varepsilon} - 1}{ke^{\varepsilon} + (M - k)} \mathbb{E}\left[\sum_{m \in T_k(e_1, A \cdot S)} A \cdot s_m\right]$$

$$
\begin{aligned}
&= r_k \frac{e^{\varepsilon}-1}{ke^{\varepsilon}+(M-k)} \mathbb{E}\left[\sum_{m\in \arg\max_k(\{\langle e_1, As_1\rangle, \ldots, \langle e_1, As_M\rangle\})} A\cdot s_m\right] \\
&= e_1.
\end{aligned}
$$

By focusing on the first coordinate of the above equation and observing that $\langle e_1, As_M\rangle = \langle a, s_m\rangle$ where a is the first row of the rotation matrix A, we must have

$$
r_k \cdot \frac{e^{\varepsilon}-1}{ke^{\varepsilon}+(M-k)} \mathbf{E}_{a\sim \mathsf{unif}(\mathbf{S}^{d-1})}\left[\sum_{m\in \mathsf{Top}_k(\{\langle a, s_1\rangle, \ldots, \langle a, s_M\rangle\})} \langle a, s_m\rangle\right] = 1. \tag{96}
$$

Note that since A is a random orthogonal matrix drawn from the Haar measure on $SO(d)$, a is distributed uniformly over the unit sphere $\mathbb{S}^{d-1}$.

Next, observe that by definition,

$$
s_m = \frac{M}{\sqrt{M(M-1)}} e_m - \frac{1}{\sqrt{M(M-1)}} \mathbf{1}_M,
$$

where $\mathbf{1}_M = (\underbrace{1, 1, \ldots, 1}_{M \text{entries}}, 0, \ldots, 0) \in \{0,1\}^d$ (that is, $(\mathbf{1}_M)_m = \mathbb{1}_{\{m\leq M\}}$). Therefore,

$$
\langle a, s_m\rangle = \frac{M}{\sqrt{M(M-1)}} a_m - \frac{1}{\sqrt{M(M-1)}} \langle a, \mathbf{1}_M\rangle,
$$

and hence plugging in (96) yields

$$
\begin{aligned}
& r_k \cdot \frac{e^{\varepsilon}-1}{ke^{\varepsilon}+(M-k)} \mathbf{E}_{a\sim \mathsf{unif}(\mathbf{S}^{d-1})}\left[\sum_{m\in \mathsf{Top}_k(\{\langle a, s_1\rangle, \ldots, \langle a, s_M\rangle\})} \langle a, s_m\rangle\right] \\
=& r_k \cdot \frac{e^{\varepsilon}-1}{ke^{\varepsilon}+(M-k)} \cdot \frac{M}{\sqrt{M(M-1)}} \mathbf{E}_{a\sim \mathsf{unif}(\mathbf{S}^{d-1})}\left[\sum_{i=1}^k a_{(i|M)} - \frac{k}{M}\langle a, \mathbf{1}_M\rangle\right] \\
=& r_k \cdot \frac{e^{\varepsilon}-1}{ke^{\varepsilon}+(M-k)} \cdot \sqrt{\frac{M}{M-1}} \cdot \underbrace{\mathbf{E}_{a\sim \mathsf{unif}(\mathbf{S}^{d-1})}\left[\sum_{i=1}^k a_{(i|M)}\right]}_{:=C_k},
\end{aligned}
$$

where (1) $a_{(i|M)}$ denotes the i-th largest entry of the first M coordinates of a and (2) the last equality holds since a is uniformly distributed over $\mathbb{S}^{d-1}$. □

