Optimal Multidimensional Differentially Private Mechanisms in the Large-Composition Regime

Wael Alghamdi*[1], Shahab Asoodeh[2], Flavio P. Calmon[1], Juan Felipe Gomez[1], Oliver Kosut[3] and Lalitha Sankar[3]
[1] School of Engineering and Applied Sciences, Harvard University
(emails: alghamdi@g.harvard.edu, flavio@seas.harvard.edu, juangomez@g.harvard.edu)
[2] Department of Computing and Software, McMaster University (email: asoodehs@mcmaster.ca)
[3] School of Electrical, Computer, and Energy Engineering, Arizona State University (emails: {okosut, lsankar}@asu.edu)

2023 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-7554-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISIT54713.2023.10206658

***Abstract*—We construct vector differentially-private (DP) mechanisms that are asymptotically optimal in the limit of the number of compositions growing without bound. First, we derive via the central limit theorem a reduction from DP to a KL-divergence minimization problem. Second, we formulate the general theory of spherically-symmetric DP mechanisms in the large-composition regime. Specifically, we show that additive, continuous, spherically-symmetric DP mechanisms are optimal if one considers a spherically-symmetric cost (e.g., bounded noise variance) and an ℓ^2 sensitivity metric. We then formulate a finite-dimensional problem that produces noise distributions that can get arbitrarily close to optimal among monotone mechanisms. Finally, we demonstrate numerically that our proposed mechanism achieves better DP parameters than the vector Gaussian mechanism for the same variance constraint.**

The full proofs can be found in the extended version at [1].

I. Introduction

Differential privacy (DP) has emerged as the predominant measure for data privacy. DP is achieved by applying a randomized mechanism to functional queries from a dataset, in such a way so that any given element of the underlying dataset cannot be confirmed with certainty [2]. Any random mechanism will therefore introduce distortion on the query output, reducing utility. Thus, it is natural to ask how to design mechanisms that achieve the optimal trade-off between privacy and utility. A number of works [3]–[10] have sought optimal DP mechanisms in a variety of settings. However, these works all focus on the *single-shot* setting, in which a single mechanism is applied to a single query.

In contrast, many applications requiring data privacy involve processing data over a large number of steps. For example, when training a machine learning (ML) model with DP, one typically uses stochastic gradient descent with hundreds or thousands of iterations, where the DP mechanism is applied to each iteration [11]. In the DP literature, each iteration involving a mechanism is a *composition*, so an application involving many iterations is in the *large-composition regime*. Thus, the law of large numbers applies to the statistical analysis, which means that summary statistics can be used to accurately approximate the achieved DP. This view was taken in [12], which studied optimal scalar-valued DP mechanisms in the large-composition regime. It was shown that if the mechanism is designed by minimizing a Kullback-Leibler (KL) divergence subject to an expected cost constraint, then this mechanism will be optimal in the sense that any other mechanism will achieve worse DP for sufficiently many compositions.

However, many applications involve *vector* queries, not scalar queries. In the ML training example, the query is typically the gradient of the underlying model, which is a vector with length equal to the number of model parameters. In this paper, we study optimal mechanisms in the large-composition regime for vector queries. Our contributions are:

- As in [12], we find that the problem can be posed as a KL-divergence minimization problem, wherein the mechanism that minimizes this problem will outperform any other, given sufficiently many compositions.
- We show that a vector query is in general different from multiple scalar queries, and that the optimal mechanisms do not typically involve adding independent noise to each element of the vector.
- We show that for a spherically-symmetric cost function (e.g., ℓ^2 constraint) and an ℓ^2 sensitivity, the optimal mechanism is additive, continuous, and spherically-symmetric.
- We show that for any monotone spherically-symmetric noise mechanism, the worst-case shift is the maximum shift, where by "worst-case" we mean that it gives the maximum value for any f-divergence [13]. This means that minimizing the worst-case DP requires only focusing on the maximum shift. This result is crucial to accounting for multidimensional mechanisms, as it reduces the expectation to a 2-dimensional integral. This result also motivates our focus on *monotone* mechanisms, as they are dramatically simpler to analyze, despite not necessarily being optimal.
- While the optimal monotone mechanism has no closed-form expression, we formulate a finite-dimensional convex optimization problem (using the 2-dimensional integral form) that can be solved efficiently to find mechanisms arbitrarily close to optimal among monotone mechanisms.

*Corresponding author, and the remaining authors are in alphabetical order.

