The Optimal Noise-Adding Mechanism in Differential Privacy

Quan Geng and Pramod Viswanath, *Fellow, IEEE*

Abstract—Differential privacy is a framework to quantify to what extent individual privacy in a statistical database is preserved while releasing useful aggregate information about the database. In this paper, within the classes of mechanisms oblivious of the database and the queries beyond the global sensitivity, we characterize the fundamental tradeoff between privacy and utility in differential privacy, and derive the optimal ϵ-differentially private mechanism for a single real-valued query function under a very general utility-maximization (or cost-minimization) framework. The class of noise probability distributions in the optimal mechanism has staircase-shaped probability density functions which are symmetric (around the origin), monotonically decreasing and geometrically decaying. The staircase mechanism can be viewed as a geometric mixture of uniform probability distributions, providing a simple algorithmic description for the mechanism. Furthermore, the staircase mechanism naturally generalizes to discrete query output settings as well as more abstract settings. We explicitly derive the parameter of the optimal staircase mechanism for ℓ_1 and ℓ_2 cost functions. Comparing the optimal performances with those of the usual Laplacian mechanism, we show that in the high privacy regime (ϵ is small), the Laplacian mechanism is asymptotically optimal as $\epsilon \to 0$; in the low privacy regime (ϵ is large), the minimum magnitude and second moment of noise are $\Theta(\Delta e^{(-\epsilon/2)})$ and $\Theta(\Delta^2 e^{(-2\epsilon/3)})$ as $\epsilon \to +\infty$, respectively, while the corresponding figures when using the Laplacian mechanism are Δ/ϵ and $2\Delta^2/\epsilon^2$, where Δ is the sensitivity of the query function. We conclude that the gains of the staircase mechanism are more pronounced in the moderate-low privacy regime.

Index Terms—Data privacy, randomized algorithm.

I. Introduction

DIFFERENTIAL privacy is a formal framework to quantify to what extent individual privacy in a statistical database is preserved while releasing useful aggregate information about the database. The key idea of differential privacy is that the presence or absence of any individual data

Q. Geng was with the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Urbana, IL 61801 USA. He is now with Google Inc., New York, NY 10011 USA (e-mail: gengquanshine@gmail.com).

P. Viswanath is with the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Urbana, IL 61801 USA (e-mail: pramodv@illinois.edu).

in the database should not affect the final released statistical information significantly, and thus it can give strong privacy guarantees against an adversary with arbitrary auxiliary information. For motivation and background of differential privacy, we refer the readers to the survey [1] by Dwork.

Since its introduction in [2] by Dwork et. al., differential privacy has spawned a large body of research in differentially private data-releasing mechanism design and performance analysis in various settings. Differential privacy is a privacy-preserving constraint imposed on the query output releasing mechanisms, and to make use of the released information, it is important to understand the fundamental tradeoff between utility(accuracy) and privacy.

In many existing works on studying the tradeoff between accuracy and privacy in differential privacy, the usual metric of accuracy is in terms of the variance, or magnitude expectation of the noise added to the query output. For example, Hardt and Talwar [3] study the tradeoff between privacy and error for answering a set of linear queries over a histogram in a differentially private way, where the error is defined as the worst expectation of the ℓ^2-norm of the noise among all possible query output. [3] derives lower and upper bounds on the error given the differential privacy constraint. Nikolov *et al.* [4] extend the result on the tradeoff between privacy and error to the case of (ϵ, δ)-differential privacy. Li *et al.* [5] study how to optimize linear counting queries under differential privacy, where the error is measured by the mean squared error of query output estimates, which corresponds to the variance of the noise added to the query output to preserve differential privacy.

More generally, the error can be a general function depending on the additive noise (distortion) to the query output. Ghosh *et al.* [6] study a very general utility-maximization framework for a single count query with sensitivity one under differential privacy, where the utility (cost) function can be a general function depending on the noise added to the query output. [6] shows that there exists a universally optimal mechanism (adding geometric noise) to preserve differential privacy for a general class of utility functions under a Bayesian framework. Brenner and Nissim [7] show that for general query functions, no universally optimal differential privacy mechanisms exist. Gupte and Sundararajan [8] generalize the result of [6] to a minimax setting.

In this work, within the classes of mechanisms oblivious of the database and the queries beyond the global sensitivity, we study the fundamental tradeoff between utility and privacy under differential privacy, and derive the optimal differentially

private mechanism for a single real-valued query function, where the utility model is a min-max function which is the same as the one adopted in [8] and similar to the Bayesian framework used in [6], and the real-valued query function can have arbitrary sensitivity. Our results can be viewed as a generalization of [6] and [8] to general real-valued query functions with arbitrary sensitivity in a min-max utility framework. We discuss the relations of our work and the existing works in detail in Section I-D. An early version of this work has been presented in part in the conference paper [9].