G Proof of Lemma 3.8

Proof. First of all, observe that

$$
\begin{aligned}
& \mathbb{E}_{a\sim \mathsf{unif}(\mathbb{S}^{d-1})}\left[\sum_{i=1}^k a_{(i|M)}\right] \\
&= \mathbb{E}_{a\sim \mathsf{unif}(\mathbb{S}^{d-1})}\left[\mathbb{E}\left[\sum_{i=1}^k a_{(i|M)} \middle| \sum_{i=1}^M a_i^2\right]\right] \\
&\overset{\text{(a)}}{=} \mathbb{E}_{a\sim \mathsf{unif}(\mathbb{S}^{d-1})}\left[\sqrt{\sum_{i=1}^M a_i^2} \cdot \mathbb{E}_{(a'_1, \ldots, a'_M)\sim \mathsf{unif}(\mathbb{S}^{M-1})}\left[\sum_{i=1}^k a'_{(i)}\right]\right] \\
&= \underbrace{\mathbb{E}_{a\sim \mathsf{unif}(\mathbb{S}^{d-1})}\left[\sqrt{\sum_{i=1}^M a_i^2}\right]}_{\text{(i)}} \cdot \underbrace{\mathbb{E}_{(a'_1, \ldots, a'_M)\sim \mathsf{unif}(\mathbb{S}^{M-1})}\left[\sum_{i=1}^k a'_{(i)}\right]}_{\text{(ii)}},
\end{aligned}
$$

where (a) holds due to the spherical symmetry of a. Next, we bound (i) and (ii) separately.

Claim G.1 (Bounding (i))**.** *For any $d \geq M > 2$, it holds that*

$$\sqrt{\frac{M-2}{d-2}} \leq \mathbb{E}_{a\sim \mathsf{unif}(\mathbb{S}^{d-1})}\left[\sqrt{\sum_{i=1}^{M} a_i^2}\right] \leq \sqrt{\frac{M}{d-2}}. \tag{97}$$

Proof of Claim G.1. Observe that when a is distributed uniformly over $\mathbb{S}^{d-1}$, it holds that

$$(a_1, a_2, ..., a_d) \stackrel{d}{=} \left(\frac{Z_1}{\sqrt{\sum_{i=1}^d Z_i^2}}, \frac{Z_2}{\sqrt{\sum_{i=1}^d Z_i^2}}, ..., \frac{Z_d}{\sqrt{\sum_{i=1}^d Z_i^2}}\right),$$

where $A \stackrel{d}{=} B$ denotes A and B have the same distribution, and $Z_1, ..., Z_d \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. As a result, we must have

$$\mathbb{E}_{a\sim \mathsf{unif}(\mathbb{S}^{d-1})}\left[\sqrt{\sum_{i=1}^{M} a_i^2}\right] = \mathbb{E}_{Z_1,...,Z_M \stackrel{\text{i.i.d.}}{\sim}\mathcal{N}(0,1)}\left[\sqrt{\frac{\sum_{i=1}^M Z_i^2}{\sum_{i=1}^M Z_i^2 + \sum_{i'=M+1}^d Z_{i'}^2}}\right].$$

By Jensen's inequality, it holds that

$$\begin{aligned}
&\mathbb{E}_{Z_1,...,Z_M \stackrel{\text{i.i.d.}}{\sim}\mathcal{N}(0,1)}\left[\sqrt{\frac{\sum_{i=1}^M Z_i^2}{\sum_{i=1}^M Z_i^2 + \sum_{i'=M+1}^d Z_{i'}^2}}\right] \\
&= \mathbb{E}_{Z_1,...,Z_M \stackrel{\text{i.i.d.}}{\sim}\mathcal{N}(0,1)}\left[\sqrt{\frac{1}{1 + \frac{\sum_{i'=M+1}^d Z_{i'}^2}{\sum_{i=1}^M Z_i^2}}}\right] \\
&\stackrel{\text{(a)}}{\geq} \sqrt{\frac{1}{1 + \mathbb{E}_{Z_1,...,Z_M \stackrel{\text{i.i.d.}}{\sim}\mathcal{N}(0,1)}\left[\frac{\sum_{i'=M+1}^d Z_{i'}^2}{\sum_{i=1}^M Z_i^2}\right]}} \\
&\stackrel{\text{(b)}}{=} \sqrt{\frac{1}{1 + \frac{d-M}{M-2}}} \\
&= \sqrt{\frac{M-2}{d-2}},
\end{aligned}$$

where (a) holds since $x \mapsto \sqrt{1/(1+x)}$ is a convex mapping for $x > 0$, and (b) holds due to the fact that $\sum_i Z_i^2$ follows from a χ^2 distribution and that the ratio of two independent χ^2 random variables follows an F-distribution.

