Optimal Algorithms for Mean Estimation under Local Differential Privacy

Hilal Asi [* 1]   Vitaly Feldman [* 2]   Kunal Talwar [* 2]

Abstract

We study the problem of mean estimation of ℓ_2-bounded vectors under the constraint of local differential privacy. While the literature has a variety of algorithms that achieve the asymptotically optimal rates for this problem, the performance of these algorithms in practice can vary significantly due to varying (and often large) hidden constants. In this work, we investigate the question of designing the protocol with the smallest variance. We show that PrivUnit (Bhowmick et al., 2018) with optimized parameters achieves the optimal variance among a large family of locally private randomizers. To prove this result, we establish some properties of local randomizers, and use symmetrization arguments that allow us to write the optimal randomizer as the optimizer of a certain linear program. These structural results, which should extend to other problems, then allow us to show that the optimal randomizer belongs to the PrivUnit family.

We also develop a new variant of PrivUnit based on the Gaussian distribution which is more amenable to mathematical analysis and enjoys the same optimality guarantees. This allows us to establish several useful properties on the exact constants of the optimal error as well as to numerically estimate these constants.

1. Introduction

Mean estimation is one of the most fundamental problems in machine learning and is the building block of a countless number of algorithms and applications including stochastic optimization (Duchi, 2018), federated learning (Bonawitz et al., 2017) and others. However, it is now evident that

standard algorithms for this task may leak sensitive information about users' data and compromise their privacy. This had led to the development of numerous algorithms for estimating the mean while preserving the privacy of users. The most common models for privacy are either the central model where there exists a trusted curator or the local model where such trusted curator does not exist.

In this work, we study the problem of mean estimation in the local privacy model. More specifically, we have n users each with a vector v_i in the Euclidean unit ball in $\mathbb{R}^d$. Each user will use a randomizer $\mathcal{R} : \mathbb{R}^d \to \mathcal{Z}$ to privatize their data where $\mathcal{R}$ must satisfy ε-differential privacy, namely, for any v_1 and v_2, $P(\mathcal{R}(v_1) = u)/P(\mathcal{R}(v_2) = u) \le e^\varepsilon$. Then, we run an aggregation method $\mathcal{A} : \mathcal{Z}^n \to \mathbb{R}^d$ such that $\mathcal{A}(\mathcal{R}(v_1), \dots, \mathcal{R}(v_n))$ provides an estimate of $\frac{1}{n} \sum_{i=1}^n v_i$. Our goal in this work is to characterize the optimal protocol (pair of randomizer $\mathcal{R}$ and aggregation method $\mathcal{A}$) for this problem and study the resulting optimal error.

Due to its importance and many applications, the problem of private mean estimation in the local model has been studied by numerous papers (Bhowmick et al., 2018; Feldman & Talwar, 2021; Chen et al., 2020). As a result, a clear understanding of the asymptotic optimal rates has emerged, showing that the optimal squared error is proportional to $\Theta(\frac{d}{n \min(\varepsilon, \varepsilon^2)})$: Duchi et al. (2018); Bhowmick et al. (2018) developed algorithms that obtain this rate and (Duchi & Rogers, 2019) proved corresponding lower bounds. Subsequent papers (Feldman & Talwar, 2021; Chen et al., 2020) have developed several other algorithms that achieve the same rates.

However, these optimality results do not give a clear characterization of which algorithm will enjoy better performance in practice. Constant factors here matter more than they do in run time or memory, as ε is typically limited by privacy constraints, and increasing the sample size by collecting data for more individuals is often infeasible or expensive. The question of finding the randomizer with the smallest error is therefore of great interest.

1.1. Our contributions

Motivated by these limitations, we investigate strict optimality for the problem of mean estimation with local privacy. We study the family of non-interactive and unbiased proto-

---

*Equal contribution [1]Stanford University, part of this work performed while interning at Apple [2]Apple, USA.. Correspondence to: Hilal Asi <asi@stanford.edu>, Vitaly Feldman <vitaly.edu@gmail.com>, Kunal Talwar <ktalwar@apple.com>.

cols, that is, a protocol is a pair of local private randomizer $\mathcal{R} : \mathbb{S}^{d-1} \to \mathcal{Z}$ and an aggregation method $\mathcal{A} : \mathcal{Z}^n \to \mathbb{R}^d$ where the protocol outputs $\mathcal{A}(\mathcal{R}(v_1), \dots, \mathcal{R}(v_n))$ such that $\mathbb{E}[\mathcal{A}(\mathcal{R}(v_1), \dots, \mathcal{R}(v_n))] = \frac{1}{n}\sum_{i=1}^n v_i$. We measure the error of a private protocol in terms of its (worst case) mean squared error

$$\mathsf{Err}_n(\mathcal{A}, \mathcal{R}) =$$

$$\sup_{v_1,\dots,v_n \in \mathbb{S}^{d-1}} \mathbb{E}\left[\left\|\mathcal{A}(\mathcal{R}(v_1), \dots, \mathcal{R}(v_n)) - \frac{1}{n}\sum_{i=1}^n v_i\right\|_2^2\right].$$

We obtain the following results.

First, we show that PrivUnit of Bhowmick et al. (2018) with optimized parameters is optimal amongst a large family of protocols. Our strategy for proving optimality consists of two main steps: first, we show that for non-interactive protocols, additive aggregation with a certain randomizer attains the optimal error. Then, for protocols with additive aggregation, we show that PrivUnit obtains the optimal error. Our proof builds on establishing several new properties of the optimal local randomizer, which allow us to express the problem of designing the optimal randomizer as a linear program. This in turn helps characterize the structure of optimal randomizers and allows us to show that there is an optimal randomizer which is an instance of PrivUnit.

Finding the exact constants in the error of PrivUnit is mathematically challenging. Our second contribution is to develop a new algorithm PrivUnitG that builds on the Gaussian distribution and attains the same error as PrivUnit up to a $(1 + o(1))$ multiplicative factor as $d \to \infty$. In contrast to PrivUnit, we show that the optimal parameters of PrivUnitG are independent of the dimension, hence enabling efficient calculation of the constants for high dimensional settings. Moreover, PrivUnitG is amenable to mathematical analysis which yields several properties on the constants of the optimal error.

1.2. Related work

Local privacy is perhaps one of the oldest forms of privacy and dates back to Warner (1965) who used it to encourage truthfulness in surveys. This definition resurfaced again in the contex of modern data analysis by Evfimievski et al. (2003) and was related to differential privacy in the seminal work of Dwork et al. (2006). Local privacy has attracted a lot of interest, both in the academic community (Beimel et al., 2008; Duchi et al., 2018; Bhowmick et al., 2018), and in industry where it has been deployed in several industrial applications (Erlingsson et al., 2014; Apple Differential Privacy Team, 2017). Recent work in the Shuffle model of privacy (Bittau et al., 2017; Cheu et al., 2019; Erlingsson et al., 2019; Balle et al., 2019; Feldman et al., 2022) has led to increased interest in the local model with moderate

values of the local privacy parameter, as they can translate to small values of central ε under shuffling.

The problem of locally private mean estimation has received a great deal of attention in the past decade (Duchi et al., 2018; Bhowmick et al., 2018; Duchi & Rogers, 2019; Erlingsson et al., 2020; Agarwal et al., 2018; Girgis et al., 2020; Chen et al., 2020; Gandikota et al., 2019; Feldman & Talwar, 2021). Duchi et al. (2018) developed asymptotically optimal procedures for estimating the mean when $\varepsilon \leq 1$, achieving expected squared error $O(\frac{d}{n\varepsilon^2})$. Bhowmick et al. (2018) proposed a new algorithm that is optimal for $\varepsilon \geq 1$ as well, achieving error $O(\frac{d}{n\min(\varepsilon,\varepsilon^2)})$. These rates are optimal as Duchi & Rogers (2019) show tight lower bounds which hold for interactive protocols. There has been more work on locally private mean estimation that studies the problem with additional constraints such as communications cost (Erlingsson et al., 2020; Feldman & Talwar, 2021; Chen et al., 2020).

Ye & Barg (2017; 2018) study (non-interactive) locally private estimation problems with discrete domains and design algorithms that achieve optimal rates. These optimality results are not restricted to the family of unbiased private mechanisms. However, in contrast to our work, these results are only asymptotic hence their upper and lower bounds are matching only as the number of samples goes to infinity.

