Optimal Noise-Adding Mechanism in Additive Differential Privacy

Quan Geng
Google AI

Wei Ding
Google AI

Ruiqi Guo
Google AI

Sanjiv Kumar
Google AI

Abstract

We derive the optimal $(0,\delta)$-differentially private query-output independent noise-adding mechanism for single real-valued query function under a general cost-minimization framework. Under a mild technical condition, we show that the optimal noise probability distribution is a uniform distribution with a probability mass at the origin. We explicitly derive the optimal noise distribution for general ℓ^p cost functions, including ℓ^1 (for noise magnitude) and ℓ^2 (for noise power) cost functions, and show that the probability concentration on the origin occurs when $\delta > \frac{p}{p+1}$. Our result demonstrates an improvement over the existing Gaussian mechanisms by a factor of two and three for $(0,\delta)$-differential privacy in the high privacy regime in the context of minimizing the noise magnitude and noise power, and the gain is more pronounced in the low privacy regime. Our result is consistent with the existing result for $(0,\delta)$-differential privacy in the discrete setting, and identifies a probability concentration phenomenon in the continuous setting.

1 Introduction

Differential privacy, introduced by Dwork et al. (2006b), is a framework to quantify to what extent individual privacy in a statistical dataset is preserved while releasing useful aggregate information about the dataset. Differential privacy provides strong privacy guarantees by requiring the near-indistinguishability of whether an individual is in the dataset or not based on the released information. For more motivation and background of differential privacy, we refer the readers to the survey by Dwork (2008) and the book by Dwork and Roth (2014).

The classic differential privacy is called ϵ-differential privacy, which imposes an upper bound e^{ϵ} on the multiplicative distance of the probability distributions of the randomized query outputs for any two neighboring datasets, and the standard approach for preserving ϵ-differential privacy is to add a Laplacian noise to the query output. Since its introduction, differential privacy has spawned a large body of research in differentially private data-releasing mechanism design, and the noise-adding mechanism has been applied in many machine learning algorithms to preserve differential privacy, e.g., logistic regression (Chaudhuri and Monteleoni, 2008), empirical risk minimization (Chaudhuri et al., 2011), online learning (Jain et al., 2012), statistical risk minimization (Duchi et al., 2012), deep learning (Shokri and Shmatikov, 2015; Abadi et al., 2016; Phan et al., 2016; Agarwal et al., 2018), hypothesis testing (Sheffet, 2018), matrix completion (Jain et al., 2018), expectation maximization (Park et al., 2017), and principal component analysis (Chaudhuri et al., 2012; Ge et al., 2018).

To fully make use of the randomized query outputs, it is important to understand the fundamental trade-off between privacy and utility (accuracy). Ghosh et al. (2009) studied a very general utility-maximization framework for a single count query with sensitivity one under ϵ-differential privacy. Gupte and Sundararajan (2010) derived the optimal noise probability distributions for a single count query with sensitivity one for minimax (risk-averse) users. Geng and Viswanath (2016b) derived the optimal ϵ-differentially private noise adding mechanism for single real-valued query function with arbitrary query sensitivity, and show that the optimal noise distribution has a staircase-shaped probability density function. Geng et al. (2015) generalized the result in Geng and Viswanath (2016b) to two-dimensional query output space for the ℓ^1 cost function, and show the optimality of a two-dimensional staircase-shaped probability density function. Soria-Comas and Domingo-Ferrer (2013) also independently derived the staircase-shaped noise probability distri-

Proceedings of the 22$^{\text{nd}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2019, Naha, Okinawa, Japan. PMLR: Volume 89. Copyright 2019 by the author(s).

bution under a different optimization framework.

A relaxed notion of ϵ-differential privacy is (ϵ, δ)-differential privacy, introduced by Dwork et al. (2006a). The common interpretation of (ϵ, δ)-differential privacy is that it is ϵ-differential privacy "except with probability δ" (Mironov, 2017). The standard approach for preserving (ϵ, δ)-differential privacy is the Gaussian mechanism, which adds a Gaussian noise to the query output. Geng and Viswanath (2016a) studied the trade-off between utility and privacy for a single *integer-valued* query function in (ϵ, δ)-differential privacy. Geng and Viswanath (2016a) show that for ℓ^1 and ℓ^2 cost functions, the discrete uniform noise distribution is optimal for $(0, \delta)$-differential privacy when the query sensitivity is one, and is *asymptotically* optimal as $\delta \to 0$ for arbitrary query sensitivity. Geng et al. (2018) extend the result for single real-valued query functions under (ϵ, δ)-differential privacy and show that the truncated Laplacian mechanism is asymptotically optimal in various high privacy regimes. Balle and Wang (2018) improved the classic analysis of the Gaussian mechanism for (ϵ, δ)-differential in the high privacy regime ($\epsilon \to 0$), and develops an optimal Gaussian mechanism whose variance is calibrated directly using the Gaussian cumulative density function instead of a tail bound approximation.

$(\epsilon, 0)$-differential privacy and $(0, \delta)$-differential privacy can be viewed as two special cases of the commonly used (ϵ, δ)-differential privacy paradigm. While $(\epsilon, 0)$-differential privacy is well studied and exact optimality result has been obtained, little is known about $(0, \delta)$-differential privacy. Characterizing the privacy-utility tradeoff in $(0, \delta)$-differential privacy is important towards understanding the fundamental privacy and utility tradeoff in (ϵ, δ)-differential privacy.