On the other hand, it also holds that

$$\begin{aligned}
&\mathbb{E}_{Z_1,...,Z_M \stackrel{\text{i.i.d.}}{\sim}\mathcal{N}(0,1)}\left[\sqrt{\frac{\sum_{i=1}^M Z_i^2}{\sum_{i=1}^M Z_i^2 + \sum_{i'=M+1}^d Z_{i'}^2}}\right] \\
&\stackrel{\text{(a)}}{\leq} \sqrt{\mathbb{E}_{Z_1,...,Z_M \stackrel{\text{i.i.d.}}{\sim}\mathcal{N}(0,1)}\left[\frac{\sum_{i=1}^M Z_i^2}{\sum_{i=1}^M Z_i^2 + \sum_{i'=M+1}^d Z_{i'}^2}\right]} \\
&= \sqrt{\mathbb{E}_{Z_1,...,Z_M \stackrel{\text{i.i.d.}}{\sim}\mathcal{N}(0,1)}\left[1 - \frac{\sum_{i=M+1}^d Z_{i'}^2}{\sum_{i=1}^M Z_i^2 + \sum_{i'=M+1}^d Z_{i'}^2}\right]} \\
&= \sqrt{1 - \mathbb{E}_{Z_1,...,Z_M \stackrel{\text{i.i.d.}}{\sim}\mathcal{N}(0,1)}\left[\frac{1}{1 + \frac{\sum_{i=1}^M Z_i^2}{\sum_{i=M+1}^d Z_{i'}^2}}\right]}
\end{aligned}$$

$$\overset{(b)}{\leq} \sqrt{1-\frac{1}{1+\mathbb{E}_{Z_1,...,Z_M\overset{\text{i.i.d.}}{\sim}\mathcal{N}(0,1)}\left[\frac{\sum_{i=1}^M Z_i^2}{\sum_{i=M+1}^d Z_{i'}^2}\right]}}$$

$$\overset{(c)}{=} \sqrt{1-\frac{1}{1+\frac{M}{d-M-2}}}$$

$$= \sqrt{\frac{M}{d-2}},$$

where (a) holds since $\sqrt{\cdot}$ is concave, (b) holds since $x \mapsto \frac{1}{1+x}$ is convex, and (c) again is due to the fact that the ratio of two independent χ^2 random variables follows an F-distribution.

Claim G.2 (Bounding (ii))**.** *As long as*

- $k \geq 400 \cdot \log 10$,
- $\log(M/k) \geq \left(\frac{10^3 \pi \log 2}{9}\right)^2$,

it holds that

$$\sqrt{\frac{k\log\left(\frac{M}{k}\right)}{24\pi\log 2M}} \leq \mathbb{E}_{(a_1',...,a_M')\sim\mathsf{unif}(\mathbb{S}^{M-1})}\left[\sum_{i=1}^k a_{(i)}'\right] \leq \sqrt{\frac{4k\log M}{M}}. \tag{98}$$

Proof of Claim G.2. We start by re-writing a':

$$(a_1', a_2', ..., a_M') \overset{d}{=} \left(\frac{Z_1}{\sqrt{\sum_{i=1}^M Z_i}}, \frac{Z_2}{\sqrt{\sum_{i=1}^M Z_i}}, ..., \frac{Z_M}{\sqrt{\sum_{i=1}^M Z_i}}\right).$$