While there are several results in differential privacy that establish asymptotically matching lower and upper bounds for various problems of interest, strict optimality results are few. While there are some results known for the one-dimensional problem (Ghosh et al., 2009; Gupte & Sundararajan, 2010), some of which extend to a large class of utility functions, such universal mechanisms are known not to exist for multidimensional problems (Brenner & Nissim, 2010). (Geng et al., 2015; Kairouz et al., 2016) show that for certain loss functions, one can phrase the problem of designing optimal local randomizers as linear programs, whose size is exponential in the size of the input domain.

2. Problem setting and preliminaries

We begin in this section by defining local differential privacy. To this end, we say that two probability distributions P and Q are (ε, δ)-close if for every event E

$$e^{-\varepsilon}(P(E) - \delta) \leq Q(E) \leq e^\varepsilon P(E) + \delta.$$

We say two random variables are (ε, δ)-close if their distributions are (ε, δ)-close.

We can now define local DP randomizers.

Definition 2.1. A randomized algorithm $\mathcal{R} : X \to Y$ is (replacement) (ε, δ)-DP local randomizer if for all $x, x' \in X$, $\mathcal{R}(x)$ and $\mathcal{R}(x')$ are (ε, δ)-close.

In this work, we will primarily be interested in *pure* DP randomizers, i.e. those which satisfy $(\varepsilon, 0)$-DP. We abbreviate this as ε-DP. In the setting of local randomizers, the difference between (ε, δ)-DP and pure DP is not significant; indeed any (ε, δ)-DP local randomizer can be converted (Feldman et al., 2022; Cheu & Ullman, 2021) to one that satisfies ε-DP while changing the distributions by a statistical distance of at most $O(\delta)$.

The main problem we study in this work is locally private mean estimation. Here, we have n unit vectors $v_1, \ldots, v_n \in \mathbb{R}^d$, i.e $v_i \in \mathbb{S}^{d-1}$. The goal is to design (locally) private protocols that estimate the mean $\frac{1}{n} \sum_{i=1}^{n} v_i$. We focus on the setting of non-interactive private protocols: such a protocol consists of a pair of private local randomizer $\mathcal{R} : \mathbb{S}^{d-1} \rightarrow \mathcal{Z}$ and aggregation method $\mathcal{A} : \mathcal{Z}^n \rightarrow \mathbb{R}^d$ where the final output is $\mathcal{A}(\mathcal{R}(v_1), \ldots, \mathcal{R}(v_n))$. We require that the output is unbiased, that is, $\mathbb{E}[\mathcal{A}(\mathcal{R}(v_1), \ldots, \mathcal{R}(v_n))] = \frac{1}{n} \sum_{i=1}^{n} v_i$, and wish to find private protocols that minimize the variance

$$\mathsf{Err}_n(\mathcal{A}, \mathcal{R}) =$$
$$\sup_{v_1, \ldots, v_n \in \mathbb{S}^{d-1}} \mathbb{E}\left[\left\| \mathcal{A}(\mathcal{R}(v_1), \ldots, \mathcal{R}(v_n)) - \frac{1}{n} \sum_{i=1}^{n} v_i \right\|_2^2 \right].$$

Note that in the above formulation, the randomizer $\mathcal{R}$ can have arbitrary domains (not necessarily $\mathbb{R}^d$), and the aggregation method can be arbitrary as well. However, one important special family of private protocols, which we term *canonical private protocols*, are protocols where the local randomizer $\mathcal{R} : \mathbb{S}_2^{d-1} \rightarrow \mathbb{R}^d$ has outputs in $\mathbb{R}^d$ and the aggregation method is the simple additive aggregation $\mathcal{A}^+(z_1, \ldots, z_n) = \frac{1}{n} \sum_{i=1}^{n} z_i$. In addition to being a natural family of protocols, canonical protocols are (i) simple and easy to implement, and (ii) achieve the smallest possible variance amongst the family of all possible unbiased private protocols, as we show in the subsequent sections.

Notation We let $\mathbb{S}^{d-1} = \{u \in \mathbb{R}^d : \|u\|_2 = 1\}$ denote the unit sphere, and $R \cdot \mathbb{S}^{d-1}$ denote the sphere of radius $R > 0$. Whenever clear from context, we use the shorter notation $\mathbb{S}$. Given a random variable V, we let f_V denote the probability density function of V. For a randomizer $\mathcal{R}$ and input v, $f_{\mathcal{R}(v)}$ denotes the probability density function of the random variable $\mathcal{R}(v)$. For a Gaussian random variable $V \sim \mathsf{N}(0, \sigma^2)$ with $\sigma > 0$, we let $\phi_\sigma : \mathbb{R} \rightarrow \mathbb{R}$ denote the probability density function of V and $\Phi_\sigma : \mathbb{R} \rightarrow [0, 1]$ denote its cumulative distribution function. For ease of notation, we write ϕ and Φ when $\sigma = 1$. Given two random variables V and U, we say that $V \stackrel{d}{=} U$ if V and U has the same distribution, that is, $f_V = f_U$. Finally, we let $e_i \in \mathbb{R}^d$ denote the standard basis vectors and $O(d) = \{U \in \mathbb{R}^{d \times d} : UU^T = I\}$ denote the subspace of orthonormal matrices of dimension d.

3. Optimality of PrivUnit

In this section, we prove our main optimality results showing that PrivUnit with additive aggregation achieves the optimal error among the family of unbiased locally private procedures. More precisely, we show that for any ε-DP local randomizer $\mathcal{R} : \mathbb{R}^d \rightarrow \mathcal{Z}$ and any aggregation method $\mathcal{A} : \mathcal{Z}^n \rightarrow \mathbb{R}^d$ that is unbiased,

$$\mathsf{Err}_n(\mathcal{A}^+, \mathrm{PrivUnit}) \leq \mathsf{Err}_n(\mathcal{A}, \mathcal{R}).$$

We begin in Section 3.1 by introducing the algorithm PrivUnit and stating its optimality guarantees in Section 3.2. To prove the optimality result, we begin in Section 3.3 by showing that there exists a canonical private protocol that achieves the optimal error, then in Section 3.2 we show that PrivUnit is the optimal local randomizer in the family of canonical protocols.

3.1. PrivUnit

We begin by introducing PrivUnit which was developed by Bhowmick et al. (2018). Given an input vector $v \in \mathbb{S}^{d-1}$ and letting $W \sim \mathsf{Uni}(\mathbb{S}^{d-1})$, PrivUnit$(p, \gamma)$ has the following distribution (up to normalization)

$$\mathrm{PrivUnit}(p, \gamma) \sim \begin{cases} W \mid \langle W, v \rangle \geq \gamma & \text{with prob. } p \\ W \mid \langle W, v \rangle < \gamma & \text{with prob. } 1 - p \end{cases}$$

A normalization factor is needed to obtain the correct expectation. We provide full details in Algorithm 1.

The following theorem states the privacy guarantees of PrivUnit. Theorem 1 in (Bhowmick et al., 2018) provides privacy guarantees based on several mathematical approximations which may not be tight. For our optimality results, we require the following exact privacy guarantee of PrivUnit.

Theorem 3.1. *Bhowmick et al., 2018, Theorem 1 Let $q = P(W_1 \leq \gamma)$ where $W \sim \mathsf{Uni}(\mathbb{S}^{d-1})$. If $\frac{p}{1-p} \frac{q}{1-q} \leq e^\varepsilon$ then PrivUnit(p, γ) is an ε-DP local randomizer.*

Throughout the paper, we will sometimes use the equivalent notation PrivUnit(p, q) which describes running PrivUnit(p, γ) with $q = P(W_1 \leq \gamma)$ as in Theorem 3.1.

3.2. Optimality

Asymptotic optimality of PrivUnit has already been established by prior work. Bhowmick et al. (2018) show that the error of PrivUnit is upper bounded by $O(\frac{d}{n \min(\varepsilon, \varepsilon^2)})$ for certain parameters. Moreover, Duchi & Rogers (2019) show a lower bound of $\Omega(\frac{d}{n \min(\varepsilon, \varepsilon^2)})$, implying that PrivUnit is asymptotically optimal.