This material is based upon work supported by the National Science Foundation under Grant Nos. CAREER-1845852, FAI-2040880, CIF-1900750, SCH-2205080, CIF-1922971, CIF-1901243, and CIF-2007688; by NSERC Canada; and by the U.S. Department of Energy Award No. DE-SC0022158.

- We numerically demonstrate the performance of the mechanism found via the finite-dimensional problem, where we show that it performs better than the vector Gaussian mechanism, which is typically used in state-of-the-art applications.

The paper is organized as follows. Sec. II gives the definitions for DP, the large-composition regime, and poses the KL-divergence optimization problem. Sec. III proves the optimality of additive, continuous, and spherically symmetric mechanisms under certain conditions. Sec. IV describes how to compute privacy statistics for these symmetric mechanisms. Sec. V gives the finite-dimensional optimization problem, and proves that these mechanisms come arbitrarily close to optimal. Sec. VI illustrates some experimental results.

A. Notation and Assumptions

We fix a Euclidean space $\mathbb{R}^m$ throughout, and an m-dimensional random vector X. Denote by λ and $\|\cdot\|$ the Lebesgue measure and ℓ^2 norm, respectively, on $\mathbb{R}^m$. For a probability measure P on $\mathbb{R}^m$ and $c:\mathbb{R}^m\to\mathbb{R}$, the expectation is denoted by $\mathbb{E}_P[c]:=\int_{\mathbb{R}^m} c(x)\,dP(x)$. The shift operator is denoted by $(T_x r)(A):=r(A-x)$. For probability measures P,Q over $\mathbb{R}^m$, the KL-divergence is denoted by $D(P\,\|\,Q)$, the variance of the information density by

$$\mathrm{V}(P\,\|\,Q):=\mathbb{E}_P\left[\left(\log\frac{dP}{dQ}-D(P\,\|\,Q)\right)^2\right], \tag{1}$$

and the E_γ-divergence is defined for $\gamma\ge 0$ as

$$\mathsf{E}_\gamma(P\,\|\,Q):=\sup_{A\text{ Borel}} P(A)-\gamma Q(A)=\mathbb{E}_Q\left[\left(\frac{dP}{dQ}-\gamma\right)^+\right],$$

where $a^+:=\max(0,a)$. A probability measure P over $\mathbb{R}^m$ is spherically-symmetric if $P(\{Ux:x\in B\})=P(B)$ for any Borel $B\subset\mathbb{R}^m$ and every orthogonal matrix $U\in\mathbb{R}^{m\times m}$.

In the DP problem we consider, we restrict the sensitivity to be ℓ^2 sensitivity. The cost functions c we consider will satisfy mild assumptions that will be explicitly invoked when used.

II. Differential Privacy and The Large-Composition Regime

Let d be a dataset containing private information of several individuals and $f(d)\in\mathbb{R}^m$ be the response to a query f about this dataset (e.g., $f(d)$ could be the proportion of individuals in d having a particular property). To maintain the privacy of individuals, a typical approach is to pass $x=f(d)$ through a privacy-preserving mechanism $P_{Y|X}$ and release $Y\sim P_{Y|X=x}$. The de-facto standard definition for privacy is *differential privacy* (DP) [2]: $P_{Y|X}$ is said to be (ε,δ)-DP for $\varepsilon\ge 0$ and $\delta\in[0,1]$ if

$$\sup_{\substack{x,x'\in\mathbb{R}^m\\ \|x-x'\|\le s}} \mathsf{E}_{e^\varepsilon}(P_{Y|X=x}\parallel P_{Y|X=x'})\le\delta, \tag{2}$$

where $\|\cdot\|$ is a norm and s is the sensitivity of the query f defined as the maximum of $\|f(d)-f(d')\|$ over all pairs of datasets d and d' that differ in one entry.

In some applications (e.g., training deep models), a dataset d might receive k sequential queries $f_1,\dots,f_k$. As before, the privacy of individuals in d can be maintained by using a mechanism $P_{Y|X}$ for k times to generate the k-tuple $Y^k=(Y_1,\dots,Y_k)$ from the k-tuple $(f_1(d),\cdots,f_k(d))$. Let $P_{Y|X}^{\circ k}$ denote the resulting k-fold mechanism obtained by $P_{Y|X}$. In information-theoretic parlance, $P_{Y|X}^{\circ k}$ is a memoryless channel with $(f_1(d),\cdots,f_k(d))$ as the input and Y^k as the output.