1.1 Our Contributions

In this work, we characterize the fundamental tradeoff between privacy and utility in $(0, \delta)$-differential privacy for a single read-valued query function. Within the class of query-output independently noise-adding mechanisms, we derive the optimal noise distribution for $(0, \delta)$-differential privacy under a general cost-minimization framework similar to Ghosh et al. (2009); Gupte and Sundararajan (2010); Geng and Viswanath (2014); Geng et al. (2015); Geng and Viswanath (2016a). Under a mild technical condition on the noise probability distribution[1], we show that the optimal noise probability distribution is a uniform distribution with a probability mass at the origin, which can be viewed as the distribution of the product of a uniform random variable and a Bernoulli random variable. The probability mass on the origin can be zero or non-zero, depending on the value of δ. We explicitly derive the optimal noise distribution for general ℓ^p cost functions, including ℓ^1 (for noise magnitude) and ℓ^2 (for noise power) cost functions, and show that the probability concentration on the origin occurs when $\delta > \frac{p}{p+1}$.

Compared with the improved Gaussian mechanisms for $(0, \delta)$-differential privacy (Balle and Wang, 2018), our result demonstrates a two-fold and three-fold improvement in the high privacy regime in the context of minimizing the noise magnitude and noise power, respectively. The improvement is more pronounced in the low privacy regime.

Comparing the exact optimality results of ϵ-differential privacy and $(0, \delta)$-differential privacy, we show that given the same amount of privacy constraint, $(0, \delta)$-differential privacy yields a higher utility than ϵ-differential privacy in the high privacy regime.

Our result is consistent with the existing result for $(0, \delta)$-differential privacy in the discrete setting (Geng and Viswanath, 2016a) which shows that the discrete uniform distribution is optimal for an *integer-valued* query function when the query sensitivity is one, and asymptotically optimal as $\delta \to 0$ for general query sensitivity. Interestingly, our result identifies a probability concentration phenomenon in the continuous setting for single *real-valued* query function.

1.2 Organization

The paper is organized as follows. In Section 2, we give some preliminaries on differential privacy, and formulate the trade-off between privacy and utility under $(0, \delta)$-differential privacy for a single real-valued query function as a functional optimization problem. Section 3 presents the optimal noise probability distribution preserving $(0, \delta)$-differential privacy, subject to a mild technical condition. Section 4 applies our main result to a class of momentum cost functions, and derives the explicit forms of the optimal noise probability distributions with minimum noise magnitude and noise power, respectively. Section 5 compares our result with the improved Gaussian mechanism in the context of minimizing noise magnitude and noise power.

[1] In this work, we assume that the noise probability distribution has higher probability over the small noise than the big noise. This condition is satisfied by virtually all probability distributions used in differential privacy, including the uniform distribution, the Laplacian distribution, the truncated Laplacian distribution, the Gaussian distribution, and the staircase distribution. While the optimality result in this paper depends on this assumption, we believe this assumption can be done away.

2 Problem Formulation

In this section, we first give some preliminaries on differential privacy, and then formulate the trade-off between privacy and utility under $(0, \delta)$-differential privacy for a single real-valued query function as a functional optimization problem.

2.1 Background on Differential Privacy

Consider a real-valued query function

$$q : \mathcal{D} \to \mathbb{R},$$

where $\mathcal{D}$ is the set of all possible datasets. The real-valued query function q will be applied to a dataset, and the query output is a real number. Two datasets $D_1, D_2 \in \mathcal{D}$ are called neighboring datasets if they differ in at most one element, i.e., one is a proper subset of the other and the larger dataset contains just one additional element (Dwork, 2008). A randomized query-answering mechanism $\mathcal{K}$ for the query function q will randomly output a number with probability distribution depending on query output $q(D)$, where D is the dataset.

Definition 1 (ϵ-differential privacy (Dwork, 2008)). *A randomized mechanism $\mathcal{K}$ gives ϵ-differential privacy if for all data sets D_1 and D_2 differing on at most one element, and any measurable set $S \subset Range(\mathcal{K})$,*

$$Pr[\mathcal{K}(D_1) \in S] \le e^{\epsilon}\ Pr[\mathcal{K}(D_2) \in S], \tag{1}$$

where $\mathcal{K}(D)$ is the random output of the mechanism $\mathcal{K}$ when the query function q is applied to the dataset D.

The differential privacy constraint (1) imposes an upper bound e^{ϵ} on the multiplicative distance of the two probability distributions. It essentially requires that for all neighboring datasets, the probability distributions of the output of the randomized mechanism should be approximately the same. Therefore, for any individual record, its presence or absence in the dataset will not significantly affect the output of the mechanism, which makes it hard for adversaries with arbitrary background knowledge to make inference on any individual from the released query output information. The parameter $\epsilon \in (0, +\infty)$ quantifies how private the mechanism is: the smaller ϵ is, the more private the randomized mechanism is.

The standard approach to preserving ϵ-differential privacy is to perturb the query output by adding a random noise with Laplacian distribution proportional to the sensitivity Δ of the query function q, where the sensitivity of a real-valued query function is defined as:

Definition 2 (Query Sensitivity (Dwork, 2008)). *For a real-valued query function $q : \mathcal{D} \to \mathbb{R}$, the sensitivity of q is defined as*

$$\Delta := \max_{D_1, D_2 \in \mathcal{D}} |q(D_1) - q(D_2)|,$$

for all D_1, D_2 differing in at most one element.

Introduced by Dwork et al. (2006a), a relaxed version of ϵ-differential privacy is (ϵ, δ)-differential privacy, which relaxes the constraint (1) with an additive term $\delta \in [0, 1]$.

Definition 3 ((ϵ, δ)-differential privacy (Dwork et al., 2006a)). *A randomized mechanism $\mathcal{K}$ gives (ϵ, δ)-differential privacy if for all data sets D_1 and D_2 differing on at most one element, and all $S \subset Range(\mathcal{K})$,*

$$Pr[\mathcal{K}(D_1) \in S] \le e^{\epsilon}\ Pr[\mathcal{K}(D_2) \in S] + \delta.$$

In the special case where $\epsilon = 0$, the constraint for $(0, \delta)$-differential privacy is

$$\Pr[\mathcal{K}(D_1) \in S] \le \Pr[\mathcal{K}(D_2) \in S] + \delta. \tag{2}$$

It is ready to see that $(0, \delta)$-differential privacy puts an upper bound δ on the *additive* distance of the two probability distributions.