A. Background on Differential Privacy

The basic problem setting in differential privacy for statistical database is as follows: suppose a dataset curator is in charge of a statistical database which consists of records of many individuals, and an analyst sends a query request to the curator to get some aggregate information about the whole database. Without any privacy concerns, the database curator can simply apply the query function to the dataset, compute the query output, and send the result to the analyst. However, to protect the privacy of individual data in the dataset, the dataset curator should use a randomized query-answering mechanism such that the probability distribution of the query output does not differ too much whether any individual record is in the database or not.

Formally, consider a real-valued query function

$$q : \mathcal{D} \to \mathbb{R}, \tag{1}$$

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 [1]. A randomized query-answering mechanism $\mathcal{K}$ for the query function q will randomly output a number with probability distribution depends on the query output $q(D)$, where D is the dataset.

Definition 1 (ϵ-Differential Privacy [1]): 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 all $S \subset Range(\mathcal{K})$,

$$Pr[\mathcal{K}(D_1) \in S] \leq \exp(\epsilon) \, Pr[\mathcal{K}(D_2) \in S], \tag{2}$$

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 (2) 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.

1) Operational Meaning of ϵ-Differential Privacy in the Context of Hypothesis Testing: We first give an operational interpretation of differential privacy in the context of hypothesis testing. While this interpretation is not directly used for proving the results in this paper, it is a useful tool for building the intuition and is useful in other contexts [10].

As shown by [11], one can interpret the differential privacy constraint (2) in the context of hypothesis testing in terms of false alarm probability and missing detection probability. Indeed, consider a binary hypothesis testing problem over two neighboring datasets, $H_0 : D_1$ versus $H_1 : D_2$, where an individual's record is in D_2 only. Given a decision rule, let S be the decision region such that when the released output lies in S, H_1 will be rejected, and when the released output lies in S^C (the complement of S), H_0 will be rejected. The false alarm probability P_{FA} and the missing detection probability P_{MD} can be written as

$$P_{FA} = P(\mathcal{K}(D_1) \in S^C), \tag{3}$$
$$P_{MD} = P(\mathcal{K}(D_2) \in S). \tag{4}$$

Therefore, from (2) we get

$$1 - P_{FA} \leq e^\epsilon P_{MD}. \tag{5}$$

Thus

$$e^\epsilon P_{MD} + P_{FA} \geq 1. \tag{6}$$

Switch D_1 and D_2 in (2), and we get

$$Pr[\mathcal{K}(D_2) \in S] \leq \exp(\epsilon) \, Pr[\mathcal{K}(D_1) \in S]. \tag{7}$$

Therefore,

$$1 - P_{MD} \leq e^\epsilon P_{FA}, \tag{8}$$

and thus

$$P_{MD} + e^\epsilon P_{FA} \geq 1. \tag{9}$$

In conclusion, we have

$$e^\epsilon P_{MD} + P_{FA} \geq 1, \tag{10}$$
$$P_{MD} + e^\epsilon P_{FA} \geq 1. \tag{11}$$

The ϵ-differential privacy constraint implies that in the context of hypothesis testing, P_{FA} and P_{MD} can not be both too small.

2) Laplacian Mechanism: The standard approach to preserving ϵ-differential privacy is to perturb the query output by adding 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 [1]): For a real-valued query function $q : \mathcal{D} \to \mathbb{R}$, the sensitivity of q is defined as

$$\Delta := \max_{D_1, D_2 \in \mathcal{D}} |q(D_1) - q(D_2)|, \tag{12}$$

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

Formally, the Laplacian mechanism is:

Definition 3 (Laplacian Mechanism [2]): For a real-valued query function $q : \mathcal{D} \rightarrow \mathbb{R}$ with sensitivity Δ, Laplacian mechanism will output

$$\mathcal{K}(D) := q(D) + Lap(\frac{\Delta}{\epsilon}), \qquad (13)$$

where $Lap(\lambda)$ is a random variable with probability density function

$$f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}, \quad \forall x \in \mathbb{R}. \qquad (14)$$

Consider two neighboring datasets D_1 and D_2 where $|q(D_1) - q(D_2)| = \Delta$. It is easy to compute the tradeoff between the false alarm probability P_{FA} and the missing detection probability P_{MD} under Laplacian mechanism, which is

$$P_{MD} = \begin{cases} 1 - e^\epsilon P_{FA} & P_{FA} \in [0, \frac{1}{2}e^{-\epsilon}) \\ \frac{e^{-\epsilon}}{4P_{FA}} & P_{FA} \in [\frac{1}{2}e^{-\epsilon}, \frac{1}{2}) \\ e^{-\epsilon}(1 - P_{FA}) & P_{FA} \in [\frac{1}{2}, 1] \end{cases} \qquad (15)$$

Since its introduction in [2], the Laplacian mechanism has become the standard tool in differential privacy and has been used as the basic building block in a number of works on differential privacy analysis in other more complex problem settings, e.g., [5], [11]–[47]. Given this near-routine use of the query-output independent adding of Laplacian noise, the following two questions are natural:

- Is query-output independent perturbation optimal?
- Assuming query-output independent perturbation, is Lapacian noise distribution optimal?