It can be verified that $P_{Y|X}^{\circ k}$ is $(\varepsilon,\delta_{P_{Y|X}^{\circ k}}(\varepsilon))$-DP for any $\varepsilon\ge 0$, where

$$\delta_{P_{Y|X}^{\circ k}}(\varepsilon):=\sup_{\substack{\|u_i-v_i\|\le s\\ 1\le i\le k}} \mathsf{E}_{e^\varepsilon}\left(\prod_{i=1}^k P_{Y|X=u_i}\middle\|\prod_{i=1}^k P_{Y|X=v_i}\right), \tag{3}$$

or equivalently, $(\varepsilon_{P_{Y|X}^{\circ k}}(\delta),\delta)$-DP for any $\delta\in(0,1)$, where

$$\varepsilon_{P_{Y|X}^{\circ k}}(\delta):=\inf\left\{\varepsilon\ge 0\ :\ \delta_{P_{Y|X}^{\circ k}}(\varepsilon)\le\delta\right\}. \tag{4}$$

Computing the quantities $\delta_{P_{Y|X}^{\circ k}}(\varepsilon)$ or $\varepsilon_{P_{Y|X}^{\circ k}}(\delta)$ in closed-form expressions is intractable. Nevertheless, when k is sufficiently large (as in almost all deep learning applications), we can derive an asymptotic formula for $\varepsilon_{P_{Y|X}^{\circ k}}(\delta)$ in terms of the KL-divergence, as shown in the following theorem.

Theorem 1. *Fix a sensitivity $s>0$ and a Markov kernel $P_{Y|X}$ on $\mathbb{R}^m$ satisfying $\sup_{\|x-x'\|\le s}\mathrm{V}(P_{Y|X=x}\,\|\,P_{Y|X=x'})<\infty$. Then, for any $\delta\in(0,1/2)$, we have*

$$\lim_{k\to\infty}\frac{\varepsilon_{P_{Y|X}^{\circ k}}(\delta)}{k}=\sup_{\|x-x'\|\le s} D(P_{Y|X=x}\,\|\,P_{Y|X=x'}). \tag{5}$$

According to this theorem, the characterization of the privacy guarantee of $P_{Y|X}^{\circ k}$ in the large-composition regime reduces to computing the maximum KL-divergence $D(P_{Y|X=s}\|P_{Y|X=x'})$ over all possible choices of x and x' such that $\|x-x'\|\le s$. Given this asymptotic result, our goal is to design the "optimal" mechanism $P_{Y|X}$, that is, the mechanism with the best privacy guarantee (i.e., smallest $\varepsilon_{P_{Y|X}^{\circ k}}(\delta)$) while maintaining a desired level of "utility". To formalize the utility requirement, we consider the bound $\mathbb{E}[c(Y-x)\mid X=x]\le C$ for all $x\in\mathbb{R}^m$ and a given $C\ge 0$, where $c:\mathbb{R}^m\to\mathbb{R}_+$ is a measurable cost function. This constraint ensures that mechanism's output Y_i is reasonably close to its input $f_j(d)$.

Given the asymptotic result in Theorem 1 and the utility constraint, we can now formulate the asymptotically optimal DP mechanism with a given utility constraint as the solution of the following optimization problem:

$$\begin{aligned}\inf_{P_{Y|X}\in\mathscr{R}}\quad & \sup_{\|x-x'\|\le s} D(P_{Y|X=x}\parallel P_{Y|X=x'})\\ \text{subject to}\quad & \sup_{x\in\mathbb{R}^m}\mathbb{E}[c(Y-x)\mid X=x]\le C,\end{aligned} \tag{6}$$

where $\mathscr{R}$ denotes the set of all Markov kernels on $\mathbb{R}^m$.

III. Optimality of Additive, Continuous, Spherically Symmetric Mechanisms

We prove in this section that there is an additive mechanism $P_{Y|X=x} = T_x P$, $x \in \mathbb{R}^m$, (i.e., $P_{Y|X=x}(B) = P(B-x)$ for a measure P on $\mathbb{R}^m$ and every Borel set $B \subset \mathbb{R}^m$) that solves our main problem (6), and for which P is spherically symmetric and absolutely continuous with respect to the Lebesgue measure. We prove this result under the following assumption on the cost function c.