In this section, we prove that additive aggregation applied with PrivUnit with the best choice of parameters p, γ is

truly optimal, that is, it outperforms any unbiased private algorithm. The following theorem states our optimality result for PrivUnit.

Theorem 3.2. *Let* $\mathcal{R} : \mathbb{S}^{d-1} \to \mathcal{Z}$ *be an* ε-*DP local randomizer, and* $\mathcal{A} : \mathcal{Z}^n \to \mathbb{R}^d$ *be an aggregation procedure such that* $\mathbb{E}[\mathcal{A}(\mathcal{R}(v_1), \ldots, \mathcal{R}(v_n))] = \frac{1}{n} \sum_{i=1}^{n} v_i$ *for all* $v_1, \ldots, v_n \in \mathbb{S}^{d-1}$. *Then there is* $p^\star \in [0,1]$ *and* $\gamma^\star \in [0,1]$ *such that PrivUnit$(p^\star, \gamma^\star)$ is* ε-*DP local randomizer and*

$$\mathsf{Err}(\mathcal{A}^+, PrivUnit(p^\star, \gamma^\star)) \leq \mathsf{Err}(\mathcal{A}, \mathcal{R}).$$

The proof of Theorem 3.2 will proceed in two steps: first, in Section 3.3 (Proposition 3.3), we show that there exists an optimal private procedure that is canonical, then in Section 3.2 (Proposition 3.4) we prove that PrivUnit is the optimal randomizer in this family. Theorem 3.2 is a direct corollary of these two propositions.

3.3. Optimality of canonical protocols

In this section, we show that there exists a canonical private protocol that achieves the optimal error. In particular, we have the following result. We defer the proof to Appendix A.2.

Proposition 3.3. *Let* $(\mathcal{R}, \mathcal{A})$ *be such that* $\mathcal{R} : \mathbb{S}^{d-1} \to \mathcal{Z}$ *is* ε-*DP local randomizer and* $\mathbb{E}[\mathcal{A}(\mathcal{R}(v_1), \ldots, \mathcal{R}(v_n))] = \frac{1}{n} \sum_{i=1}^{n} v_i$ *for all* $v_1, \ldots, v_n \in \mathbb{S}^{d-1}$. *Then there a canonical randomizer* $\mathcal{R}' : \mathbb{S}^{d-1} \to \mathbb{R}^d$ *that is* ε-*DP local randomizer and*
$$\mathsf{Err}_n(\mathcal{A}, \mathcal{R}) \geq \mathsf{Err}_n(\mathcal{A}^+, \mathcal{R}').$$

3.4. Optimality of PrivUnit among canonical randomizers

In this section, we show that PrivUnit achieves the optimal error in the family of canonical randomizers. To this end, first note that for additive aggregation $\mathcal{A}^+$, we have $\mathsf{Err}_n(\mathcal{A}^+, \mathcal{R}) = \mathsf{Err}_1(\mathcal{A}^+, \mathcal{R})/n$. Denoting $\mathsf{Err}(\mathcal{R}) = \mathsf{Err}_1(\mathcal{A}^+, \mathcal{R})$ for canonical randomizers, we have the following optimality result.

Proposition 3.4. *Let* $\mathcal{R} : \mathbb{S}^{d-1} \to \mathbb{R}^d$ *be an* ε-*DP local randomizer such that* $\mathbb{E}[\mathcal{R}(v)] = v$ *for all* $v \in \mathbb{S}^{d-1}$. *Then there is* $p^\star \in [0,1]$ *and* $\gamma^\star \in [0,1]$ *such that PrivUnit$(p^\star, \gamma^\star)$ is* ε-*DP local randomizer and*

$$\mathsf{Err}(PrivUnit(p^\star, \gamma^\star)) \leq \mathsf{Err}(\mathcal{R}).$$

The proof of Proposition 3.4 builds on a sequence of lemmas, each of which allows to simplify the structure of an optimal algorithm. We begin with the following lemma which show that there exists an optimal algorithm which is invariant to rotations.

Lemma 3.5 (Rotation-Invariance Lemma). *Let* $\mathcal{R}$: $\mathbb{S}^{d-1} \to \mathbb{R}^d$ *be an* ε-*DP local randomizer such that*

$\mathbb{E}[\mathcal{R}(v)] = v$ *for all* $v \in \mathbb{S}^{d-1}$. *There exists an* ε-*DP local randomizer* $\mathcal{R}'$ *such that*

1. $\mathbb{E}[\mathcal{R}'(v)] = v$ *for all* $v \in \mathbb{S}^{d-1}$

2. $\mathsf{Err}(\mathcal{R}') \leq \mathsf{Err}(\mathcal{R})$

3. $\mathsf{Err}(\mathcal{R}') = \mathsf{Err}(\mathcal{R}', v)$ *for all* $v \in \mathbb{S}^{d-1}$

4. *For any* $v, v_0 \in \mathbb{S}^{d-1}$, *there is an orthonormal matrix* $V \in \mathbb{R}^{d \times d}$ *such that* $\mathcal{R}'(v) \overset{d}{=} V\mathcal{R}'(v_0)$.

5. $\frac{f_{\mathcal{R}'(v)}(u_1)}{f_{\mathcal{R}'(v)}(u_2)} \leq e^\varepsilon$ *for any* $v \in \mathbb{S}^{d-1}$ *and* $u_1, u_2 \in \mathbb{R}^d$ *with* $\|u_1\|_2 = \|u_2\|_2$

Proof. Given $\mathcal{R}$, we define $\mathcal{R}'$ as follows. First, sample a random rotation matrix $U \in \mathbb{R}^{d \times d}$ where $U^T U = I$, then set

$$\mathcal{R}'(x) = U^T \mathcal{R}(Ux).$$

The randomizer $\mathcal{R}'$ satisfies all of our desired properties. We defer the full proof to Appendix A. $\square$

Lemma 3.5 implies that we can restrict our attention to algorithms have the same density for all inputs up to rotations and hence allows to study their behavior for a single input. Moreover, as we show in the following lemma, given a randomizer that works for a single input, we can extend it to achieve the same error for all inputs. To facilitate notation, we say that a density $f : \mathbb{R}^d \to \mathbb{R}_+$ is ε-indistinguishable if $\frac{f'(u_1)}{f'(u_2)} \leq e^\varepsilon$ for all $u_1, u_2 \in \mathbb{R}^d$ such that $\|u_1\|_2 = \|u_2\|_2$.

Lemma 3.6. *Fix* $v_0 = e_1 \in \mathbb{S}^{d-1}$. *Let* $f : \mathbb{R}^d \to \mathbb{R}_+$ *be an* ε-*indistinguishable density function with corresponding random variable* $\mathcal{R}$ *such that that* $\mathbb{E}[\mathcal{R}] = e_1$. *There exists an* ε-*DP local randomizer* $\mathcal{R}'$ *such that* $\mathsf{Err}(\mathcal{R}', v) = \mathbb{E}[\|\mathcal{R} - e_1\|_2^2]$ *and* $\mathbb{E}[\mathcal{R}'(v)] = v$ *for all* $v \in \mathbb{S}^{d-1}$.

Proof. The proof is similar to the proof of Lemma 3.5. For any $v \in \mathbb{S}^{d-1}$, we let $U(v) \in \mathbb{R}^{d \times d}$ be an orthonormal matrix such that $v_0 = U(v)v$. Then, following Lemma 3.5, we define $\mathcal{R}'(v) = U^T(v)\mathcal{R}$. The claim immediately follows. $\square$

Lemma 3.5 and Lemma 3.6 imply that we only need to study the behavior of randomizer for a fixed input. Henceforth, we will fix the input to $v = e_1$ and investigate properties of the density given e_1.

Given $v = (v_1, v_2, \ldots, v_d)$ we define its reflection to be $v^- = (v_1, -v_2, \ldots, -v_d)$. The next lemma shows that we can assume that the densities at v and v^- are equal for some optimal algorithm.