In this work we answer the above two questions. Our main result is that given an ϵ-differential privacy constraint, within the classes of mechanisms oblivious of the database and the queries beyond the global sensitivity, under a general utility-maximization (equivalently, cost-minimization) model:

- adding query-output independent noise is indeed optimal (under a mild technical condition),[1]
- the optimal noise distribution is *not* the Laplacian distribution; instead, the optimal one has a *staircase-shaped* probability density function.

These results are derived under the following settings:

- either the mechanism is oblivious of the database, or the domain of the query output is the entire real line or the set of integers;
- nothing more about the query function is known beyond its global sensitivity;

If any of these conditions are violated (the output domain has sharp boundaries, or the local sensitivity deviates from the global sensitivity [48], or we are restricted to specific query functions [23]), then the optimal privacy mechanism need not be data or query-output dependent.

B. Problem Formulation

We formulate a utility-maximization (cost-minimization) problem under the differential privacy constraint.

[1] We refer the readers to Theorem 3 for the exact technical condition. We emphasize that this technical condition is distinct from the settings and assumptions below, and we believe that this technical condition can be done away with.

1) Differential Privacy Constraint: A general randomized releasing mechanism $\mathcal{K}$ is a family of noise probability distributions indexed by the query output (denoted by t), i.e.,

$$\mathcal{K} = \{\mathcal{P}_t : t \in \mathbb{R}\}, \qquad (16)$$

and given dataset D, the mechanism $\mathcal{K}$ will release the query output $t = q(D)$ corrupted by additive random noise with probability distribution $\mathcal{P}_t$:

$$\mathcal{K}(D) = t + X_t, \qquad (17)$$

where X_t is a random variable with probability distribution $\mathcal{P}_t$.

In this work we consider the classes of mechanisms which are oblivious of the database and the queries beyond the global sensitivity. To satisfy the differential privacy constraint (2), given any $t_1 \in \mathbb{R}$, the mechanism has to consider the case there might exist a dataset D_1 such that $q(D_1) = t_1$, and for any t_2 such that $|t_1 - t_2| \leq \Delta$, where Δ is the global sensitivity of $q(\cdot)$, there might exist a neighboring dataset D_2 such that $q(D_2) = t_2$.

Therefore, for any $t_1, t_2 \in \mathbb{R}$ such that $|t_1 - t_2| \leq \Delta$ (corresponding to the query outputs for two neighboring datasets), we have

$$\mathcal{P}_{t_1}(S) \leq e^\epsilon \mathcal{P}_{t_2}(S + t_1 - t_2), \quad \forall \text{ measurable set } S \subset \mathbb{R}, \quad (18)$$

where for any $t \in \mathbb{R}$, $S + t := \{s + t | s \in S\}$.

2) Utility Model: The utility model we use in this work is under a min-max framework, which is also used in the works Gupte and Sundararajan [8], and Brenner and Nissim [7], while Ghosh *et al.* [6] uses a Bayesian utility model.

Consider a cost function $\mathcal{L}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$, which is a function of the additive noise. Given additive noise x, the cost is $\mathcal{L}(x)$. Given query output $t \in \mathbb{R}$, the additive noise is a random variable with probability distribution $\mathcal{P}_t$, and thus the expectation of the cost is

$$\int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_t(dx). \qquad (19)$$

The objective is to minimize the worst case cost among all possible query output $t \in \mathbb{R}$, i.e.,

$$\underset{t \in \mathbb{R}}{\text{minimize}} \sup \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_t(dx). \qquad (20)$$

3) Optimization Problem: Combining the differential privacy constraint (18) and the objective function (20), we formulate a functional optimization problem:

$$\underset{\{\mathcal{P}_t\}_{t \in \mathbb{R}}}{\text{minimize}} \underset{t \in \mathbb{R}}{\sup} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_t(dx) \qquad (21)$$

$$\text{s.t.} \ \forall \text{ measurable set } S \subseteq \mathbb{R}, \quad \forall |t_1 - t_2| \leq \Delta,$$

$$\mathcal{P}_{t_1}(S) \leq e^\epsilon \mathcal{P}_{t_2}(S + t_1 - t_2). \qquad (22)$$