This yields that

$$(a_{(1)}', a_{(2)}', ..., a_{(k)}') \overset{d}{=} \left(\frac{Z_{(1)}}{\sqrt{\sum_{i=1}^M Z_i^2}}, \frac{Z_{(2)}}{\sqrt{\sum_{i=1}^M Z_i^2}}, ..., \frac{Z_{(k)}}{\sqrt{\sum_{i=1}^M Z_i^2}}\right),$$

and hence

$$\mathbb{E}_{(a_1',...,a_M')\sim\mathsf{unif}(\mathbb{S}^{M-1})}\left[\sum_{i=1}^k a_{(i)}'\right] = \mathbb{E}_{Z_1,...,Z_M\overset{\text{i.i.d.}}{\sim}\mathcal{N}(0,1)}\left[\frac{1}{\sqrt{\sum_{i=1}^M Z_i^2}}\sum_{i=1}^k Z_{(i)}\right].$$

Upper bound. To upper bound the above, observe that

$$\mathbb{E}_{Z_1,...,Z_M\overset{\text{i.i.d.}}{\sim}\mathcal{N}(0,1)}\left[\frac{1}{\sqrt{\sum_{i=1}^M Z_i^2}}\sum_{i=1}^k Z_{(i)}\right] \leq k\mathbb{E}_{Z_1,...,Z_M\overset{\text{i.i.d.}}{\sim}\mathcal{N}(0,1)}\left[\frac{1}{\sqrt{\sum_{i=1}^M Z_i^2}}Z_{(1)}\right].$$

Let $\mathcal{E}_1 \coloneqq \left\{(Z_1,...,Z_M) \middle| \sum_{i=1}^M Z_i^2 \leq M(1-\gamma)\right\}$ where $\gamma > 0$ will be optimized later. Then it holds that

$$\Pr\{\mathcal{E}_1\} \leq e^{-\frac{M\gamma^2}{4}}. \tag{99}$$

On the other hand, the Borell-TIS inequality ensures

$$\Pr\left\{\left|Z_{(1)} - \mathbb{E}\left[Z_{(1)}\right]\right| > \xi\right\} \leq 2e^{-\frac{\xi^2}{2\sigma^2}}, \tag{100}$$

where $Z_i \sim \mathcal{N}(0, \sigma^2)$ (in our case, $\sigma = 1$). Since $\mathbb{E}\left[Z_{(1)}\right] \leq \sqrt{2\log M}$, it holds that

$$\Pr\left\{Z_{(1)} \geq \sqrt{2\log M} + \xi\right\} \leq 2e^{-\xi^2}.$$

Therefore, define $\mathcal{E}_2 \coloneqq \left\{Z_{(1)} \geq \sqrt{2\log M} + \xi\right\}$ and we obtain

$$\begin{aligned}
&\mathbb{E}_{Z_1,\dots,Z_M \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)}\left[\frac{1}{\sqrt{\sum_{i=1}^M Z_i^2}} \sum_{i=1}^k Z_{(i)}\right] \\
&\leq k\mathbb{E}_{Z_1,\dots,Z_M \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)}\left[\frac{1}{\sqrt{\sum_{i=1}^M Z_i^2}} Z_{(1)}\right] \\
&\leq k \cdot \left(\mathbb{E}\left[\frac{Z_{(1)}}{\sqrt{\sum_{i=1}^M Z_i^2}}\middle|\mathcal{E}_1 \cap \mathcal{E}_2\right] + \sup_{z_1,\dots,z_m}\left(\frac{z_{(1)}}{\sqrt{\sum_{i=1}^M z_i^2}}\right) \cdot \Pr\left(\mathcal{E}_1^c \cup \mathcal{E}_2^c\right)\right) \\
&\leq k \cdot \left(\frac{\sqrt{2\log M + \xi}}{M(1-\gamma)} + 1 \cdot \left(e^{-M\gamma^2/4} + 2e^{-\xi^2}\right)\right) \\
&\leq k \cdot \left(\frac{\sqrt{2\log M + \sqrt{\log(M)}}}{0.9 \cdot M} + 1 \cdot \left(e^{-M/400} + 2/M\right)\right) \\
&= \Theta\left(\frac{k\sqrt{\log M}}{M}\right),
\end{aligned}$$

where the last inequality holds by picking $\gamma = 0.1$ and $\xi = \sqrt{\log M}$.