Lemma 3.7 (Reflection Symmetry). *Let $f : \mathbb{R}^d \to \mathbb{R}_+$ be an ε-indistinguishable density function with corresponding random variable $\mathcal{R}$ such that that $\mathbb{E}[\mathcal{R}] = e_1$. There is $f' : \mathbb{R}^d \to \mathbb{R}_+$ with corresponding random variable $\mathcal{R}'$ that satisfies the same properties such that $\mathsf{Err}(\mathcal{R}') \leq \mathsf{Err}(\mathcal{R})$ and $f'(u) = f'(u^-)$ for all $u \in \mathbb{R}^d$.*

Proof. We define $f'(u) = \frac{f(u)+f(u^-)}{2}$ for all $u \in \mathbb{S}^{d-1}$. First, it is immediate to see that $f(u) = f(u^-)$ for all $u \in \mathbb{R}^d$. Moreover, we have

$$\frac{f'(u_1)}{f'(u_2)} = \frac{f(u_1) + f(u_1^-)}{f(u_2) + f(u_2^-)}$$
$$\leq \max\left(\frac{f(u_1)}{f(u_2)}, \frac{f(u_1^-)}{f(u_2^-)}\right) \leq e^{\varepsilon}.$$

Note also that $\mathbb{E}[\mathcal{R}'] = e_1$ since the marginal distribution of the first coordinate in the output did not change and it is clear that for other coordinates the expectation is zero as $u + u^- = c \cdot e_1$ for any $u \in \mathbb{R}^d$. Finally, note that $\mathsf{Err}(\mathcal{R}', e_1) = \mathsf{Err}(\mathcal{R}, e_1)$ since $\|u - e_1\|_2 = \|u^- - e_1\|_2$ for all $u \in \mathbb{R}^d$. $\qquad\square$

We also have the following lemma which shows that the optimal density f outputs vectors on a sphere with some fixed radius.

Lemma 3.8. *Let $f : \mathbb{R}^d \to \mathbb{R}_+$ be an ε-indistinguishable density function with corresponding random variable $\mathcal{R}$ such that that $\mathbb{E}[\mathcal{R}] = e_1$. For any $\tau > 0$, there exists an ε-indistinguishable density $f' : \mathbb{R}^d \to \mathbb{R}_+$ with corresponding random variable $\mathcal{R}'$ such that $\|\mathcal{R}'\|_2 = C$ for some $C > 0$, $\mathbb{E}[\mathcal{R}'] = e_1$ and $\mathsf{Err}(\mathcal{R}') \leq \mathsf{Err}(\mathcal{R}) + \tau$.*

Proof. By Lemma 3.7, we can assume without loss of generality that $\mathcal{R}$ satisfies reflection symmetry, that is $f(u) = f(u^-)$. We think of the density f as first sampling a radius R then sampling a vector $u \in \mathbb{R}^d$ of radius R. We also assume that $\mathcal{R}$ has bounded radius; otherwise as $\mathsf{Err}(\mathcal{R}) = \mathbb{E}[\|\mathcal{R}^2\|_2] - 1$ is bounded, we can project the output of $\mathcal{R}$ to some large radius $R_\tau > 0$ while increasing the error by at most $\tau > 0$ for any τ. Similarly, we can assume that the output has radius at least r_τ while increasing the error by at most τ. Let f_R denote the distribution of the radius, and $f_{u|R=r}$ be the conditional distribution of the output given the radius is r. In this terminology $\mathbb{E}[\mathcal{R}] = e_1$ implies that

$$\mathsf{Err}(\mathcal{R}, e_1) = \mathbb{E}[\|\mathcal{R} - e_1\|_2^2]$$
$$= \mathbb{E}[\|\mathcal{R}\|_2^2 + \|e_1\|_2^2 - 2\langle\mathcal{R}, e_1\rangle]$$
$$= \mathbb{E}[\|\mathcal{R}\|_2^2] - 1.$$

For the purpose of finding the optimal algorithm, we need $\mathcal{R}$ that minimizes $\mathbb{E}[\|\mathcal{R}\|_2^2]$. Denote $W_r = \mathbb{E}[\langle\mathcal{R}, e_1\rangle \mid R = r]$

and set

$$C_{\max} = \sup_{r \in [r_\tau, R_\tau]} \frac{W_r}{r}.$$

Noting that $\mathbb{E}[\langle\mathcal{R}, e_1\rangle] = 1$, we have

$$\mathbb{E}[\|\mathcal{R}\|_2^2] = \frac{\mathbb{E}[\|\mathcal{R}\|_2^2]}{\mathbb{E}[\langle\mathcal{R}, e_1\rangle]^2}$$
$$= \frac{\mathbb{E}[\|\mathcal{R}\|_2^2]}{\left(\int_{r=0}^{\infty} f_r W_r \, dr\right)^2}$$
$$= \frac{\mathbb{E}[\|\mathcal{R}\|_2^2]}{\left(\int_{r=0}^{\infty} f_r r (W_r/r) \, dr\right)^2}$$
$$\geq \frac{\mathbb{E}[\|\mathcal{R}\|_2^2]}{\left(\int_{r=0}^{\infty} f_r r C_{\max} \, dr\right)^2}$$
$$\geq \frac{1}{C_{\max}^2} \frac{\mathbb{E}[\|\mathcal{R}\|_2^2]}{\mathbb{E}[\|\mathcal{R}\|_2]^2} \geq \frac{1}{C_{\max}^2}.$$

Now consider $r_{\max} > 0$ that has $C_{\max} = W_{r_{\max}}/r_{\max}$; $r_{\max}$ exists as $\mathcal{R}$ has outputs in $[r_\tau, R_\tau]$. Let $f_{\max}$ denote the conditional distribution of $\mathcal{R}$ given that $R = r_{\max}$ and let $\mathcal{R}_{\max}$ denote the corresponding randomizer. We define a new randomizer $\mathcal{R}'$ as follows

$$\mathcal{R}' = \frac{1}{r_{\max} C_{\max}} \mathcal{R}_{\max},$$

with corresponding density f'. Note that f' is ε-indistinguishable as f is ε-indistinguishable and the conditional distributions given different radii are disjoint which implies $f_{\max}$ is ε-indistinguishable. Moreover $f_{\max}(u) = f_{\max}(u^-)$ which implies that $\mathbb{E}[\mathcal{R}'] = \frac{1}{r_{\max} C_{\max}} \mathbb{E}[R \mid R = r_{\max}] = e_1$. Finally, note that $\mathcal{R}'$ satisfies

$$\mathbb{E}[\|\mathcal{R}'\|_2^2] = \frac{1}{r_{\max}^2 C_{\max}^2} \mathbb{E}[\|\mathcal{R}_{\max}\|_2^2]$$
$$= \frac{1}{r_{\max}^2 C_{\max}^2} \mathbb{E}[\|\mathcal{R}\|_2^2 \mid R = r_{\max}]$$
$$= \frac{1}{C_{\max}^2} \leq \mathbb{E}[\|\mathcal{R}\|_2^2].$$

The claim follows. $\qquad\square$

Before we present our main proposition which formulates the linear program that finds the optimal minimizer, we need the following key property which allows to describe the privacy guarantee as a linear constraint. We remark that such a lemma can easily be proven for deletion DP, so that our results would extend to that definition.

Lemma 3.9. *Let $\mathcal{R} : \mathbb{S}^{d-1} \to \mathbb{R}^d$ be an ε-DP local randomizer. There is $\rho : \mathbb{R}^d \to \mathbb{R}_+$ such that for all $v \in \mathbb{S}^{d-1}$ and $u \in \mathbb{R}$*

$$e^{-\varepsilon/2} \leq \frac{f_{\mathcal{R}(v)}(u)}{\rho(u)} \leq e^{\varepsilon/2}.$$

Moreover, if $\mathcal{R}$ satisfies the properties of Lemma 3.5 (invariance) then $\rho(u_1) = \rho(u_2)$ for $\|u_1\|_2 = \|u_2\|_2$.