Assumption 1. *The cost function $c : \mathbb{R}^m \to \mathbb{R}$ satisfies:*

- Spherical symmetry: *there is a function $\widetilde{c} : \mathbb{R}_+ \to \mathbb{R}$ such that $c(x) = \widetilde{c}(\|x\|)$ for all $x \in \mathbb{R}^m$.*
- Positivity: *$c(x) \geq 0$ for all $x \in \mathbb{R}^m$, and $c(0) = 0$.*
- Monotonicity: *$c(x) \leq c(x')$ if $\|x\| \leq \|x'\|$.*
- Unboundedness: *$c(x) \to \infty$ as $\|x\| \to \infty$.*
- Lower-semicontinuity: *c is lower semicontinuous.*

A natural choice of cost function is positive powers of the the quadratic cost $c(x) = \|x\|^\alpha$, $\alpha > 0$, but we allow $c(x)$ to be any function that satisfies the above assumptions.

Let the subset $\mathscr{P} \subset \mathscr{R}$ consist of all Markov kernels $P_{Y|X}$ on $\mathbb{R}^m$ that satisfy the cost constraint in (6), i.e., for which

$$\sup_{x \in \mathbb{R}^m} \mathbb{E}[c(Y-x) \mid X = x] \leq C. \tag{7}$$

Then, we denote the result of the optimization (6) by

$$\mathrm{KL}^\star := \inf_{P_{Y|X} \in \mathscr{P}} \sup_{x,x' \in \mathbb{R}^m : \|x - x'\| \leq s} D(P_{Y|X=x} \| P_{Y|X=x'}). \tag{8}$$

We consider the subset of $\mathscr{P}$ consisting of those mechanisms $P_{Y|X}$ that are additive, i.e., $Y = X + Z$ for a noise Z independent of X. For such an additive mechanism $P_{Y|X=x} = T_x P$ (so $Z \sim P$), the KL-divergence can be rewritten as

$$D(P_{Y|X=x} \parallel P_{Y|X=x'}) = D(P \parallel T_{x'-x} P). \tag{9}$$

Therefore, we consider the set $\mathscr{P}_{\text{add}}$ consisting of Borel probability measures P on $\mathbb{R}^m$ that satisfy the cost constraint

$$\mathbb{E}_P[c] \leq C, \tag{10}$$

and denote the optimal value they can achieve in (6) by

$$\mathrm{KL}^\star_{\text{add}} := \inf_{P \in \mathscr{P}_{\text{add}}} \sup_{a \in \mathbb{R}^m : \|a\| \leq s} D(P \parallel T_a P). \tag{11}$$

As additive mechanisms are a subset of all mechanisms, we trivially have $\mathrm{KL}^\star \leq \mathrm{KL}^\star_{\text{add}}$. Conversely, we show in Theorem 2 below that when equality holds, the optimal value $\mathrm{KL}^\star$ is achievable by an additive mechanism that has a PDF.

Theorem 2. *If the cost function c satisfies Assumption 1, then additive mechanisms are optimal for the KL-divergence optimization* (6)*, i.e.,*

$$\mathrm{KL}^\star = \mathrm{KL}^\star_{\text{add}}. \tag{12}$$

Further, there exists at least one probability measure $P^\star \in \mathscr{P}_{\text{add}}$ achieving the optimal value $\mathrm{KL}^\star$, which is necessarily absolutely continuous with respect to the Lebesgue measure.

Remark 1. This theorem is proved using the same proof for its 1-dimensional instantiation in [12, Theorem 1], *mutatis mutandis*.

In addition to optimality of additive continuous mechanisms for (6) shown in Theorem 2, we show in Theorem 3 below that it also suffices to consider spherically-symmetric mechanisms. In other words, we may assume that the noise Z has a PDF p_Z that is spherically symmetric.

Theorem 3. *We may choose the optimal mechanism $P^\star$ in Theorem 2 to be spherically symmetric, i.e., with p denoting its PDF, we have that $p(z) = \widetilde{p}(\|z\|)$ for some function $\widetilde{p}$.*

Remark 2. The 1-dimensional version of this result follows from joint convexity of the KL-divergence, since the performance of a given mechanism p can only be improved by the even mechanism $(p(x) + p(-x))/2$. In the multi-dimensional setting, a more delicate argument using the Haar measure is necessary.

IV. Isotropic and Monotone DP Mechanisms

We lay the groundwork for a special class of isotropic DP mechanisms along with a provable method for computing their associated privacy parameters (ε, δ). Consider a DP mechanism that, given X, releases $Y = X + Z$ for a noise Z that is independent of X. We have shown in Section III that it suffices to consider continuous and spherically symmetric Z. It is not hard to see that any spherically symmetric random vector Z can be written in the form

$$Z = R \cdot U \tag{13}$$

where U is a uniformly distributed random vector over the unit $(m-1)$-sphere in $\mathbb{R}^m$, and R is a nonnegative scalar random variable (not necessarily independent of U). In fact, we may set $R = \|Z\|$ and $U = Z/\|Z\|$.