Lower bound. The analysis of the lower bound is more sophisticated. To begin with, let

$$\mathcal{E}_M \coloneqq \left\{(Z_1, ..., Z_M)\middle|\sum_{i=1}^M Z_i^2 \in [M(1-\gamma), M(1+\gamma)]\right\}$$

denote the good event such that the denominator of our target is well-controlled, where $\gamma > 0$ again will be optimized later. By the concentration of χ^2 random variables, it holds that

$$\Pr\left\{\mathcal{E}_M^c\right\} \leq e^{-\frac{M}{2}(\gamma - \log(1+\gamma))} + e^{-\frac{M\gamma^2}{4}} \leq e^{-\frac{M}{2}\left(1-\frac{1}{\sqrt{1+\gamma}}\right)\gamma} + e^{-\frac{M\gamma^2}{4}} \leq 2e^{-\frac{M\gamma^2}{4}}. \tag{101}$$

Next, to lower bound $\sum_{i=1}^k Z_{(i)}$, we partition $(Z_1, Z_2, ..., Z_M)$ into k blocks $B_1, B_2, ..., B_k$ where each block contains at least $N = \lfloor M/k \rfloor$ samples: $B_j \coloneqq [(j-1) \cdot N + 1 : j \cdot N]$ for $j \in [k-1]$ and $B_k = [M] \setminus \left(\bigcup_{j=1}^{k-1} B_j\right)$. Define $\tilde{Z}_{(1)}^{(j)}$ be the maximum samples in the j-th block: $\tilde{Z}_{(1)}^{(j)} \coloneqq \max_{i \in B_j} Z_i$. Then, it is obvious that

$$\sum_{i=1}^k Z_{(i)} \geq \sum_{j=1}^k \tilde{Z}_{(1)}^{(j)}.$$

To this end, we define $\mathcal{E}_1$ to be the good event that 90% of $\tilde{Z}_{(1)}^{(j)}$'s are large enough (i.e., concentrated to the expectation):

$$\mathcal{E}_1 \coloneqq \left\{\left|\left\{j \in [k]\middle|\tilde{Z}_{(1)}^{(j)} \geq \frac{\sqrt{\log N}}{\sqrt{\pi \log 2}} - \log 100\right\}\right| > 0.9k\right\}.$$

Note that by the Borell-TIS inequality, for any $j \in [k]$,

$$\Pr\left\{\tilde{Z}_{(1)}^{(j)} \geq \frac{\sqrt{\log N}}{\sqrt{\pi \log 2}} - \xi\right\} \geq 1 - 2e^{-\xi^2},$$

so setting $\xi = \log 100$ implies $\Pr\left\{\tilde{Z}^{(j)}_{(1)} \geq \frac{\sqrt{\log N}}{\sqrt{\pi \log 2}} - \xi\right\} \geq 0.98$. Since blocks are independent with each other, applying Hoeffding's bound yields

$$\Pr\{\mathcal{E}_1\} \geq 1 - \Pr\{\mathsf{Binom}(k, 0.98) \leq 0.9\} \geq 1 - e^{-k(0.08)^2} \geq 0.9,$$

when $k \geq 400 \cdot \log 10 \geq \log 10/0.08^2$.