Proof. Define

$$\rho(u) = \sqrt{\inf_{v \in \mathbb{S}^{d-1}} f_{\mathcal{R}(v)}(u) \cdot \sup_{v \in \mathbb{S}^{d-1}} f_{\mathcal{R}(v)}(u)}.$$

Note that for all $v \in \mathbb{S}^{d-1}$,

$$\frac{f_{\mathcal{R}(v)}(u)}{\sqrt{\inf_{v \in \mathbb{S}^{d-1}} f_{\mathcal{R}(v)}(u) \cdot \sup_{v \in \mathbb{S}^{d-1}} f_{\mathcal{R}(v)}(u)}}$$

$$= \sqrt{\frac{f_{\mathcal{R}(v)}(u)}{\inf_{v \in \mathbb{S}^{d-1}} f_{\mathcal{R}(v)}(u)}} \sqrt{\frac{f_{\mathcal{R}(v)}(u)}{\sup_{v \in \mathbb{S}^{d-1}} f_{\mathcal{R}(v)}(u)}} \leq e^{\varepsilon/2}.$$

The second direction follows similarly. The second part of the claim follows as for any $u_1, u_2 \in \mathbb{R}^d$ such that $\|u_1\|_2 = \|u_2\|_2$, if $f_{R(v_1)}(u_1) = t$ for any mechanism that satisfies the properties of Lemma 3.5 then there is v_2 such that $f_{R(v_2)}(u_2) = t$. The definition of ρ now implies that $\rho(u_1) = \rho(u_2)$. □

We are now ready to present our main step towards proving the optimality result. The following proposition formulates the problem of finding the optimal algorithm as a linear program. As a result, we show that there is an optimal algorithm whose density function has at most two different probabilities.

Proposition 3.10. *Let $\mathcal{R} : \mathbb{S}^{d-1} \to \mathbb{R}^d$ be an ε-DP local randomizer such that $\mathbb{E}[\mathcal{R}(v)] = v$ for all $v \in \mathbb{S}^{d-1}$. For any $\tau > 0$, there exist constants $C, p > 0$ and an ε-DP local randomizer $\mathcal{R}' : \mathbb{S}^{d-1} \to C \cdot \mathbb{S}^{d-1}$ such that $\mathbb{E}[\mathcal{R}'(v)] = v$ for all $v \in \mathbb{S}^{d-1}$, $\mathsf{Err}(\mathcal{R}') \leq \mathsf{Err}(\mathcal{R}) + \tau$, $f_{\mathcal{R}'(v)}(u) = f_{\mathcal{R}'(v)}(u^-)$, and $f_{\mathcal{R}'(v)}(u) \in \{e^{-\varepsilon/2}, e^{\varepsilon/2}\}p$ for all $u \in C \cdot \mathbb{S}^{d-1}$.*

Proof. The proof will proceed by formulating a linear program which describes the problem of finding the optimal randomizer and then argue that minimizer of this program must satisfy the desired conditions. To this end, first we use the properties of the optimal randomizer from the previous lemmas to simplify the linear program. Lemma 3.8 implies that there exists an optimal randomizer $\mathcal{R} : \mathbb{S}^{d-1} \to C \cdot \mathbb{S}^{d-1}$ for some $C > 0$ that is also invariant under rotations (satisfies the conclusions of Lemma 3.5). Moreover, Lemma 3.9 implies that the density function $f_{R(v)}$ has for some $p > 0$

$$e^{-\varepsilon/2}p \leq f_{R(v)}(u) \leq e^{\varepsilon/2}p.$$

Adding the requirement of unbiasedness, and noticing that for such algorithms the error is $C^2 - 1$, this results in the following minimization problem where the variables are

C and the density functions $f_v : C\mathbb{S}^{d-1} \to \mathbb{R}_+$ for all $v \in \mathbb{S}^{d-1}$

$$\arg \min_{C, f_v : C\mathbb{S}^{d-1} \to \mathbb{R}_+ \text{ for all } v \in \mathbb{S}^d} C \qquad (A)$$

subject to

$$e^{-\varepsilon/2}p \leq f_v(u) \leq e^{\varepsilon/2}p, \quad v \in \mathbb{S}^{d-1}, \ u \in C\mathbb{S}^{d-1}$$

$$\int_{C\mathbb{S}^{d-1}} f_v(u)u\,du = v, \quad v \in \mathbb{S}^{d-1}$$

$$\int_{C\mathbb{S}^{d-1}} f_v(u)du = 1, \quad v \in \mathbb{S}^{d-1}$$

Lemma 3.5 and Lemma 3.6 also show that the optimal algorithm is invariant under rotations, and that we only need to find the output distribution f with respect to a fixed input $v = e_1$. Moreover, Lemma 3.7 says that can assume that $f_{e_1}(u) = f_{e_1}(u^-)$ for all u. We also work now with the normalized algorithm $\hat{\mathcal{R}}(v) = \mathcal{R}(v)/C$ (that is, the output on the unit sphere). Note that for $\hat{\mathcal{R}}$ we have $\mathbb{E}[\hat{\mathcal{R}}(v)] = v/C$. Denoting $\alpha = 1/C$, this results in the following linear program (LP)

$$\arg \max_{\alpha, p, f_{e_1} : \mathbb{S}^{d-1} \to \mathbb{R}_+} \alpha \qquad (B)$$

subject to

$$e^{-\varepsilon/2}p \leq f_{e_1}(u) \leq e^{\varepsilon/2}p, \quad u \in \mathbb{S}^{d-1}$$

$$f_{e_1}(u) = f_{e_1}(u^-), \quad u \in \mathbb{S}^{d-1}$$

$$\int_{C\mathbb{S}^{d-1}} f_{e_1}(u)u\,du = \alpha e_1,$$

$$\int_{C\mathbb{S}^{d-1}} f_{e_1}(u)du = 1.$$

We need to show that most of the inequality constraints $e^{-\varepsilon/2}p \leq f_{e_1}(u) \leq e^{\varepsilon/2}p$ must be tight at one of the two extremes. To this end, we approximate the LP (B) using a finite number of variables by discretizing the density function f_{e_1}. We assume we have a $\delta/2$-cover $S = \{u_1, \dots, u_K\}$ of $\mathbb{S}^{d-1}$. We assume without loss of generality that if $u_i \in S$ then $u_i^- \in S$ and we also write $S = S_0 \cup S_1$ where $S_0 = S_1^-$ and $S_0 \cap S_1 = \emptyset$. Let $B_i = \{w \in \mathbb{S}^{d-1} : \|w - u_i\|_2 \leq \|w - u_j\|_2\}$, $V_i = \int_{u \in B_i} 1\,du$, and $\bar{u}_i = \mathbb{E}_{U \sim \mathsf{Uni}(\mathbb{S}^{d-1})}[U \mid U \in B_i]$. Now we limit our linear program to density functions that are constant over each B_i, resulting in the following LP

$$\arg \max_{\alpha, f_{e_1} : \mathbb{S}^{d-1} \to \mathbb{R}_+} \alpha \qquad (C)$$

subject to

$$e^{-\varepsilon/2}p \leq f_{e_1}(u) \leq e^{\varepsilon/2}p, \quad u \in S_0$$

$$\sum_{u \in S_0} f_{e_1}(u)V_i\bar{u}_i + \sum_{u \in S_1} f_{e_1}(u^-)V_i\bar{u}_i = \alpha e_1,$$

$$\sum_{u \in S_0} f_{e_1}(u)V_i + \sum_{u \in S_1} f_{e_1}(u^-)V_i = 1.$$

6

Let $\alpha_1^\star$ and $\alpha_2^\star$ denote the maximal values of (B) and (C), respectively. Each solution to (C) is also a solution to (B) hence $\alpha_1^\star \geq \alpha_2^\star$. Moreover, given $\delta > 0$, let f be a solution of (B) that obtains $\alpha \geq \alpha_1^\star - \delta$ and let $\mathcal{R}$ be the corresponding randomizer. We can now define a solution for the discrete program (C) by setting for $u \in B_i$,

$$\hat{f}(u) = \frac{1}{V_i} \int_{w \in B_i} f_{e_1}(w) dw$$

Equivalently, we can define $\hat{\mathcal{R}}$ as follows: first run $\mathcal{R}$ to get u and find B_i such that $u \in B_i$. Then return a vector uniformly at random from B_i. Note that $\hat{f}$ clearly satisfies the first and third constraints in (C). As for the second constraint, it follows since $\hat{f}_{e_1}(u) = \hat{f}_{e_1}(u^-)$ which implies that $\sum_{u \in S} f_{e_1}(u) V_i \bar{u}_i = \hat{\alpha} v$ for some $\hat{\alpha} > 0$. It remains to show that $\hat{\alpha} \geq \alpha_1^\star - 2\delta$. The above representation of $\hat{\mathcal{R}}$ shows that $\|\mathbb{E}[\hat{\mathcal{R}} - \mathcal{R}]\| \leq \delta$ and therefore we have $\hat{\alpha} \geq \alpha_1^\star - 2\delta$.