In the remainder of the paper, we will only consider "monotone" mechanisms, defined as follows.

Definition 1. We say that a random vector Z is *monotone* if it has a PDF $p(z)$ such that for every $z \in \mathbb{R}^m$ and $t \in [0, 1)$, we have $p(tz) \geq p(z)$.

Remark 3. Note that a random vector is monotone *and* spherically-symmetric if its PDF can be written $p(z) = \widetilde{p}(\|z\|)$ such that $\widetilde{p} : \mathbb{R}_+ \to \mathbb{R}_+$ is non-increasing. One example is the Gaussian mechanism $Z \sim \mathcal{N}(0, \sigma^2 I_m)$. We focus on monotone mechanisms since, for such mechanisms, it is tractable to both do DP accounting (see Lemma 1) as well as solve the KL-divergence optimization (6) (see Proposition 1).

The following lemma shows that accounting for monotone spherically-symmetric DP mechanisms reduces to computing the E_γ divergence at the maximal shift. This property is known to hold for the Gaussian mechanism [11].

Lemma 1. *If $Z \sim p$ is a monotone spherically-symmetric random vector, and $\gamma \geq 1$, then $a \mapsto \mathsf{E}_\gamma(p \| T_a p)$ is spherically*

symmetric and increasing in the norm of $\|a\|$. In particular, for any $s > 0$ we have

$$\max_{\|a\| \le s} \mathsf{E}_\gamma(p \,\|\, T_a p) = \mathsf{E}_\gamma(p \,\|\, T_{s\boldsymbol{e}_1} p). \tag{14}$$

We generalize Lemma 1 in another dimension, namely, we show next that the same result holds for any f-divergence. Specializing this result to the KL-divergence will help simplify the numerical implementation of the DP mechanism we introduce in the next section.

Proposition 1. *Let $f : (0,\infty) \to \mathbb{R}$ be a convex function satisfying $f(1) = 0$. For any monotone spherically-symmetric random vector $Z \sim p$, the mapping $a \mapsto D_f(p\|T_a p)$ is spherically symmetric and increasing in the norm of $\|a\|$. In particular, for any $s > 0$ we have*

$$\max_{\|a\| \le s} D_f(p \,\|\, T_a p) = D_f(p \,\|\, T_{s\boldsymbol{e}_1} p). \tag{15}$$

Remark 4. The above results show that monotonicity facilitates DP accounting—indeed, accounting for multidimensional non-monotonic mechanisms presents a significant challenge, since, as suggested by (3), one must maximize over all possible $u_i - v_i$ for $1 \le i \le k$. Thus, in the next section we restrict attention to the subclass of monotone mechanisms. Recall that Theorem 3 shows that, among all additive mechanisms, the optimal one is spherically-symmetric. It can be seen that spherically-symmetric mechanisms would still be optimal among all monotone mechanisms for the KL-divergence problem (6). Indeed, as in the proof of Theorem 3, if p is the PDF of an optimal mechanism that is monotone but not necessarily spherically-symmetric, then constructing the PDF

$$q(z) := \int_{\mathrm{O}(m)} p(Uz)\, d\mu(z) \tag{16}$$

(where μ the Haar measure over the orthogonal group $\mathrm{O}(m)$) we see that q is the PDF of a monotone spherically-symmetric mechanism that performs at least as well as p for the problem (6) (hence, optimally among monotone mechanisms). In the sequel, we will denote the optimal value achievable by a monotone mechanism for the problem (6) by $\mathrm{KL}^\star_{\text{monotone}}$.

V. Proposed Optimal Mechanisms

We construct in this section a family of monotone and spherically-symmetric mechanisms that are optimal among all monotone mechanisms for the KL-divergence optimization in (11), i.e., they achieve $\mathrm{KL}^\star_{\text{monotone}}$ (see Remark 4). In view of Lemma 1, DP accounting is possible for the family of mechanisms we construct. In addition, we use Proposition 1 to show that our proposed construction is numerically tractable. We also demonstrate in the next section via numerical experiments that our proposed mechanism achieves improved DP parameters in comparison to the Gaussian mechanism that has the same variance (i.e., with $c(x) = \|x\|^2$).[1]

[1]Without loss of generality, we fix the sensitivity $s = 1$, which is allowed as we may simply replace the cost $c(x)$ with $c(sx)$.