2.2 $(0, \delta)$-Differential Privacy Constraint on the Noise Probability Distribution

A standard approach for preserving differential privacy is query-output independent noise-adding mechanisms, where a random noise is added to the query output. Given a dataset D, a query-output independent noise-adding mechanism $\mathcal{K}$ will release the query output $t = q(D)$ corrupted by an additive random noise X with probability distribution $\mathcal{P}$:

$$\mathcal{K}(D) = t + X. \tag{3}$$

The $(0, \delta)$-differential privacy constraint (2) on $\mathcal{K}$ is that for any $t_1, t_2 \in \mathbb{R}$ such that $|t_1 - t_2| \le \Delta$ (corresponding to the query outputs for two neighboring datasets),

$$\mathcal{P}(S) \le \mathcal{P}(S + t_1 - t_2) + \delta, \forall \text{ measurable set } S \subset \mathbb{R},$$

where $\forall t \in \mathbb{R}$, $S + t$ is defined as the set $\{s + t \mid s \in S\}$.

Equivalently, the $(0, \delta)$-differential privacy constraint on the noise probability distribution $\mathcal{P}$ is

$$\mathcal{P}(S) \le \mathcal{P}(S + d) + \delta, \forall\ |d| \le \Delta, \text{measurable set } S \subset \mathbb{R}. \tag{4}$$

2.3 Utility Model

Consider a cost function $\mathcal{L}(\cdot) : \mathbb{R} \to \mathbb{R}$, which is a function of the additive noise in the query-output noise-adding mechanism. Given an additive noise x, the cost is $\mathcal{L}(x)$, and thus the expectation of the cost over $\mathcal{P}$ is

$$\int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx). \tag{5}$$

Our objective is to minimize the expectation of the cost over the noise probability distribution for preserving $(0, \delta)$-differential privacy.

2.4 Optimization Problem

Combining the differential privacy constraint (4) and the objective function (5), we formulate a functional optimization problem:

$$\underset{\mathcal{P}}{\text{minimize}} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) \tag{6}$$

subject to $\forall$ measurable set $S \subseteq \mathbb{R}$, $\forall |d| \leq \Delta$.

$$|\mathcal{P}(S) - \mathcal{P}(S + d)| \leq \delta \tag{7}$$

3 Main Result

In this section, we solve the functional optimization problem (6) for $(0, \delta)$-differential privacy, and present our main result in Theorem 2. Under a mild technical condition on the probability distribution (see Property 2), we show that the optimal noise probability distribution is a uniform distribution with a probability mass at the origin, which can be viewed as the distribution of the product of a uniform random variable and a Bernoulli random variable.

We assume that the cost function $\mathcal{L}(\cdot)$ satisfies a natural property.

Property 1. *$\mathcal{L}(x)$ is a symmetric function, and monotonically increasing for $x \geq 0$, i.e, $\mathcal{L}(x)$ satisfies*

$$\mathcal{L}(x) = \mathcal{L}(-x), \forall\, x \in \mathbb{R},$$

and

$$\mathcal{L}(x) \leq \mathcal{L}(y), \forall\, 0 \leq x \leq y.$$

First, we show that without loss of generality, we only need to consider symmetric noise probability distributions.

Lemma 1. *Given a noise probability distribution $\mathcal{P}$ satisfying (7), there exists a probability distribution $\hat{\mathcal{P}}$ such that $\hat{\mathcal{P}}$ satisfies (7), and*

$$\hat{\mathcal{P}}(S) = \hat{\mathcal{P}}(-S), \forall \textit{ measurable set } S \subseteq \mathbb{R},$$

and

$$\int_{x \in \mathbb{R}} \mathcal{L}(x)\hat{\mathcal{P}}(dx) = \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx).$$

Proof. Define $\hat{\mathcal{P}}$ as follows: $\forall$ measurable set$S \subseteq \mathbb{R}$,

$$\hat{\mathcal{P}}(S) := \frac{\mathcal{P}(S) + \mathcal{P}(-S)}{2}.$$

It is ready to see $\hat{\mathcal{P}}$ is a symmetric probability distribution. As the loss function $\mathcal{L}(\cdot)$ is symmetric, we have $\int_{x \in \mathbb{R}} \mathcal{L}(x)\hat{\mathcal{P}}(dx) = \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx)$.

Next we show that $\hat{\mathcal{P}}$ also satisfies the differential privacy constraint. Indeed, $\forall$ *measurable set* $S \subseteq \mathbb{R}$ and d such that $|d| \leq \Delta$, we have

$$\begin{aligned}
&|\hat{\mathcal{P}}(S) - \hat{\mathcal{P}}(S + d)| \\
&= |\frac{\mathcal{P}(S) + \mathcal{P}(-S)}{2} - \frac{\mathcal{P}(S + d) + \mathcal{P}(-S - d)}{2}| \\
&\leq \frac{|\mathcal{P}(S) - \mathcal{P}(S + d)|}{2} + \frac{|\mathcal{P}(-S) + \mathcal{P}(-S - d)|}{2} \\
&\leq \frac{\delta}{2} + \frac{\delta}{2} \\
&= \delta.
\end{aligned}$$

□

Due to Lemma 1, we can restrict ourselves to symmetric noise probability distributions.

As the loss function $\mathcal{L}(\cdot)$ is monotonically increasing as the noise becomes bigger, we impose a mild and natural condition on the symmetric noise probability contribution, which requires the noise probability distribution to have bigger probability measure on the small noise than the large noise. More precisely,

Property 2. *Given a symmetric probability measure $\mathcal{P} \in \mathcal{P}_{sym}$, $\mathcal{P}$ is monotonically decreasing if*

$$\mathcal{P}(S) \geq \mathcal{P}(S + a), \forall a \geq 0, \textit{ measurable set } S \subseteq [0, +\infty)$$

Property 2 is satisfied by a large class of probability distributions, including the uniform distribution, the Laplacian distribution and the Gaussian distribution.