C. An Overview of Our Results

1) Optimal Noise Probability Distribution: When the query output domain is the real line or the set of integers, we show (subject to some mild technical conditions on the family of differentially private mechanisms) that adding query-output-independent noise is optimal. Thus we only need to study what
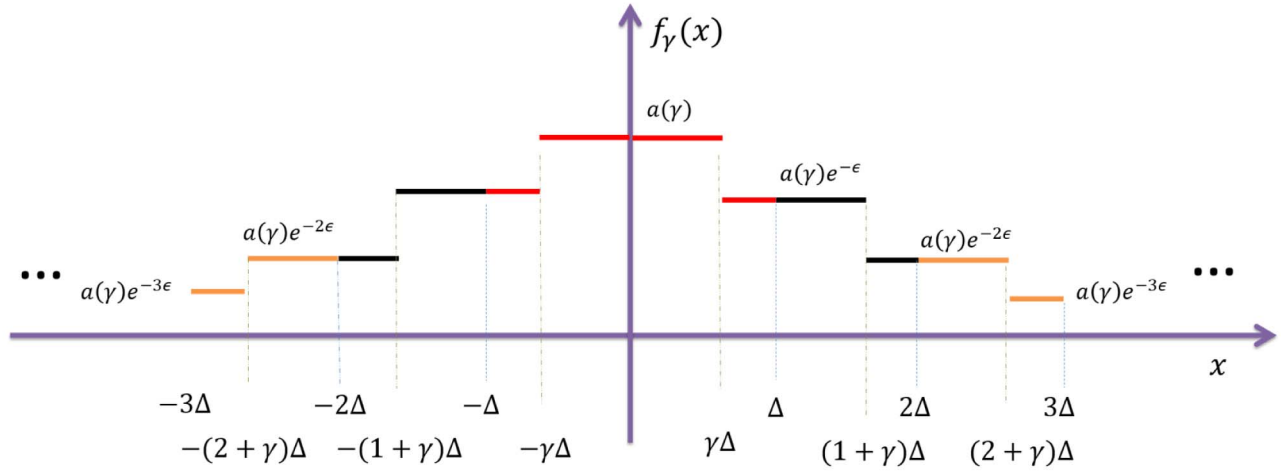
Fig. 1.   The Staircase-Shaped Probability Density Function $f_\gamma(x)$.

the optimal noise probability distribution is. Let $\mathcal{P}$ denote the probability distribution of the noise added to the query output. Then the optimization problem (21) and (22) is reduced to

$$\underset{\mathcal{P}}{\text{minimize}} \quad \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) \tag{23}$$

$$\text{s.t. } \forall \text{ measurable set } S \subseteq \mathbb{R}, \quad \forall |d| \leq \Delta,$$

$$\mathcal{P}(S) \leq e^\epsilon \mathcal{P}(S + d). \tag{24}$$

Consider a staircase-shaped probability distribution with probability density function (p.d.f.) $f_\gamma(\cdot)$ defined as

$$f_\gamma(x) = \begin{cases} a(\gamma) & x \in [0, \gamma\,\Delta) \\ e^{-\epsilon} a(\gamma) & x \in [\gamma\,\Delta, \Delta) \\ e^{-k\epsilon} f_\gamma(x - k\Delta) & x \in [k\Delta, (k+1)\Delta) \text{ for } k \in \mathbb{N} \\ f_\gamma(-x) & x < 0 \end{cases} \tag{25}$$

where

$$a(\gamma) \triangleq \frac{1 - e^{-\epsilon}}{2\Delta(\gamma + e^{-\epsilon}(1 - \gamma))} \tag{26}$$

is a normalizing constant to make $\int_{x \in \mathbb{R}} f_\gamma(x)dx = 1$.

Our main result is

Theorem 1: If the cost function $\mathcal{L}(\cdot)$ is symmetric and increasing, and $\sup_{x \geq T} \frac{\mathcal{L}(x+1)}{\mathcal{L}(x)} < +\infty$ for some $T > 0$, the optimal noise probability distribution has a staircase-shaped probability density function $f_{\gamma^*}(\cdot)$, where

$$\gamma^* = \underset{\gamma \in [0,1]}{\arg\min} \int_{x \in \mathbb{R}} \mathcal{L}(x) f_\gamma(x)dx. \tag{27}$$