As the search space for the KL-divergence optimization (11) is infinite-dimensional, we resort to a quantization approach. Our construction can be seen as a generalization of the 1-dimensional approach in [12]. We fix a large enough ball, which we divide into spherical shells of fixed small enough width. We require that the mechanism be constant over the individual spherical shells. Then, we impose geometric tails outside the fixed large ball. Formally, we introduce the following construction.

Definition 2. Fix two positive integers n and N, a constant $r \in (0,1)$, and a vector $\boldsymbol{p} = (p_0, p_1, \dots, p_N) \in [0,\infty)^{N+1}$ with $p_0 \ge \cdots \ge p_N$. Consider the partition of $\mathbb{R}$ by intervals $\{\mathcal{J}_{i,n} := \left[\frac{i}{n}, \frac{i+1}{n}\right)\}_{i \in \mathbb{N}}$. We define the piecewise-constant function

$$\widetilde{f}_{n,r,\boldsymbol{p}}(\rho) := \begin{cases} p_i, & \text{if } \rho \in \mathcal{J}_{i,n}, \text{with } i < N, \\ p_N r^{i-N}, & \text{if } \rho \in \mathcal{J}_{i,n}, \text{with } i \ge N. \end{cases} \tag{17}$$

We also define the density $f_{n,r,\boldsymbol{p}} : \mathbb{R}^m \to [0,\infty)$ by

$$f_{n,r,\boldsymbol{p}}(x) := \widetilde{f}_{n,r,\boldsymbol{p}}(\|x\|), \tag{18}$$

and associate with $f_{n,r,\boldsymbol{p}}$ the Borel measure $P_{n,r,\boldsymbol{p}}$ given by

$$P_{n,r,\boldsymbol{p}}(B) := \int_B f_{n,r,\boldsymbol{p}}(x)\, dx. \tag{19}$$

We show next that the optimal distribution among the family introduced in Definition 2 can be found via a simple finite-dimensional convex optimization problem. For each $(n, N, r) \in \mathbb{N}^2 \times (0,1)$, let $\mathscr{F}_{n,N,r}$ denote the family of mechanisms

$$\mathscr{F}_{n,N,r} := \left\{P_{n,r,\boldsymbol{p}} \;;\; \boldsymbol{p} \in [0,\infty)^{N+1}, P_{n,r,\boldsymbol{p}}(\mathbb{R}^m) = 1\right\}. \tag{20}$$

Denote also the optimal value

$$\mathrm{KL}^\star_{n,N,r}(C) := \inf_{\substack{P \in \mathscr{F}_{n,N,r} \\ \mathbb{E}_P[c] \le C}} \sup_{\|a\| \le 1} D(P \,\|\, T_a P). \tag{21}$$

To state our next result more compactly, we introduce the following shorthands. For each $s, \rho, \theta \ge 0$, let $H(s,\rho,\theta)$ denote the area of the triangle with side lengths s, ρ, and θ, i.e., $H(s,\rho,\theta) = 0$ if there is no triangle with such side lengths, and otherwise

$$H(s,\rho,\theta) := \frac{1}{4}\sqrt{(s+\rho+\theta)(s+\rho-\theta)(s-\rho+\theta)(-s+\rho+\theta)}. \tag{22}$$

For each $i, j, n \in \mathbb{N}$, denote the constant

$$\gamma_{i,j,n} := \int_{\mathcal{J}_{i,n}} \int_{\mathcal{J}_{j,n}} \theta\rho \cdot H(1,\rho,\theta)^{m-3}\, d\theta\, d\rho. \tag{23}$$

Also, denote the constants

$$c_{i,n} := \int_{\|x\| \in \mathcal{J}_{i,n}} c(x)\, dx. \tag{24}$$

Denote the open balls

$$\mathcal{B}(\rho) := \{x \in \mathbb{R}^m \,:\, \|x\| < \rho\}. \tag{25}$$

For integers $i \geq 0$ and $n \geq 1$, denote the volume of the spherical shell

$$v_{i,n} = \lambda\left(\mathcal{B}\left(\frac{i+1}{n}\right) \setminus \mathcal{B}\left(\frac{i}{n}\right)\right). \tag{26}$$

Denote also the volume of the unit ball

$$V_m := \lambda(\mathcal{B}(1)) = \frac{\pi^{m/2}}{\Gamma\left(\frac{m}{2}+1\right)}. \tag{27}$$

The following result shows that the optimization (21) required to numerically construct our proposed mechanism (i.e., finding the vector $\boldsymbol{p} \in \mathbb{R}_+^{N+1}$ for a fixed choice of (n, N, r)) can be carried out as a finite-dimensional convex optimization problem.