Let $\mathcal{P}_{\text{sym, mon}}$ denote the set of symmetric probability measures which are monotonically decreasing.

Lemma 2. *Given $\mathcal{P} \in \mathcal{P}_{sym,\ mon}$, then for any $t \neq 0$, $\mathcal{P}(\{t\}) = 0$, i.e., $\mathcal{P}$ cannot have a non-zero probability mass on any singular point except the origin $t = 0$.*

Proof. Suppose there exists $t \neq 0$ such that $\mathcal{P}(\{t\}) \neq 0$. Since $\mathcal{P}$ is symmetric, we can assume $t > 0$. Since $\mathcal{P}$ is monotonically decreasing in $[0, +\infty)$, for any

$t' \in (0,t), \mathcal{P}(\{t'\}) \geq \mathcal{P}(\{t\})$, which implies $\mathcal{P}((0,t)) = +\infty$. This contradicts with the fact that $\mathcal{P}$ is a probability measure. □

Within the classes of monotonically decreasing probability distributions, we identify a sufficient and necessary condition for preserving $(0,\delta)$-differential privacy.

Theorem 1. *Given $\mathcal{P} \in \mathcal{P}_{sym,\ mon}$, $\mathcal{P}$ satisfies the $(0,\delta)$-differential privacy constraint* (7)*, if and only if*

$$\mathcal{P}([-\frac{\Delta}{2}, \frac{\Delta}{2}]) \leq \delta.$$

Proof. First we show that it is a necessary condition. Assume $\mathcal{P}$ satisfies the $(0,\delta)$-differential privacy constraint (7). Consider $S = [-\frac{\Delta}{2}, +\infty)$ and $d = \Delta$ in (7), and we have

$$|\mathcal{P}([-\frac{\Delta}{2}, +\infty)) - \mathcal{P}([\frac{\Delta}{2}, +\infty))| \leq \delta,$$

and thus $\mathcal{P}([-\frac{\Delta}{2}, \frac{\Delta}{2})) \leq \delta$. Due to Lemma 2, $\mathcal{P}([-\frac{\Delta}{2}, \frac{\Delta}{2}]) = \mathcal{P}([-\frac{\Delta}{2}, \frac{\Delta}{2})) \leq \delta$.

Next we show that it is a sufficient condition. Assume $\mathcal{P}([-\frac{\Delta}{2}, \frac{\Delta}{2}]) \leq \delta$. As $\mathcal{P}$ is symmetric and the differential privacy constraint (7) applies to all measurable subset $S \subseteq \mathbb{R}$ and all d such that $|d| \leq \Delta$, it is equivalent to show that $\mathcal{P}(S) - \mathcal{P}(S+d) \leq \delta, \forall d \in (0,\Delta]$.

Since $\mathcal{P}$ is symmetric and monotonically decreasing in $[0,+\infty)$, $\mathcal{P}(S) - \mathcal{P}(S+d)$ is maximized when $S = [-\frac{d}{2}, +\infty)$. Therefore, $\forall d \in (0,\Delta]$,

$$\begin{aligned}
&\mathcal{P}(S) - \mathcal{P}(S+d) \\
&\leq \mathcal{P}([-\frac{d}{2}, +\infty)) - \mathcal{P}([-\frac{d}{2}, +\infty) + d) \\
&= \mathcal{P}([-\frac{d}{2}, \frac{d}{2})) \\
&\leq \mathcal{P}([-\frac{\Delta}{2}, \frac{\Delta}{2})) \\
&\leq \delta.
\end{aligned}$$

This concludes the proof of Theorem 1. □

Consider a class of probability distributions $\{P_\alpha\}$ parameterized by $\alpha \in [0,\delta)$, where $\mathcal{P}_\alpha$ is defined as

$$\mathcal{P}_\alpha(\{0\}) = \alpha$$

and except the point $t = 0$, $\mathcal{P}_\alpha$ has a uniform probability distribution over the set $[-\frac{1-\alpha}{\delta-\alpha}\frac{\Delta}{2}, \frac{1-\alpha}{\delta-\alpha}\frac{\Delta}{2}] \setminus \{0\}$ with probability density $\frac{\delta-\alpha}{\Delta}$ (see Figure 1).

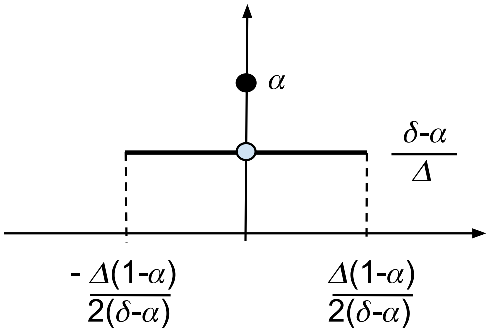
α
$\frac{\delta\text{-}\alpha}{\Delta}$
$-\frac{\Delta(1\text{-}\alpha)}{2(\delta\text{-}\alpha)}$
$\frac{\Delta(1\text{-}\alpha)}{2(\delta\text{-}\alpha)}$

Figure 1: Probability distribution of $\mathcal{P}_\alpha$. $\mathcal{P}_\alpha$ has a probability mass $\alpha \in [0,\delta)$ at the origin, and has a uniform distribution over $[-\frac{1-\alpha}{\delta-\alpha}\frac{\Delta}{2}, \frac{1-\alpha}{\delta-\alpha}\frac{\Delta}{2}] \setminus \{0\}$ with probability density $\frac{\delta-\alpha}{\Delta}$.