Corollary 2: If the cost function $\mathcal{L}(\cdot)$ is symmetric and increasing, and $\sup_{x \geq T} \frac{\mathcal{L}(x+1)}{\mathcal{L}(x)} < +\infty$ for some $T > 0$, among the classes of ϵ-differentially private mechanisms which are oblivious of the database and the queries beyond the global sensitivity Δ, the optimal query-output independent noise-adding mechanism (minimizing the worse-case cost) is to add a noise with staircase-shaped probability density function $f_{\gamma^*}(\cdot)$, where γ^* is defined in (27).
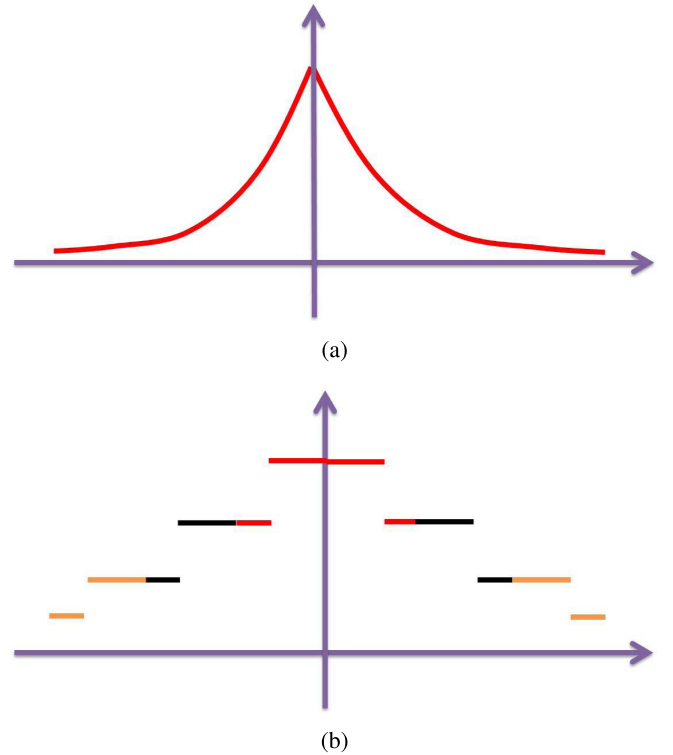

Fig. 2.   Probability Density Functions of Laplacian Mechanism and Staircase Mechanism. (a) Laplace Mechanism. (b) Staircase Mechanism.

We plot the probability density functions of Laplace mechanism and staircase mechanism in Figure 2. Figure 1 in Section III gives a precise description of the staircase mechanism.

The staircase mechanism is specified by three parameters: ϵ, Δ, and γ^* which is determined by ϵ and the cost function $\mathcal{L}(\cdot)$. For certain classes of cost functions, there are closed-form expressions for the optimal γ^*.

2) Applications: Minimum Noise Magnitude and Noise Power: We apply our main result Theorem 4 to two typical cost functions $\mathcal{L}(x) = |x|$ and $\mathcal{L}(x) = x^2$, which measure noise magnitude and noise power, respectively. We derive the

closed-form expressions for the optimal parameters γ^* for these two cost functions. Comparing the optimal performances with those of the Laplacian mechanism, we show that in the high privacy regime (ϵ is small), the Laplacian mechanism is asymptotically optimal as $\epsilon \to 0$; in the low privacy regime (ϵ is large), the minimum expectation of noise amplitude and minimum noise power are $\Theta(\Delta e^{-\frac{\epsilon}{2}})$ and $\Theta(\Delta^2 e^{-\frac{2\epsilon}{3}})$ as $\epsilon \to +\infty$, while the expectation of noise amplitude and power using the Laplacian mechanism are $\frac{\Delta}{\epsilon}$ and $\frac{2\Delta^2}{\epsilon^2}$, respectively, where Δ is the sensitivity of the query function. We conclude that the gains are more pronounced in the low privacy regime.

In certain applications we need to consider the optimal noise distribution for differential privacy with large value of ϵ. For instance, in [49] a large ϵ-DP algorithm is running on a random subsample from a dataset results in an effective value of ϵ which is much lower. In these settings, the staircase mechanism will improve on the Laplacian mechanism significantly.

3) Extension to the Discrete Setting: Since for many important practical applications query functions are integer-valued, we also derive the optimal differentially private mechanisms for answering a single integer-valued query function. We show that adding query-output independent noise is optimal under a mild technical condition, and the optimal noise probability distribution has a staircase-shaped probability mass function, which can be viewed as the discrete variant of the staircase mechanism in the continuous setting.

This result helps us directly compare our work and the existing works [6], [8] on integer-valued query functions. Our result shows that for integer-valued query function, the optimal noise probability mass function is also staircase-shaped, and in the case the sensitivity $\Delta = 1$, the optimal probability mass function is reduced to the geometric distribution, which was derived in [6] and [8]. Therefore, this result can be viewed as a generalization of [6] and [8] in the discrete setting for query functions with arbitrary sensitivity.