To finish the proof, it remains to show that the discrete LP (C) has a solution that satisfies the desired properties. Note that as this is a linear program with K variables and $2K + d + 2$ constraints, K linearly independent constraints must be satisfied Bertsimas & Tsitsiklis, 1997, theorem 2.3), which shows that for at least $K - d - 2$ of the sets B_i we have $f(u_i) \in \{e^\varepsilon, e^{-\varepsilon}\} p$.

To finish the proof, we need to manipulate the probabilities for the remaining $d - 2$ sets to satisfy our desired requirements. As these sets have small probability, this does not change the accuracy by much and we just need to do this manipulation carefully so as to preserve reflection symmetry and unbiasedness. The full details are tedious and we present them in Appendix A.4. □

Given the previous lemmas, we are now ready to finish the proof of Proposition 3.4.

Proof. Fix $\tau > 0$ and an unbiased ε-DP local randomizer $\mathcal{R}^\star$. Proposition 3.10 shows that there exists $\mathcal{R} : \mathbb{S}^{d-1} \to r_{\max} \mathbb{S}^{d-1}$ that is ε-DP, unbiased, reflection symmetric ($f_{\mathcal{R}(e_1)}(u) = f_{\mathcal{R}(e_1)}(u^-)$ for all u), and satisfies $f_{\mathcal{R}(e_1)}(u) \in \{e^{-\varepsilon/2}, e^{\varepsilon/2}\} p$. Morevoer $\mathsf{Err}(\mathcal{R}) \leq \mathsf{Err}(\mathcal{R}^\star) + \tau$. We will transform $\mathcal{R}$ into an instance of PrivUnit while maintaining the same error as $\mathcal{R}$.

To this end, if $\mathcal{R}$ is an instance of PrivUnit then we are done. Otherwise let $f = f_{\mathcal{R}(e_1)}$, $S_0(t) = \{u : f(u) = pe^{\varepsilon/2}, \langle u, e_1 \rangle \leq t\}$ and $S_1(t) = \{u : f(u) = pe^{-\varepsilon/2}, \langle u, e_1 \rangle \geq t\}$. Consider $t \in [-1, 1]$ that solves the following minimization problem

$$\underset{t \in [-1,1]}{\text{minimize}} \int_{S_0(t)} f(u) du + \int_{S_1(t)} f(u) du$$

Let $p^\star$ be the value of the above minimization problem and $t^\star$ the corresponding minimizer. Let $p_0 = \int_{S_0(t^\star)} f(u) du$ and $p_1 = \int_{S_1(t^\star)} f(u) du$. Assume without loss of generality that $p_0 \leq p_1$ (the other direction follows from identical arguments). Let $\hat{S} \subseteq S_1(t^\star)$ be such that $\hat{S} = \hat{S}^-$ and $\int_{\hat{S}} f(u) du = p_0$. We define $\tilde{f}$ by swapping the probabilities on $\hat{S}$ and $S_0(t^\star)$, that is,

$$\tilde{f}(u) = \begin{cases} f(u) & \text{if } u \notin S_0(t^\star) \cup \hat{S} \\ pe^{-\varepsilon/2} & \text{if } u \in S_0(t^\star) \\ pe^{\varepsilon/2} & \text{if } u \in \hat{S} \end{cases}$$

Clearly $\tilde{f}$ still satisfies all of our desired properties and has $\mathbb{E}_{U \sim \tilde{f}}[\langle U, e_1 \rangle] \geq \mathbb{E}_{U \sim f}[\langle U, e_1 \rangle]$ as we have that $\langle u_1, e_1 \rangle \geq \langle u_2, e_1 \rangle$ for $u_1 \in \hat{S}$ and $u_2 \in S_0(t^\star)$. Note also that $\tilde{f}(u) = pe^{-\varepsilon/2}$ for u such that $\langle u, e_1 \rangle \leq t$. Moreover, for u such that $\langle u, e_1 \rangle \geq t$, we have that $\tilde{f}(u) = pe^{-\varepsilon/2}$ only if $u \in B := S_1 \setminus \hat{S}$. Let δ be such that the set $A = \{u : t^\star \leq \langle u, e_1 \rangle \leq t^\star + \delta\}$ has $\int_{u \in A} \tilde{f}(u) du = \int_{u \in B} \tilde{f}(u) du$. We now define

$$\hat{f}(u) = \begin{cases} pe^{\varepsilon/2} & \text{if } \langle u, e_1 \rangle \geq t^\star + \delta \\ pe^{-\varepsilon/2} & \text{if } \langle u, e_1 \rangle \leq t^\star + \delta \end{cases}$$

Clearly, $\hat{f}(u)$ is an instance of PrivUnit. Now we prove that it satisfies all of our desired properties. First, note that we can write $\hat{f}$ as

$$\hat{f}(u) = \begin{cases} \tilde{f}(u) & \text{if } u \notin A \cup B \\ \tilde{f}(u) & \text{if } u \in A \cap B \\ pe^{-\varepsilon/2} & \text{if } u \in A \setminus B \\ pe^{\varepsilon/2} & \text{if } u \in B \setminus A \end{cases}$$

This implies that $\int_{\mathbb{S}^{d-1}} \hat{f} du = 1$. Moreover, $\hat{f}$ is ε-indistiguishable by definition. Finally, note that $\mathbb{E}_{U \sim \hat{f}}[\langle U, e_1 \rangle] \geq \mathbb{E}_{U \sim \tilde{f}}[\langle U, e_1 \rangle]$ as $\langle u_1, e_1 \rangle \geq \langle u_2, e_1 \rangle$ for $u_1 \in B \setminus A$ and $u_2 \in A \setminus B$. Let $\hat{\mathcal{R}}$ be the randomizer that corresponds to $\hat{f}$. We define $\mathcal{R}' = \frac{1}{\mathbb{E}_{U \sim \hat{f}}[\langle U, e_1 \rangle]} \hat{\mathcal{R}}$. We have that $\mathbb{E}[\mathcal{R}'] = e_1$ and that

$$\mathsf{Err}(\mathcal{R}') = \frac{1}{\mathbb{E}_{U \sim \hat{f}}[\langle U, e_1 \rangle]^2} - 1$$

$$\leq \frac{1}{\mathbb{E}_{U \sim f}[\langle U, e_1 \rangle]^2} - 1 = \mathsf{Err}(\mathcal{R}).$$

As $\mathcal{R}'$ is an instance of PrivUnit, the claim follows. □

4. PrivUnitG: an optimal algorithm based on Gaussian distribution

In this section, we develop a new variant of PrivUnit, namely PrivUnitG, based on the Gaussian distribution. PrivUnitG

essentially provides an easy-to-analyze approximation of the optimal algorithm PrivUnit. This enables to efficiently find accurate approximations of the optimal parameters $p^\star$ and $q^\star$. In fact, we show that these parameters are independent of the dimension which is computationally valuable. Moreover, building on PrivUnitG, we are able to analytically study the constants that characterize the optimal loss.

For a Gaussian random variable $U = \mathsf{N}(0, 1/d)$ and input vector v, PrivUnitG has the following distribution (up to normalization constants)

$$\text{PrivUnitG} \sim \begin{cases} U \mid \langle U, v \rangle \geq \gamma & \text{with probability } p \\ U \mid \langle U, v \rangle < \gamma & \text{with probability } 1 - p \end{cases}$$

We present the full details including the normalization constants in Algorithm 2. We usually use the notation PrivUnitG(p, q) which means applying PrivUnitG with p and $\gamma = \Phi^{-1}(q)/\sqrt{d}$.

The following proposition gives the privacy and utility guarantees for PrivUnitG. The r.v. α is defined (see Algorithm 2) as $\alpha = \langle U, v \rangle$ where U is drawn from PrivUnitG(v). We define $m = \sigma \phi(\gamma/\sigma) \left(\frac{p}{1-q} - \frac{1-p}{q} \right)$ with $\sigma^2 = 1/d$ and $\gamma = \sigma \cdot \Phi^{-1}(q)$. We defer the proof to Appendix B.2.