Let $\mathcal{SP}$ denote the set of all symmetric and monotonically decreasing probability distributions satisfying the $(0,\delta)$-differential privacy (7). Our main result on the optimal noise probability distribution is:

Theorem 2. *If the cost function $\mathcal{L}(x)$ satisfies Property 1, then for any $\Delta > 0$ and $0 < \delta < 1$,*

$$\inf_{\mathcal{P} \in \mathcal{SP}} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) = \inf_{\alpha \in [0,\delta)} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}_\alpha(dx).$$

Proof. First note that for any $\alpha \in [0,\delta)$, $\mathcal{P}_\alpha$ is symmetric and monotonically decreasing in $[0,+\infty)$, and

$$\mathcal{P}_\alpha([-\frac{\Delta}{2}, \frac{\Delta}{2}]) = \alpha + \frac{\delta-\alpha}{\Delta}\Delta = \delta.$$

Therefore, due to Theorem 1, $\mathcal{P}_\alpha$ satisfies the $(0,\delta)$-differential privacy constraint (7), and thus $\mathcal{P}_\alpha \in \mathcal{SP}$.

Applying a similar argument as in Lemma 20 of Geng and Viswanath (2016b), we can use a sequence of symmetric and piece-wise linear probability density function with probability mass concentration in the origin to approximate any $\mathcal{P} \in \mathcal{SP}$ (see Figure 2). More precisely, given a probability distribution $\mathcal{P} \in \mathcal{SP}$ which may have non-zero probability mass at $x = 0$, for positive integer $i \in N$, define the probability distribution $\mathcal{P}_i$ as follows:

$$\mathcal{P}_i(\{0\}) := \mathcal{P}(\{0\})$$

and over the set $\mathbb{R}\setminus\{0\}$, $\mathcal{P}_i$ has a symmetric probability density function $f_i(x)$ with

$$f_i(x) = \begin{cases} a_k \triangleq \frac{\mathcal{P}((k\frac{\Delta}{2i}, (k+1)\frac{\Delta}{2i}])}{\frac{\Delta}{2i}} & x \in (k\frac{\Delta}{2i}, (k+1)\frac{\Delta}{2i}], \\ f_i(-x) & x < 0 \end{cases}$$

It is easy to see that $\mathcal{P}_i([-\frac{\Delta}{2}, \frac{\Delta}{2}]) = \mathcal{P}([-\frac{\Delta}{2}, \frac{\Delta}{2}]) \leq \delta$, and thus due to Theorem 1, $\mathcal{P}_i \in \mathcal{SP}$. Due to the definition of Riemann-Stieltjes integral, we have

$$\lim_{i \to +\infty} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_i(dx) = \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}(dx)$$

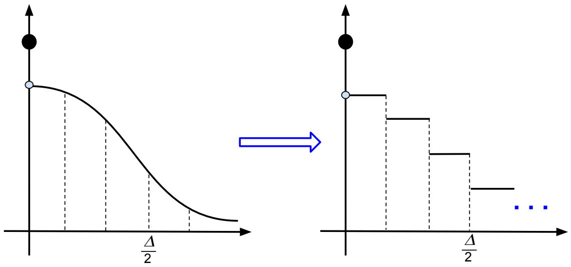

Figure 2: Discretize a probability distribution with a piecewise-constant probability density function and a probability mass at the origin.

Therefore, we only need to consider probability distributions with a probability mass on the origin and a symmetric monotonically decreasing piecewise-constant probability density function on $\mathbb{R} \setminus \{0\}$.

First we show that without loss of generality, we can assume the probability density function in $[\frac{\Delta}{2}, +\infty)$ is a step function, i.e., there exists a $t^* > \frac{\Delta}{2}$ such that probability density function is a constant in $[\frac{\Delta}{2}, t^*]$ and is zero in $(t^*, +\infty)$. Indeed, we can re-arrange the probability distribution in $[\frac{\Delta}{2}, +\infty)$ to make the probability density function to be uniform within certain interval $[\frac{\Delta}{2}, t^*]$ with the probability density the same as the previous bucket (see Figure 3). This will not increase the cost, due to the fact that $\mathcal{L}(\cdot)$ is a monotonically increasing function on $[0, +\infty)$. Since we are not changing the probability distribution in $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$, due to Theorem 1, the probability distribution after the re-arrangement also satisfies the differential privacy constraint (7).

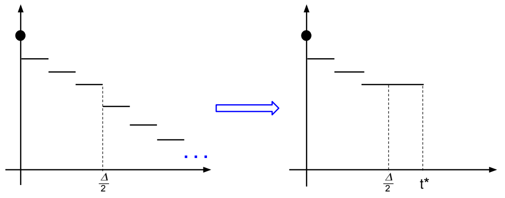

Figure 3: Re-arrange the probability distribution in $[\frac{\Delta}{2}, +\infty)$ to be a step.

Then we show that the probability distribution in $(0, \frac{\Delta}{2})$ shall be uniform as well. Indeed, if the distribution in $(0, \frac{\Delta}{2})$ is not uniform, we can decrease the probability density over $(0, \frac{\Delta}{2})$ to be the same as the point $\frac{\Delta}{2}$, and move the extra probability mass to the origin point (see Figure 4). Due to the fact that $\mathcal{L}(\cdot)$ is a monotonically increasing function on $[0, +\infty)$, this will not increase the cost. As $\mathcal{P}([-\frac{\Delta}{2}, \frac{\Delta}{2}])$ is unchanged, the new probability distribution satisfies the differential privacy constraint (7).

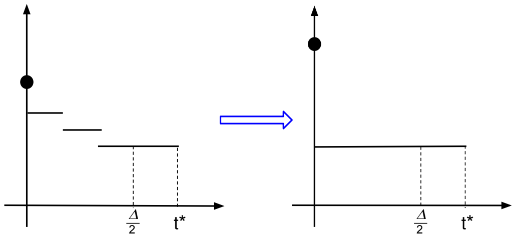

Figure 4: Re-arrange the probability distribution in $(0, \frac{\Delta}{2})$ to be uniform and put the extra probability mass at the origin.