Next, we define a "not-too-bad" event where $\sum_{j=1}^k \tilde{Z}^{(j)}_{(1)}$ is not catastrophically small:

$$\mathcal{E}_2 \coloneqq \left\{\sum_{j=1}^k \tilde{Z}^{(j)}_{(1)} \geq -\frac{k}{\sqrt{M}}\xi\right\},$$

for some $\xi > 0$ to be optimized later. Observe that $\mathcal{E}_2$ holds with high probability:

$$\begin{aligned}\Pr\{\mathcal{E}_2\} &\overset{\text{(a)}}{\geq} \Pr\left\{\frac{k}{M}\sum_{i=1}^M Z_i \geq -\frac{k}{\sqrt{M}}\xi\right\} \\ &\overset{\text{(b)}}{\geq} 1 - e^{-\xi^2/2},\end{aligned}$$

where (a) holds since the each of the top-k values must be greater than k times the average, and (b) holds due to the Hoeffding's bound on the sum of i.i.d. Gaussian variables.

Lastly, a trivial bound implies that

$$\inf_{a \in \mathbb{S}^{M-1}} \sum_{i=1}^k a_{(i)} \geq -\frac{k}{\sqrt{M}}.$$

Now, we are ready to bound $\mathbb{E}_{Z_1,\dots,Z_M \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)}\left[\frac{1}{\sqrt{\sum_{i=1}^M Z_i^2}}\sum_{i=1}^k Z_{(i)}\right]$. We begin by decomposing it into three parts:

$$\begin{aligned}\mathbb{E}_{Z_1,\dots,Z_M \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)}\left[\frac{\sum_{i=1}^k Z_{(i)}}{\sqrt{\sum_{i=1}^M Z_i^2}}\right] &= \Pr\{\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_M\} \cdot \mathbb{E}\left[\frac{\sum_{i=1}^k Z_{(i)}}{\sqrt{\sum_{i=1}^M Z_i^2}}\middle| \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_M\right] \\ &\quad + \Pr\{\mathcal{E}_1^c \cap \mathcal{E}_2 \cap \mathcal{E}_M\} \cdot \mathbb{E}\left[\frac{\sum_{i=1}^k Z_{(i)}}{\sqrt{\sum_{i=1}^M Z_i^2}}\middle| \mathcal{E}_1^c \cap \mathcal{E}_2 \cap \mathcal{E}_M\right] \\ &\quad + \Pr\{\mathcal{E}_2^c \cup \mathcal{E}_M^c\} \cdot \mathbb{E}\left[\frac{\sum_{i=1}^k Z_{(i)}}{\sqrt{\sum_{i=1}^M Z_i^2}}\middle| \mathcal{E}_2^c \cup \mathcal{E}_M^c\right].\end{aligned}$$

We bound these three terms separately. To bound the first one, observe that condition on $\mathcal{E}_1 \cap \mathcal{E}_2$, $\sum_{i=1}^k Z_{(i)} \geq \tilde{Z}^{(j)}_{(1)} \geq 0.9k\sqrt{\frac{\log N}{\pi \log 2}} - \frac{k}{\sqrt{M}}\gamma$. As a result,

$$\begin{aligned}&\Pr\{\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_M\} \cdot \mathbb{E}\left[\frac{\sum_{i=1}^k Z_{(i)}}{\sqrt{\sum_{i=1}^M Z_i^2}}\middle| \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_M\right] \\ &\geq \frac{0.9k\sqrt{\frac{\log N}{\pi \log 2}} - \frac{k}{\sqrt{M}}\gamma}{\sqrt{M(1+\gamma)}} \cdot \left(1 - \left(0.1 + e^{-\xi^2/2} + 2e^{-M\gamma^2/4}\right)\right).\end{aligned} \tag{102}$$