Theorem 4. *The optimization* (21) *can be rewritten as*

$$\underset{\boldsymbol{p}\in(0,\infty)^{N+1}}{\text{minimize}} \quad A_m \sum_{i,j\geq 0} \gamma_{i,j,n}\; p_i \log \frac{p_i}{p_j} \tag{28}$$

$$\text{subject to} \quad \sum_{i\geq 0} p_i v_{i,n} = 1 \tag{29}$$

$$\sum_{i\geq 0} p_i c_{i,n} \leq C, \tag{30}$$

where $A_m = 2^{m-3}(m-1)V_{m-1}$ *and* $p_i = p_N r^{i-N}$ *for* $i > N$.

See Figure 1 for the result of numerically solving the optimization problem in Theorem 4 in $m = 10$ dimensions. Since this mechanism is spherically-symmetric, we plot its radius' density (i.e., with the decomposition $Z = R \cdot U$ as in (13), we plot the PDF of R) and compare to the multivariate Gaussian radial distribution.[2]

Finally, we prove optimality of our proposed mechanisms introduced in Definition 2 for the optimization problem (11) among monotone mechanisms (see Remark 4).

Theorem 5. *Suppose* $c : \mathbb{R}^m \to \mathbb{R}$ *satisfies Assumption 1, and suppose* c *is also continuous and that, for some* $\alpha, \beta > 0$, $c(x) \sim \beta\|x\|^\alpha$ *as* $\|x\| \to \infty$. *With the optimal value obtainable by the families* $\mathscr{F}_{n,N,r}$ *denoted by*

$$\widehat{\mathrm{KL}}(C) := \lim_{\theta\to 0^+} \inf_{(n,N,r)\in\mathbb{N}^2\times(0,1)} \mathrm{KL}^\star_{n,N,r}(C+\theta), \tag{31}$$

we have the equality $\mathrm{KL}^\star_{\text{monotone}} = \widehat{\mathrm{KL}}$.

VI. Numerical Experiments

We apply state-of-the-art accounting methods and privacy-amplification techniques to simulate a real-world application for the proposed mechanism in Definition 2. In particular, we subsample our mechanism, following standard practice in the DP machine learning literature for amplifying privacy guarantees [11], [14], [15]. Moreover, we use the arbitrary-accuracy FFT-based numerical accountant introduced in [16] to compute tight privacy bounds for finite compositions.

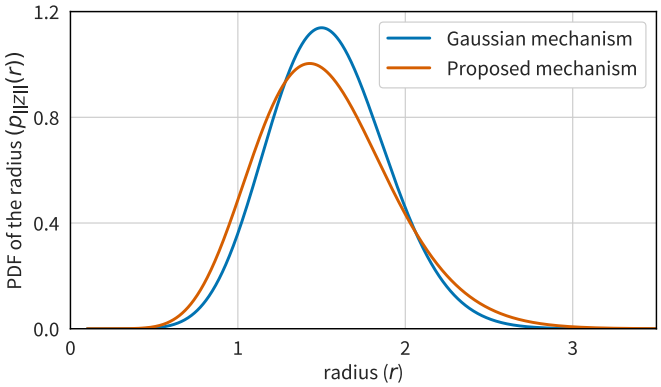

Fig. 1. The proposed mechanism and Gaussian mechanism, both in $m = 10$ dimensions and with a quadratic cost $\mathbb{E}[\|Z\|^2] = 2.5$. The construction parameters for the proposed mechanism are $n = 400$, $N = 1200$, and $r = 0.9$, and $\boldsymbol{p} \in (0,\infty)^{N+1}$ is found via Theorem 4 (see Definition 2).

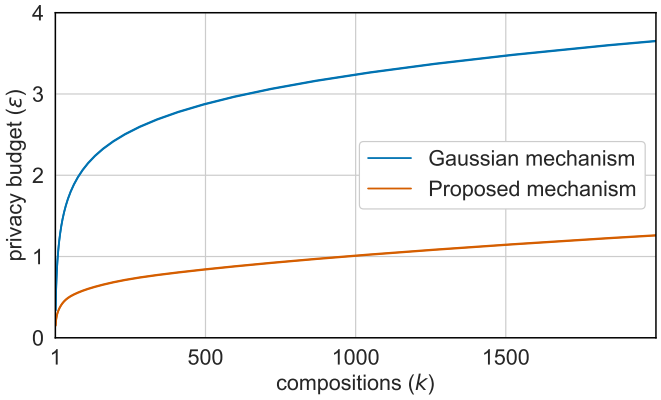

Fig. 2. Privacy budget ε versus number of the compositions, for $\mathbb{E}[\|Z\|^2] = 2.5$ (corresponding to $\sigma = 0.5$ for $\mathcal{N}(0, \sigma^2 I_{10})$), $\delta = 10^{-8}$, and subsampling rate $q = 0.001$. The proposed mechanism has 10 dimensions, and its construction parameters are $(n, N, r) = (400, 1200, 0.9)$, whereas its vector $\boldsymbol{p} \in (0,\infty)^{N+1}$ is computed numerically with the aid of Theorem 4.