In the last step, we show that $\mathcal{P}([-\frac{\Delta}{2}, \frac{\Delta}{2}])$ should be exactly δ. If it is strictly less than δ, then we can reduce the probability density over $(0, t^*)$ and increase α to make $\mathcal{P}([-\frac{\Delta}{2}, \frac{\Delta}{2}]) = \delta$. Similarly, due to the property of $\mathcal{L}(\cdot)$ and Theorem 1, we conclude that this reduces the cost while preserving the differential privacy constraint.

This concludes the proof of Theorem 2. □

A natural and simple algorithm to generate random noise with probability distribution $\mathcal{P}_\alpha$ is given in Algorithm 1.

Algorithm 1 Generation of a Random Variable of Uniform Distribution with a Probability Concentration on the Origin

```
**Input:** $\delta$, $\Delta$, and $\alpha \in [0, \delta)$.
**Output:** $X$, a random variable of uniform distribution with a probability concentration on the origin, paremeterized by $\delta, \Delta$ and $\alpha$.

Generate a binary random variable $B$ with
    $\Pr[B = 0] = \alpha,$
    $\Pr[B = 1] = 1 - \alpha.$

Generate a random variable $U$ uniformly distributed in $[-\frac{1-\alpha}{\delta-\alpha}\frac{\Delta}{2}, \frac{1-\alpha}{\delta-\alpha}\frac{\Delta}{2}]$.
$X \leftarrow B \cdot U$.
Output $X$.
```

4 Applications

In this section, we apply our main result Theorem 2 to derive an explicit expression for the parameter α in the optimal noise probability distribution $\mathcal{P}_\alpha$ for the class of ℓ^p momentum cost functions in Theorem 3. Applying Theorem 3 to the cases $p = 1$ and $p = 2$, we get the optimal noise probability distribution with minimum noise amplitude and minimum noise power for $(0, \delta)$-differential privacy, respectively.

Let $V(\mathcal{P}) := \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx)$, i.e., $V(\mathcal{P})$ denote the expectation of the cost given the noise probability distribution $\mathcal{P}$ for the cost function $\mathcal{L}(\cdot)$.

Theorem 3. *Given $0 < \delta < 1$ and the query sensitivity $\Delta > 0$. For the general momentum cost function $\ell^p(x) := |x|^p$, where $p > 0$, the optimal noise probability distribution to preserve $(0, \delta)$-differential privacy with query sensitivity Δ is $\mathcal{P}_\alpha*$ with*

$$\alpha^* = \begin{cases} 0, & \text{for } \delta \in (0, \frac{p}{p+1}] \\ (p+1)\delta - p, & \text{for } \delta \in (\frac{p}{p+1}, 1) \end{cases}$$

and the minimum cost is

$$V(\mathcal{P}_{\alpha^*}) = \begin{cases} \frac{\Delta^p}{2^p(p+1)\delta^p}, & \text{for } \delta \in (0, \frac{p}{p+1}] \\ \frac{(p+1)^p}{2^p p^p}(1-\delta)\Delta^p, & \text{for } \delta \in (\frac{p}{p+1}, 1) \end{cases}$$

Proof. It is easy to see that the momentum cost function satisfies Property 1. We can compute the cost $V(\mathcal{P}_\alpha)$ via

$$\begin{aligned} V(\mathcal{P}_\alpha) &= \int_{x\in\mathbb{R}} |x|^p \mathcal{P}_\alpha(dx) \\ &= 2\int_0^{\frac{1-\alpha}{\delta-\alpha}\frac{\Delta}{2}} x^p \frac{\delta-\alpha}{\Delta} dx \\ &= \frac{\Delta^p}{(p+1)2^p}\frac{(1-\alpha)^{p+1}}{(\delta-\alpha)^p}. \end{aligned}$$

Define $f(\alpha) := \frac{(1-\alpha)^{p+1}}{(\delta-\alpha)^p}$. As $f(\alpha)$ is a continuous function of α over $[0, \delta)$, and $f(0) = \frac{1}{\delta^p}$, and $f(\delta) = \infty$, the minimum is achieved at either $\alpha = 0$ or the point where the derivative $f'(\alpha) = 0$.

Compute the derivative of $f(\alpha)$ via

$$\begin{aligned} & f'(\alpha) \\ =& \frac{-(p+1)(1-\alpha)^p(\delta-\alpha)^p + (1-\alpha)^{p+1}p(\delta-\alpha)^{p-1}}{(\delta-\alpha)^{2p}} \\ =& \frac{(\delta-\alpha)^{p-1}(1-\alpha)^p}{(\delta-\alpha)^{2p}}(\alpha - (p+1)\delta + p). \end{aligned}$$

Set $f'(\alpha) = 0$ and we get $\alpha = (p+1)\delta - p$.

It is ready to calculate that

$$V(\mathcal{P}_\alpha) = \begin{cases} \dfrac{\Delta^p}{2^p(p+1)\delta^p}, & \text{for } \alpha = 0 \\ \dfrac{(p+1)^p}{2^p p^p}(1-\delta)\Delta^p, & \text{for } \alpha = (p+1)\delta - p \end{cases}$$

It is easy to see that when $\delta \le \frac{p}{p+1}$, at the point where the derivative is zero we have $\alpha = (p+1)\delta - p < 0$, and thus the minimum is achieved at $\alpha = 0$. When $\delta > \frac{p}{p+1}$, $(p+1)\delta - p \in (0, \delta)$, and we have $V(\mathcal{P}_0) \ge V(\mathcal{P}_{(p+1)\delta-p})$. Indeed,

$$\begin{aligned} & V(\mathcal{P}_0) \ge V(\mathcal{P}_{(p+1)\delta-p}) \\ \Leftrightarrow & \frac{\Delta^p}{2^p(p+1)\delta^p} \ge \frac{(p+1)^p}{2^p p^p}(1-\delta)\Delta^p \\ \Leftrightarrow & \frac{p^p}{(p+1)^{p+1}} \ge \delta^p(1-\delta), \end{aligned} \tag{8}$$

where (8) holds as

$$\begin{aligned} \delta^p(1-\delta) &= \frac{\delta^p(p - p\delta)}{p} \\ &\le \frac{(\frac{(p\delta+p-p\delta)}{p+1})^{p+1}}{p} \\ &= \frac{p^p}{(p+1)^{p+1}}. \end{aligned}$$

Therefore, when $\delta > \frac{p}{p+1}$, the minimum of $f(\alpha)$ is achieved at $\alpha = (p+1)\delta - p$.