To bound the second term, observe that under $\mathcal{E}_2$,

$$\sum_{i=1}^k Z_{(i)} \geq -\frac{k}{\sqrt{M}}\xi,$$

so we have

$$\begin{aligned}
&\Pr\left\{\mathcal{E}_2 \cap \mathcal{E}_1^c \cap \mathcal{E}_M\right\} \cdot \mathbb{E}\left[\frac{\sum_{i=1}^k Z_{(i)}}{\sqrt{\sum_{i=1}^M Z_i^2}} \middle| \mathcal{E}_2 \cap \mathcal{E}_1^c \cap \mathcal{E}_M\right] \\
&\geq \Pr\left\{\mathcal{E}_2 \cap \mathcal{E}_1^c \cap \mathcal{E}_M\right\} \cdot \left(-\frac{k}{\sqrt{M^2(1-\gamma)}}\xi\right) \\
&\geq \Pr\left\{\mathcal{E}_1^c\right\} \cdot \left(-\frac{k}{\sqrt{M^2(1-\gamma)}}\xi\right) \\
&\geq 0.1 \cdot \left(-\frac{\xi\sqrt{k}}{\sqrt{M^2(1-\gamma)}}\right). \qquad (103)
\end{aligned}$$

For the third term, it holds that

$$\begin{aligned}
&\Pr\left\{\mathcal{E}_2^c \cup \mathcal{E}_M^c\right\} \cdot \mathbb{E}\left[\frac{\sum_{i=1}^k Z_{(i)}}{\sqrt{\sum_{i=1}^M Z_i^2}} \middle| \mathcal{E}_2^c \cup \mathcal{E}_M^c\right] \\
&\geq \Pr\left\{\mathcal{E}_2^c \cup \mathcal{E}_M^c\right\} \cdot \inf_{a \in \mathbb{S}^{M-1}} \sum_{i=1}^k a_{(i)} \\
&\geq -\Pr\left\{\mathcal{E}_2^c \cup \mathcal{E}_M^c\right\} \cdot \frac{k}{\sqrt{M}} \\
&\geq -\left(e^{-\xi^2/2} + e^{-M\gamma^2/4}\right) \cdot \frac{k}{\sqrt{M}} \qquad (104)
\end{aligned}$$

Combining (102), (103), and (104) together, we arrive at

$$\begin{aligned}
&\mathbb{E}_{Z_1,\dots,Z_M \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)}\left[\frac{\sum_{i=1}^k Z_{(i)}}{\sqrt{\sum_{i=1}^M Z_i^2}}\right] \\
&\geq \frac{0.9k\left(\sqrt{\frac{\log N}{\pi \log 2}}\right) - \frac{k}{\sqrt{M}}\gamma}{\sqrt{M(1+\gamma)}} \cdot \left(1 - \left(0.1 + e^{-\xi^2/2} + 2e^{-M\gamma^2/4}\right)\right) - 0.1 \cdot \left(\frac{\xi\sqrt{k}}{\sqrt{M^2(1-\gamma)}}\right) \\
&\quad - \left(e^{-\xi^2/2} + e^{-M\gamma^2/4}\right) \cdot \frac{k}{\sqrt{M}}.
\end{aligned}$$

Finally, setting $\gamma = O\left(\frac{1}{\sqrt{M}}\right)$ and $\xi = O(1)$ yields the desired lower bound

$$C_{d,M,k} = \Omega\left(\frac{k\log N}{\sqrt{M}}\right).$$

□

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.

I Additional Details on Prior LDP Schemes

For completeness, we provide additional details on prior LDP mean estimation schemes in this section, including `PrivUnit` [4], `SQKR` [6], `FT21` [12], and `MMRC` [30]. We skip prior work analyzing compression-privacy-utility tradeoffs that do not specifically focus on the distributed mean estimation problem [19, 20] or others that study frequency estimation [6, 11, 30].