Proposition 4.1. *Let $p, q \in [0, 1]$ such that $\frac{p}{1-p} \frac{q}{1-q} \leq e^\varepsilon$. The algorithm PrivUnitG(p, q) is ε-DP local randomizer. Moreover, it is unbiased and has error*

$$\mathsf{Err}(PrivUnitG(p, q)) = \frac{\mathbb{E}[\alpha^2] + \frac{d-1}{d}}{\mathbb{E}[\alpha]^2} - 1.$$

Moreover, we have

$$m^2 \cdot \mathsf{Err}(PrivUnitG(p, q)) \overset{d \to \infty}{\to} 1.$$

Now we proceed to analyze the utility guarantees of PrivUnitG as compared to PrivUnit. To this end, we first define the error obtained by PrivUnitG with optimized parameters

$$\mathsf{Err}^\star_{\varepsilon,d}(\text{PrivUnitG}) = \inf_{p,q: \frac{pq}{(1-p)(1-q)} \leq e^\varepsilon} \mathsf{Err}(\text{PrivUnitG}(p, q)).$$

Similarly, we define this quantity for PrivUnit

$$\mathsf{Err}^\star_{\varepsilon,d}(\text{PrivUnit}) = \inf_{p,q: \frac{pq}{(1-p)(1-q)} \leq e^\varepsilon} \mathsf{Err}(\text{PrivUnit}(p, q)).$$

The following theorem shows that PrivUnitG enjoys the same error as PrivUnit up to small factors. We prove the theorem in Appendix B.4.

Theorem 4.2. *Let $p \in [0, 1]$ and $q \in [0, 1]$ such that PrivUnit(p, q) is ε-DP local randomizer. Then PrivUnitG(p, q) is also ε-DP local randomizer and has*

$$\frac{\mathsf{Err}(PrivUnitG(p, q))}{\mathsf{Err}(PrivUnit(p, q))} \leq 1 + O\left(\sqrt{\frac{\varepsilon + \log d}{d}} \right).$$

In particular,

$$\frac{\mathsf{Err}^\star_{\varepsilon,d}(PrivUnitG)}{\mathsf{Err}^\star_{\varepsilon,d}(PrivUnit)} \leq 1 + O\left(\sqrt{\frac{\varepsilon + \log d}{d}} \right).$$

We conduct several experiments that demonstrate that the error of both algorithms is nearly the same as we increase the dimension. We plot the ratio of the error of PrivUnitG and PrivUnit (for the same p and γ) for different epsilons and dimensions in Figure 1. These plots reaffirm the theoretical results of Theorem 4.2, that is, the ratio is smaller for large d and small ε.

4.1. Analytical expression for optimal error

We wish to understand the constants that characterize the optimal error. To this end, we build on the optimality of PrivUnitG and define the quantity $C_{\varepsilon,d}$ by

$$\mathsf{Err}^\star_{\varepsilon,d}(\text{PrivUnitG}) = C_{\varepsilon,d} \frac{d}{\varepsilon}.$$

We show that the sequence $C_{\varepsilon,d}$ has a limit as $d \to \infty$ and denote this limit by C_ε (that is, $C_{\varepsilon,d} \to C_\varepsilon$ as $d \to \infty$). Moreover, we also prove that the sequence C_ε has a limit as ε increases and let $C^\star$ be this limit ($C_\varepsilon \to C^\star$ as $\varepsilon \to \infty$).

We experimentally demonstrate the behavior of $C_{\varepsilon,d}$ and C_ε in Figure 2. These experiments show that $C^\star \approx 0.614$. We remark that as shown in (Feldman & Talwar, 2021), if $C_\varepsilon/C_{k\varepsilon}$ is close to 1, then one can get a near optimal algorithm for privacy parameter $k\varepsilon$ by repeating the algorithm for privacy parameter ε k times. The latter may be more efficient in terms of computation and this motivates understanding how quickly C_ε converges.

The following proposition shows that $C_{\varepsilon,d}$ converges as we increase the dimension d (proof in Appendix B.5).

Proposition 4.3. *Fix $\varepsilon > 0$. For any $1 \leq d_1 \leq d_2$,*

$$\Omega\left(\frac{\varepsilon + \log d_2}{d_2} + \frac{\varepsilon}{d_1} \right) \leq \left| \frac{C_{\varepsilon,d_1}}{C_{\varepsilon,d_2}} - 1 \right| \leq O\left(\frac{\varepsilon + \log d_1}{d_1} \right).$$

In particular, $C_{\varepsilon,d} \overset{d \to \infty}{\to} C_\varepsilon$.

The following proposition shows that C_ε also converges as we increase ε. We present the proof in Appendix B.6.

Proposition 4.4. *There is $C^\star > 0$ such that $\lim_{\varepsilon \to \infty} C_\varepsilon = C^\star$.*

5. Conclusions

We have shown that one can prove strict optimality for local randomizers under mild conditions, for the case of mean estimation of ℓ_2-bounded vectors. Our proof exploits the rotational symmetry of this input domain, and our approach
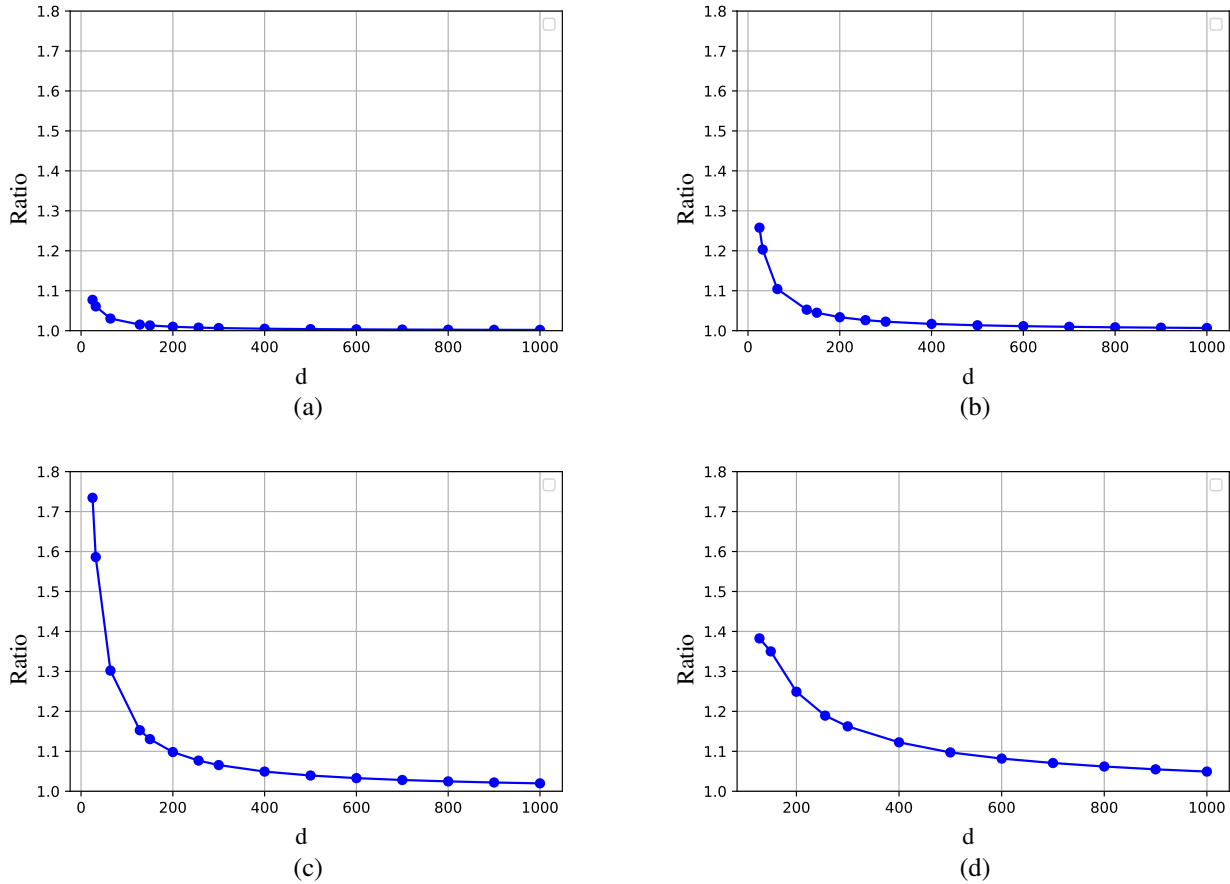
Figure 1. Ratio of the error of PrivUnitG to PrivUnit for (a) $\varepsilon = 4.0$ (b) $\varepsilon = 8.0$ (c) $\varepsilon = 16.0$ (d) $\varepsilon = 32.0$. We use the same p and γ for both algorithms by finding the best p, q that minimize PrivUnitG.
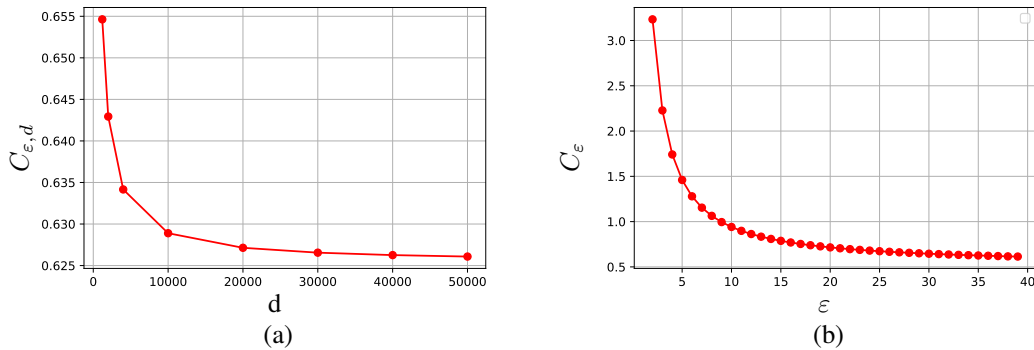