D. Connection to the Literature

In this section, we discuss the relations of our results and some directly related works in the literature, and the implications of our results on other works.

1) Laplacian Mechanism vs Staircase Mechanism: The Laplacian mechanism is specified by two parameters, ϵ and the query function sensitivity Δ. ϵ and Δ completely characterize the differential privacy constraint. On the other hand, the staircase mechanism is specified by three parameters, ϵ, Δ, and γ^* which is determined by ϵ and the utility function/cost function. For certain classes of utility functions/cost functions, there are closed-form expressions for the optimal γ^*.

From the two examples given in Section IV, we can see that although the Laplacian mechanism is not strictly optimal, in the high privacy regime ($\epsilon \to 0$), Laplacian mechanism is asymptotically optimal:

- For the expectation of noise amplitude, the additive gap from the optimal values goes to 0 as $\epsilon \to 0$,
- For noise power, the additive gap from the optimal values is upper bounded by a constant as $\epsilon \to 0$.

However, in the low privacy regime ($\epsilon \to +\infty$), the multiplicative gap from the optimal values can be arbitrarily large.
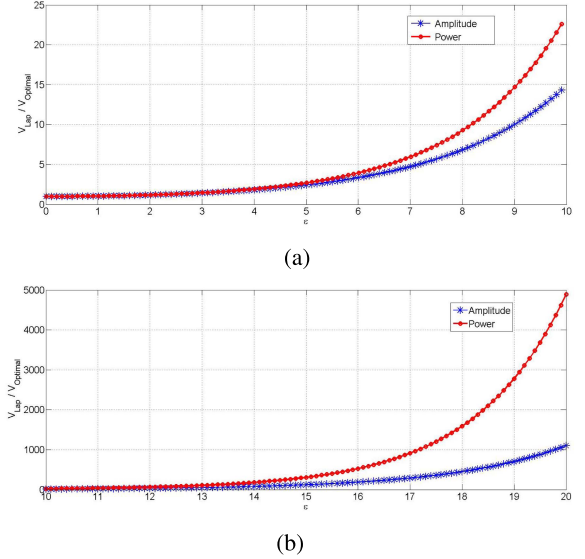

(a)

(b)

Fig. 3. Multiplicative Gain of the Staircase Mechanism over the Laplacian Mechanism. (a) $0 < \epsilon \le 10$. (b) $10 \le \epsilon \le 20$.

We conclude that in the high privacy regime, the Laplacian mechanism is nearly optimal, while in the low privacy regime significant improvement can be achieved by using the staircase mechanism. We plot the multiplicative gain of staircase mechanism over Laplacian mechanism for expectation of noise amplitude and noise power in Figure 3, where V_{Optimal} is the optimal (minimum) cost, which is achieved by staircase mechanism, and V_{Lap} is the cost of Laplacian mechanism. We can see that for $\epsilon \approx 10$, the staircase mechanism has about 15-fold and 23-fold improvement, with noise amplitude and power respectively. While $\epsilon \approx 10$ corresponds to really low privacy, our results show that low privacy can be had very cheaply (particularly when compared to the state of the art Laplacian mechanism).

Since the staircase mechanism is derived under the same problem setting as Laplacian mechanism, the staircase mechanism can be applied *wherever* Laplacian mechanism is used, and it performs *strictly better* than Laplacian mechanism (and significantly better in low privacy scenarios).

2) Relation to Shamai and Verdu, [50]: Shamai and Verdu [50] consider the minimum variance noise for a fixed value of the average of false alarm and missed detection probabilities of binary hypothesis testing. In [50], the binary hypotheses correspond to the signal being in a binary set $\{-\Delta, +\Delta\}$. Their solution involved the noise being discrete and, further, having a pmf on the integer lattice (scaled by Δ). Our setting is related, but is differentiated via the following two key distinctions:

- Instead of a constraint on the sum of false alarm and missed detection probabilities, we have constraints on symmetric weighted combinations of the two error probabilities (as in Equations (10) and (11)).
- Instead of the binary hypotheses corresponding to the signal being in a binary set $\{-\Delta, +\Delta\}$ we consider all possible binary hypotheses for the signal to be in $\{x_1, x_2\}$ where $x_1, x_2 \in [-\Delta, \Delta]$ are arbitrtary.

3) Relation to Ghosh et al. [6]: Ghosh *et al.* [6] show that for a single count query with sensitivity $\Delta = 1$, for a general class of utility functions, to minimize the expected cost under a Bayesian framework the optimal mechanism to preserve differential privacy is the geometric mechanism, which adds noise with geometric distribution.