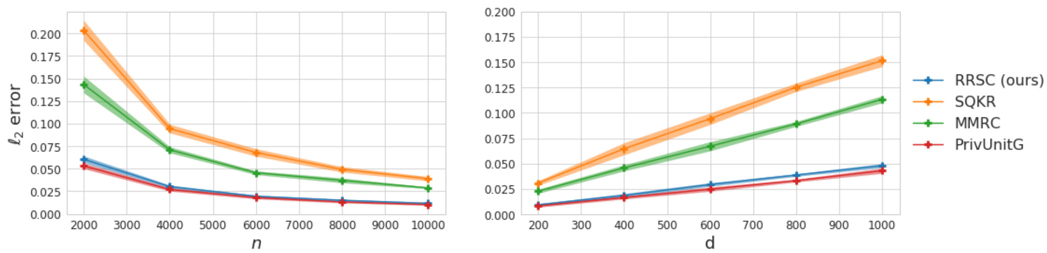

Figure 2: Comparison of `RRSC` with `SQKR` [6], `MMRC` [30], and `PrivUnitG` [2]. **(left)** ℓ_2 error vs number of users n with $d = 500$, $\varepsilon = 6$, and the number of bits is $b = \varepsilon = 6$. $k = 1$ for each n. **(right)** ℓ_2 error vs dimension d for $n = 5000$, $\varepsilon = 6$, and the number of bits is $b = \varepsilon = 6$. $k = 1$ for for each d.

I.1 `PrivUnit` [4]

[2] 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 [4, 2]). The normalization factor is required to obtain unbiasedness.

[2] 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` [6]

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 $\max_j |a_j| \leq \frac{K}{\sqrt{N}} \|v\|_2$ [23]. [27] 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[36], 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 [6], 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` [12] and `MMRC` [30]

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` [12], 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 [12]

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^ε ln(1/γ).
for j ∈ {1, ··· , J} do
    Sample a random seed s ∈ {0,1}^ℓ.
    Draw u ← p(·) using the PRG G and the random seed s.
    Sample b from Bernoulli( q(u|v) / (e^ε · p(u)) ).
    if b = 1 then
        BREAK
    end if
end for
```

Output: s

On the other hand, under `MRC` [30] the LDP mechanism is simulated via a minimal random coding technique [15]. 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 [30] 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 [30]

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^*

References

[1] N. Agarwal, A. T. Suresh, F. X. X. Yu, S. Kumar, and B. McMahan. cpsgd: Communication-efficient and differentially-private distributed sgd. *Advances in Neural Information Processing Systems*, 31, 2018.

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

[3] L. P. Barnes, H. A. Inan, B. Isik, and A. Özgür. rtop-k: A statistical estimation approach to distributed sgd. *IEEE Journal on Selected Areas in Information Theory*, 1(3):897–907, 2020.

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

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

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

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

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

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

[10] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pages 265–284. Springer, 2006.

[11] V. Feldman, J. Nelson, H. Nguyen, and K. Talwar. Private frequency estimation via projective geometry. In *International Conference on Machine Learning*, pages 6418–6433. PMLR, 2022.

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

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

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

[15] M. Havasi, R. Peharz, and J. M. Hernández-Lobato. Minimal random code learning: Getting bits back from compressed model parameters. In *International Conference on Learning Representations*, 2019.

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

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

[18] B. Isik, F. Pase, D. Gunduz, T. Weissman, and Z. Michele. Sparse random networks for communication-efficient federated learning. In *The Eleventh International Conference on Learning Representations*, 2023.

[19] B. Isik and T. Weissman. Learning under storage and privacy constraints. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 1844–1849. IEEE, 2022.

[20] B. Isik and T. Weissman. Lossy compression of noisy data for private and data-efficient learning. *IEEE Journal on Selected Areas in Information Theory*, 2023.

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

[22] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, et al. Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2):1–210, 2021.

[23] B. Kashin. Section of some finite-dimensional sets and classes of smooth functions (in russian) izv. *Acad. Nauk. SSSR*, 41:334–351, 1977.

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

[25] J. Konečnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*, 2016.

[26] Y. Lin, S. Han, H. Mao, Y. Wang, and B. Dally. Deep gradient compression: Reducing the communication bandwidth for distributed training. In *International Conference on Learning Representations*, 2018.

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

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

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

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

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

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

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

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

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

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

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

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

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