Figure 2. (a) $C_{\varepsilon,d}$ as a function of d for $\varepsilon = 35$. (b) C_{ε} as a function of ε (we obtain an accurate approximation of C_{ε} by taking a sufficiently large dimension $d = 5 \cdot 10^4$)

may be extendible to other problems that exhibit a high level of symmetry. We leave such extensions to future work.

We have also developed a new algorithm PrivUnitG, a simpler near-optimal mechanism for ℓ_2 mean estimation. It offers some advantages compared to PrivUnit and is more amenable to analysis. It is a promising candidate as a starting point to the compression approach of Feldman & Tal-war (2021) that may lead to computationally efficient low-communication randomizers that maintain near-optimality.

References

Agarwal, N., Suresh, A. T., Yu, F. X. X., Kumar, S., and McMahan, B. cpsgd: Communication-efficient

and differentially-private distributed sgd. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 31, pp. 7564–7575. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/file/21ce689121e39821d07d04faab328370-Paper.pdf.

Apple Differential Privacy Team. Learning with privacy at scale, 2017. Available at https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html.

Balle, B., Bell, J., Gascón, A., and Nissim, K. The privacy blanket of the shuffle model. In Boldyreva, A. and Micciancio, D. (eds.), *Advances in Cryptology – CRYPTO 2019*, pp. 638–667, Cham, 2019. Springer International Publishing. ISBN 978-3-030-26951-7.

Beimel, A., Nissim, K., and Omri, E. Distributed private data analysis: Simultaneously solving how and what. In *Advances in Cryptology*, volume 5157 of *Lecture Notes in Computer Science*, pp. 451–468. Springer, 2008.

Bertsimas, D. and Tsitsiklis, J. N. *Introduction to linear optimization*, volume 6. Athena Scientific Belmont, MA, 1997.

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

Bittau, A., Erlingsson, U., Maniatis, P., Mironov, I., Raghunathan, A., Lie, D., Rudominer, M., Kode, U., Tinnes, J., and Seefeld, B. Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles*, SOSP '17, pp. 441–459, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450350853. doi: 10.1145/3132747.3132769. URL https://doi.org/10.1145/3132747.3132769.

Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., and Seth, K. Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1175–1191, New York, NY, USA, 2017. ACM.

Brenner, H. and Nissim, K. Impossibility of differentially private universally optimal mechanisms. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pp. 71–80, 2010. doi: 10.1109/FOCS.2010.13.

Chen, W.-N., Kairouz, P., and Özgür, A. Breaking the communication-privacy-accuracy trilemma. *arXiv:2007.11707 [cs.LG]*, 2020.

Cheu, A. and Ullman, J. *The Limits of Pan Privacy and Shuffle Privacy for Learning and Estimation*, pp. 1081–1094. Association for Computing Machinery, New York, NY, USA, 2021. ISBN 9781450380539. URL https://doi.org/10.1145/3406325.3450995.

Cheu, A., Smith, A., Ullman, J., Zeber, D., and Zhilyaev, M. Distributed differential privacy via shuffling. In Ishai, Y. and Rijmen, V. (eds.), *Advances in Cryptology – EUROCRYPT 2019*, pp. 375–403, Cham, 2019. Springer International Publishing. ISBN 978-3-030-17653-2.

Diaconis, P. and Freedman, D. A dozen de finetti-style results in search of a theory. In *Annales de l'IHP Probabilités et statistiques*, volume 23, pp. 397–423, 1987.

Duchi, J. C. Introductory lectures on stochastic convex optimization. In *The Mathematics of Data*, IAS/Park City Mathematics Series. American Mathematical Society, 2018.

Duchi, J. C. and Rogers, R. Lower bounds for locally private estimation via communication complexity. In *Proceedings of the Thirty Second Annual Conference on Computational Learning Theory*, 2019.

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. Minimax optimal procedures for locally private estimation (with discussion). *Journal of the American Statistical Association*, 113(521):182–215, 2018.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *TCC*, pp. 265–284, 2006.

Erlingsson, U., Pihur, V., and Korolova, A. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 21st ACM Conference on Computer and Communications Security (CCS)*, 2014.

Erlingsson, U., Feldman, V., Mironov, I., Raghunathan, A., Talwar, K., and Thakurta, A. Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '19, pp. 2468–2479, USA, 2019. Society for Industrial and Applied Mathematics.

Erlingsson, Ú., Feldman, V., Mironov, I., Raghunathan, A., Song, S., Talwar, K., and Thakurta, A. Encode, shuffle, analyze privacy revisited: Formalizations and empirical evaluation. *arXiv:2001.03618 [cs.CR]*, 2020.

Evfimievski, A. V., Gehrke, J., and Srikant, R. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second Symposium on Principles of Database Systems*, pp. 211–222, 2003.

Feldman, V. and Talwar, K. Lossless compression of efficient private local randomizers. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 3208–3219. PMLR, 2021.

Feldman, V., McMillan, A., and Talwar, K. Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 954–964, 2022. doi: 10.1109/FOCS52979. 2021.00096. arXiv:2012.12803 [cs.LG].

Gandikota, V., Kane, D., Maity, R. K., and Mazumdar, A. vqsgd: Vector quantized stochastic gradient descent. *arXiv preprint arXiv:1911.07971*, 2019.

Geng, Q., Kairouz, P., Oh, S., and Viswanath, P. The staircase mechanism in differential privacy. *IEEE Journal of Selected Topics in Signal Processing*, 9(7):1176–1184, 2015. doi: 10.1109/JSTSP.2015.2425831.

Ghosh, A., Roughgarden, T., and Sundararajan, M. Universally utility-maximizing privacy mechanisms. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, STOC '09, pp. 351–360, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605585062. doi: 10. 1145/1536414.1536464. URL https://doi.org/ 10.1145/1536414.1536464.

Girgis, A. M., Data, D., Diggavi, S., Kairouz, P., and Suresh, A. T. Shuffled model of federated learning: Privacy, communication and accuracy trade-offs, 2020.

Gupte, M. and Sundararajan, M. Universally optimal privacy mechanisms for minimax agents. In *Proceedings of the Twenty-Ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '10, pp. 135–146, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781450300339. doi: 10.1145/1807085.1807105. URL https://doi. org/10.1145/1807085.1807105.

Kairouz, P., Oh, S., and Viswanath, P. Extremal mechanisms for local differential privacy. *J. Mach. Learn. Res.*, 17(1):492–542, January 2016. ISSN 1532-4435. URL http://dl.acm.org/citation. cfm?id=2946645.2946662.

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

Ye, M. and Barg, A. Asymptotically optimal private estimation under mean square loss. *arXiv:1708.00059 [math.ST]*, 2017.

Ye, M. and Barg, A. Optimal schemes for discrete distribution estimation under locally differential privacy. *IEEE Transactions on Information Theory*, 64(8):5662–5676, 2018.