In Figure 2, we fix $\delta = 10^{-8}$ and compute ε under a varying number of compositions. Under this construction, the accountant in [16] computes both upper and lower bounds on ε. This accountant allows one to set the additive error in ε and δ via the parameters $\varepsilon_{\text{error}}, \delta_{\text{error}}$. We choose $\varepsilon_{\text{error}} = 0.002$ and $\delta_{\text{error}} = 10^{-10}$, effectively making the upper and lower bounds indistinguishable (they are both plotted in Figure 2). We compare the resulting privacy curve for the proposed mechanism with that of the subsampled Gaussian mechanism, for the same dimension $m = 10$ and variance cost $\mathbb{E}[\|Z\|^2] = 2.5$. Our proposed mechanism provides stronger privacy guarantees[3] for all values of compositions $1 \leq k \leq 2000$.

[2]Note that both mechanisms in Figure 1 are monotone according to Definition 1, but this generally does not imply monotonicity of the PDF of the radial part of the random vectors.

[3]Although our proposed mechanism is not optimized for subsampling, our numerical results imply that it still outperforms the subsampled Gaussian.

REFERENCES

[1] W. Alghamdi, S. Asoodeh, F. P. Calmon, J. F. Gomez, O. Kosut, and L. Sankar, "Optimal multidimensional differentially private mechanisms in the large-composition regime," 2023. [Online]. Available: https://github.com/WaelAlghamdi/DP-Isotropic

[2] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory of Cryptography (TCC)*, Berlin, Heidelberg, 2006, pp. 265–284.

[3] A. Ghosh, T. Roughgarden, and M. Sundararajan, "Universally utility-maximizing privacy mechanisms," *SIAM Journal on Computing*, vol. 41, no. 6, pp. 1673–1693, 2012.

[4] M. Gupte and M. Sundararajan, "Universally optimal privacy mechanisms for minimax agents," in *Proceedings of the Twenty-Ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 2010, pp. 135–146.

[5] Q. Geng and P. Viswanath, "The optimal noise-adding mechanism in differential privacy," *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 925–951, 2015.

[6] Q. Geng, P. Kairouz, S. Oh, and P. Viswanath, "The staircase mechanism in differential privacy," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 7, pp. 1176–1184, 2015.

[7] J. Soria-Comas and J. Domingo-Ferrer, "Optimal data-independent noise for differential privacy," *Information Sciences*, vol. 250, pp. 200–214, 2013.

[8] Q. Geng and P. Viswanath, "Optimal noise adding mechanisms for approximate differential privacy," *IEEE Transactions on Information Theory*, vol. 62, no. 2, pp. 952–969, 2016.

[9] Q. Geng, W. Ding, R. Guo, and S. Kumar, "Tight analysis of privacy and utility tradeoff in approximate differential privacy," in *Proc. International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, S. Chiappa and R. Calandra, Eds., vol. 108, 2020, pp. 89–99.

[10] ——, "Optimal noise-adding mechanism in additive differential privacy," in *Proc. International Conference on Artificial Intelligence and Statistics*, K. Chaudhuri and M. Sugiyama, Eds., vol. 89, 2019, pp. 11–20.

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

[12] W. Alghamdi, S. Asoodeh, F. P. Calmon, O. Kosut, L. Sankar, and F. Wei, "Cactus mechanisms: Optimal differential privacy mechanisms in the large-composition regime," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 1838–1843.

[13] D. M. Sommer, E. Mohammadi, and S. Meiser, "Privacy loss classes: The central limit theorem in differential privacy," *Proceedings on Privacy Enhancing Technologies*, vol. 2019, pp. 245–269, 2019.

[14] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM J. Comput.*, vol. 40, no. 3, pp. 793–826, jun 2011.

[15] A. Beimel, K. Nissim, and U. Stemmer, "Characterizing the sample complexity of private learners," in *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*, 2013, pp. 97–110.

[16] S. Gopi, Y. T. Lee, and L. Wutschitz, "Numerical composition of differential privacy," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.