In conclusion, for the ℓ^p cost function, the optimal α is

$$\alpha^* = \begin{cases} 0, & \text{for } \delta \in (0, \frac{p}{p+1}] \\ (p+1)\delta - p, & \text{for } \delta \in (\frac{p}{p+1}, 1) \end{cases}$$

and the minimum cost is

$$V(\mathcal{P}_{\alpha^*}) = \begin{cases} \frac{\Delta^p}{2^p(p+1)\delta^p}, & \text{for } \delta \in (0, \frac{p}{p+1}] \\ \frac{(p+1)^p}{2^p p^p}(1-\delta)\Delta^p, & \text{for } \delta \in (\frac{p}{p+1}, 1) \end{cases}$$

□

Applying Theorem 3 to the cases where $p = 1$ and $p = 2$, we derive the optimal noise probability distribution for $(0, \delta)$-differential privacy with minimum noise amplitude and minimum noise power, respectively.

Corollary 4 (Optimal Noise Amplitude)**.** *Given $0 < \delta < 1$ and the query sensitivity $\Delta > 0$, to minimize the expectation of the amplitude of the noise (i.e., for the ℓ^1 cost function), the optimal noise probability distribution is $\mathcal{P}_\alpha*$ with*

$$\alpha^* = \begin{cases} 0, & \text{for } \delta \in (0, \frac{1}{2}] \\ 2\delta - 1, & \text{for } \delta \in (\frac{1}{2}, 1) \end{cases}$$

and the minimum expectation of noise amplitude is

$$V(\mathcal{P}_{\alpha^*}) = \begin{cases} \frac{\Delta}{4\delta}, & \textit{for } \delta \in (0, \frac{1}{2}] \\ (1-\delta)\Delta, & \textit{for } \delta \in (\frac{1}{2}, 1) \end{cases} \tag{9}$$

Corollary 5 (Optimal Noise Power)**.** *Given* $0 < \delta < 1$ *and the query sensitivity* $\Delta > 0$, *to minimize the expectation of the power of the noise (i.e., for the* ℓ^2 *cost function), the optimal noise probability distribution is* $\mathcal{P}_\alpha *$ *with*

$$\alpha^* = \begin{cases} 0, & \textit{for } \delta \in (0, \frac{2}{3}] \\ 3\delta - 2, & \textit{for } \delta \in (\frac{2}{3}, 1) \end{cases}$$

and the minimum expectation of noise power is

$$V(\mathcal{P}_{\alpha^*}) = \begin{cases} \frac{\Delta^2}{12\delta^2}, & \textit{for } \delta \in (0, \frac{2}{3}] \\ \frac{9}{16}(1-\delta)\Delta^2, & \textit{for } \delta \in (\frac{2}{3}, 1) \end{cases} \tag{10}$$

5 Comparison to Gaussian Mechanism

In this section, we compare our result with the classic Gaussian mechanism, which adds a Gaussian noise to preserve (ϵ, δ)-differential privacy. We show a two-fold and three-fold improvement in the high privacy regime over the improved Gaussian mechanism in Balle and Wang (2018) for minimizing the noise magnitude and the noise power, and we show that the gain is more pronounced in the low privacy regime.

A classic result on the Gaussian mechanism is that for any $\epsilon, \delta \in (0, 1)$, adding a Gaussian noise with standard deviation $\sigma = \frac{\sqrt{2\log(1.25/\delta)}}{\epsilon}\Delta$ preserves (ϵ, δ)-differential privacy (Dwork and Roth, 2014). This result does not apply to the $(0, \delta)$-differential privacy, as this would require σ to be $+\infty$ when $\epsilon = 0$.

For (ϵ, δ)-differential privacy, Balle and Wang (2018) developed an optimal Gaussian mechanism whose variance is calibrated directly using the Gaussian cumulative density function instead of a tail bound approximation. Balle and Wang (2018) show the following:

Theorem 6 (Theorem 2 in Balle and Wang (2018))**.** *A Gaussian output perturbation mechanism with* $\sigma = \frac{\Delta}{2\delta}$ *preserves* $(0, \delta)$*-differential privacy.*

It is ready to see that the Gaussian noise distribution has an expected noise amplitude $\sigma = \frac{\Delta}{2\delta}$ and an expected noise power $\sigma^2 = \frac{\Delta^2}{4\delta^2}$.

Table 1: Noise Magnitude Comparison

Gaussian	$\frac{\Delta}{2\delta}$
Optimal	$\begin{cases} \frac{\Delta}{4\delta}, & \text{for } \delta \in (0, \frac{1}{2}] \\ (1-\delta)\Delta, & \text{for } \delta \in (\frac{1}{2}, 1) \end{cases}$

Table 2: Noise Power Comparison

Gaussian	$\frac{\Delta^2}{4\delta^2}$
Optimal	$\begin{cases} \frac{\Delta^2}{12\delta^2}, & \text{for } \delta \in (0, \frac{2}{3}] \\ \frac{9}{16}(1-\delta)\Delta^2, & \text{for } \delta \in (\frac{2}{3}, 1) \end{cases}$

For comparison, our result (9) and (10) in this paper show that the minimum expected noise magnitude and noise power are $\frac{\Delta}{4\delta}$ and $\frac{\Delta^2}{12\delta^2}$ in the medium/high privacy regime ($\delta \leq \frac{1}{2}$). Therefore, our result shows a two-fold and three-fold multiplicative gain over the improved Gaussian mechanism in Balle and Wang (2018) for $(0, \delta)$-differential privacy in the high privacy regime for minimizing the noise magnitude and the noise power, respectively.