We discuss the relations and differences between [6] and our work in the following: Both [6] and our work are similar in that, given the query output, the cost function only depends on the additive noise magnitude, and is an increasing function of noise magnitude. On the other hand, there are two main differences:

- [6] works under a general Bayesian setting where an analyst holds an arbitrary prior belief about the query value and the cost function can depend on this true value, allows for arbitrary post-processing of the released query answer, and the goal is to minimize the average cost under the analyst's prior belief about the true query value. In our work, we do not impose priors on the true query output, and our goal is to minimize the worst case cost.
- [6] studies a count query where the query output is integer-valued, bounded and sensitivity is unity. In our work, we first study general real-valued query function where the query output can take any real value, and then generalize the result to discrete setting where query output is integer valued. In both cases, the sensitivity of query functions can be arbitrary, not restricted to one.

4) Relation to Gupte and Sundararajan [8]: Gupte and Sundararajan [8] derive the optimal noise probability distributions for a single count query with sensitivity $\Delta = 1$ for minimax (risk-averse) users. Their model is the same as the one in [6] except that their objective function is to minimize the worst case cost, the same as our objective. [8] shows that although there is no universally optimal solution to the minimax optimization problem in [8] for a general class of cost functions, each solution (corresponding to different cost functions) can be derived from the same geometric mechanism by randomly remapping.

As in [6] and [8] assumes the query-output is bounded. Our result shows that when the query sensitivity is one, without any boundedness knowledge about the query-output, the optimal mechanism is to add random noise with geometric distribution to the query output.

5) Relation to Brenner and Nissim [7]: While [6] shows that for a single count query with sensitivity $\Delta = 1$, there is a universally optimal mechanism for a general class of utility functions under a Bayesian framework, Brenner and Nissim [7] show that for general query functions no universally optimal mechanisms exist. Indeed, this follows directly from our results: under our optimization framework, the optimal mechanism is adding noise with staircase-shaped probability distribution which is specified by three parameters ϵ, Δ and γ^*, where in general γ^* depends on the cost function. Generally, for different cost functions, the optimal noise probability distributions have staircase-shaped probability density functions specified by different parameters. While [7] allows post-processing (e.g., random mapping of the released query output) and we do not

explicitly consider post-processing in this work, we believe that post-processing does not help in terms of improving the utility in our setting.

6) Relation to Nissim et al. [48]: Nissim *et al.* [48] show that for certain nonlinear query functions, one can improve the accuracy by adding data-dependent noise calibrated to the smooth sensitivity of the query function, which is based on the local sensitivity of the query function. In our model, we use the global sensitivity of the query function only, and assume that the local sensitivity is the same as the global sensitivity, which holds for a general class of query functions, e.g., count, sum.

7) Relation to Hardt and Talwar [3]: Hardt and Talwar [3] study the tradeoff between privacy and error for answering a set of linear queries over a histogram in a differentially private way. The error is defined as the worst expectation of the ℓ^2-norm of the noise. The lower bound given in [3] is $\Omega(\epsilon^{-1} d \sqrt{d})$, where d is the number of linear queries. An immediate consequence of our result is that for fixed d, when $\epsilon \to +\infty$, an upper bound of $\Theta(e^{-\frac{\epsilon}{3d}} d \sqrt{d})$ is achievable by adding independent staircase-shaped noise with parameter $\frac{\epsilon}{d}$ to each component.

8) Relation to Other Works: There are many existing works on studying how to improve the accuracy for answering more complex queries under differential privacy, in which the basic building block is the standard Laplacian mechanism. For example, Hay *et al.* [51] show that one can improve the accuracy for a general class of histogram queries, by exploiting the consistency constraints on the query output, and Li *et al.* [5] study how to optimize linear counting queries under differential privacy by carefully choosing the set of linear queries to be answered. In these works, the error is measured by the mean squared error of query output estimates, which corresponds to the variance of the noise added to the query output to preserve differential privacy. In terms of ϵ, the error bound in these works scales linearly to $\frac{1}{\epsilon^2}$, because of the use of the Laplacian noise. If the Laplacian distribution is replaced by staircase distribution in these works, one can improve the error bound to $\Theta(e^{-C\epsilon})$ (for some constant C which depends on the number of queries) when $\epsilon \to +\infty$ (corresponding to the low privacy regime).

E. Organization

The paper is organized as follows. We show the optimality of query-output independent perturbation in Section II, and present the optimal differentially private mechanism, staircase mechanism, in Section III. In Section IV, we apply our main result to derive the optimal noise probability distribution with minimum expectation of noise amplitude and power, respectively, and compare the performances with the Laplacian mechanism. Section V presents the asymptotic properties of γ^* in the staircase mechanism for moment cost functions, and suggests a heuristic choice of γ that appears to work well for a wide class of cost functions. Section VI generalizes the staircase mechanism for integer-valued query function in the discrete setting, and Section VII extends the staircase mechanism to the abstract setting. Section VIII concludes this paper.