In the low privacy regime, the gap is more pronounced: as $\delta \to 1$, the cost of the Gaussian mechanism converges to $\frac{\Delta}{2}$ and $\frac{\Delta^2}{4}$ for the noise magnitude and the noise power, while the cost of optimal noise from (9) and (10) converges to zero proportionally to $(1 - \delta)$.

We plot the ratio of the optimal noise magnitude and noise power over Gaussian mechanism in Fig. 5. We conclude that the derived optimal $(0, \delta)$-differential private mechanism in this work reduces the noise magnitude and noise power by $\frac{1}{2}$ and $\frac{2}{3}$ in the high privacy regime, and the improvement is more pronounced in the low privacy regime.

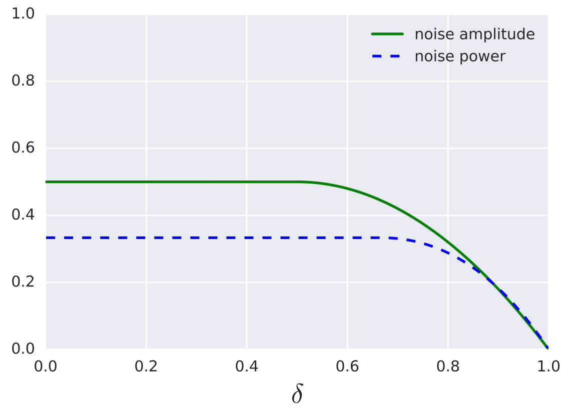

Figure 5: Ratio of the Optimal Noise Cost over the Improved Gaussian Mechanism.

Acknowledgment

We would like to thank the anonymous reviewers for their insightful comments and suggestions, which help us improve the presentation of this work.

References

Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 308–318. ACM, 2016.

Naman Agarwal, Ananda Theertha Suresh, Felix Yu, Sanjiv Kumar, and Brendan McMahan. cpSGD: Communication-efficient and differentially-private distributed SGD. In *Advances in Neural Information Processing Systems*. 2018.

Borja Balle and Yu-Xiang Wang. Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.

Kamalika Chaudhuri and Claire Monteleoni. Privacy-preserving logistic regression. In *Neural Information Processing Systems*, pages 289–296, 2008.

Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12:1069–1109, 2011.

Kamalika Chaudhuri, Anand Sarwate, and Kaushik Sinha. Near-optimal differentially private principal components. In *Advances in Neural Information Processing Systems 25*, pages 989–997. 2012.

John Duchi, Michael Jordan, and Martin Wainwright. Privacy aware learning. In *Advances in Neural Information Processing Systems*, pages 1430–1438, 2012.

Cynthia Dwork. Differential Privacy: A Survey of Results. In *Theory and Applications of Models of Computation*, volume 4978, pages 1–19, 2008.

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

Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: privacy via distributed noise generation. In *Proceedings of the 24th annual international conference on The Theory and Applications of Cryptographic Techniques*, EUROCRYPT'06, pages 486–503. Springer-Verlag, 2006a.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer Berlin / Heidelberg, 2006b.

Jason Ge, Zhaoran Wang, Mengdi Wang, and Han Liu. Minimax-optimal privacy-preserving sparse pca in distributed systems. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2018.

Quan Geng and Pramod Viswanath. The optimal mechanism in differential privacy. In *IEEE International Symposium on Information Theory (ISIT)*, pages 2371–2375, June 2014.

Quan Geng and Pramod Viswanath. Optimal noise adding mechanisms for approximate differential privacy. *IEEE Transactions on Information Theory*, 62 (2):952–969, Feb 2016a.

Quan Geng and Pramod Viswanath. The optimal noise-adding mechanism in differential privacy. *IEEE Transactions on Information Theory*, 62(2):925–951, Feb 2016b.

Quan Geng, Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The staircase mechanism in differential privacy. *IEEE Journal of Selected Topics in Signal Processing*, 9(7):1176–1184, Oct 2015.

Quan Geng, Wei Ding, Ruiqi Guo, and Sanjiv Kumar. Truncated Laplacian Mechanism for Approximate Differential Privacy. *ArXiv e-prints*, October 2018.

Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, STOC '09, pages 351–360. ACM, 2009.

Mangesh Gupte and Mukund Sundararajan. Universally optimal privacy mechanisms for minimax agents. In *Symposium on Principles of Database Systems*, pages 135–146, 2010.

Prateek Jain, Pravesh Kothari, and Abhradeep Thakurta. Differentially private online learning. In *Proceedings of the 25th Annual Conference on Learning Theory (COLT)*, 2012.

Prateek Jain, Om Dipakbhai Thakkar, and Abhradeep Thakurta. Differentially private matrix completion revisited. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.

Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275, Aug. 2017.

Mijung Park, James Foulds, Kamalika Chaudhuri, and Max Welling. DP-EM: Differentially Private Expectation Maximization. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.

Ngoc-Son Phan, Yue Wang, Xintao Wu, and Dejing Dou. Differential privacy preservation for deep autoencoders: an application of human behavior prediction. In *AAAI*, 2016.

Or Sheffet. Locally private hypothesis testing. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.

Reza Shokri and Vitaly Shmatikov. Privacy-preserving deep learning. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, pages 1310–1321. ACM, 2015.

Jordi Soria-Comas and Josep Domingo-Ferrer. Optimal data-independent noise for differential privacy. *Information Sciences*, 250:200 – 214, 2013.