II. OPTIMALITY OF QUERY-OUTPUT INDEPENDENT PERTURBATION

Recall that the optimization problem we study in this work is

$$\underset{\{\mathcal{P}_t\}_{t\in\mathbb{R}}}{\text{minimize}} \ \sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t(dx) \tag{28}$$

$$\text{s.t. } \forall \text{ measurable set } S \subseteq \mathbb{R}, \quad \forall |t_1 - t_2| \le \Delta,$$

$$\mathcal{P}_{t_1}(S) \le e^\epsilon \mathcal{P}_{t_2}(S + t_1 - t_2). \tag{29}$$

where $\mathcal{P}_t$ is the noise probability distribution when the query output is t.

Our claim is that in the optimal family of probability distributions, $\mathcal{P}_t$ can be independent of t, i.e., the probability distribution of noise is independent of the query output. We prove this claim under a technical condition which assumes that the family of probability distributions $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ indexed by the query output t is "smooth" with respect with the index t. Specifically, we restrict the family of probability measures to be piecewise constant (with arbitrarily small width) and periodic (the period can be arbitrary) in terms of the *index parameter t*.

For any positive integer n, and for any positive real number T, define

$$\mathcal{K}_{T,n} \triangleq \{\{\mathcal{P}_t\}_{t\in\mathbb{R}} \mid \{\mathcal{P}_t\}_{t\in\mathbb{R}} \text{ satisfies } (22),$$

$$\mathcal{P}_t = \mathcal{P}_{k\frac{T}{n}}, \text{ for } t \in [k\frac{T}{n}, (k+1)\frac{T}{n}), \quad k \in \mathbb{Z},$$

$$\text{and } \mathcal{P}_{t+T} = \mathcal{P}_t, \forall t \in \mathbb{R}\}. \tag{30}$$

Theorem 3: Given any family of probability distribution $\{\mathcal{P}_t\}_{t\in\mathbb{R}} \in \cup_{T>0} \cup_{n\ge1} \mathcal{K}_{T,n}$, there exists a probability distribution $\mathcal{P}^*$ such that the family of probability distributions $\{\mathcal{P}_t^*\}_{t\in\mathbb{R}}$ with $\mathcal{P}_t^* \equiv \mathcal{P}^*$ satisfies the differential privacy constraint (22) and

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t^*(dx) \le \sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t(dx). \tag{31}$$

Proof: Here we briefly discuss the main proof technique. For complete proof, see Appendix A. The proof of Theorem 3 uses two properties on the family of probability distributions satisfying differential privacy constraint (22). First, we show that for any family of probability distributions satisfying (22), any translation of the probability distributions will also preserve differential privacy, and the cost is the same. Second, we show that given a collection of families of probability distributions each of which satisfies (22), we can take a convex combination of them to construct a new family of probability distributions satisfying (22) and the new cost is not worse. Due to these two properties, given any family of probability distributions $\{\mathcal{P}_t\}_{t\in\mathbb{R}} \in \cup_{T>0}\cup_{n\ge1}\mathcal{K}_{T,n}$, one can take a convex combination of different translations of $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ to construct $\{\mathcal{P}_t^*\}_{t\in\mathbb{R}}$ with $\mathcal{P}_t^* \equiv \mathcal{P}^*$, and the cost is not worse. ∎

Here we assume the family of noise probability distributions is piecewise constant in terms of the index t (over arbitrarily small intervals with length $\frac{T}{n}$), and periodic over t (with arbitrarily large period T). These technical conditions allow our proofs to be mathematically simple, allowing for finite averaging techniques. We conjecture that the technical conditions can be done away with.

III. OPTIMAL NOISE PROBABILITY DISTRIBUTION

Due to Theorem 3, to derive the optimal randomized mechanism to preserve differential privacy, we can restrict to noise-adding mechanisms where the noise probability distribution does not depend on the query output. In this section we state our main result Theorem 4 on the optimal noise probability distribution.

Let $\mathcal{P}$ denote the probability distribution of the noise added to the query output. Then the optimization problem in (21) and (22) is reduced to

$$\underset{\mathcal{P}}{\text{minimize}} \ \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) \tag{32}$$

$$\text{s.t. } \forall \text{ measurable set } S \subseteq \mathbb{R}, \quad \forall |d| \le \Delta,$$

$$\mathcal{P}(S) \le e^\epsilon \mathcal{P}(S + d). \tag{33}$$

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

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

$$\mathcal{L}(x) = \mathcal{L}(-x), \quad \forall x \in \mathbb{R}, \tag{34}$$

and

$$\mathcal{L}(x) \le \mathcal{L}(y), \quad \forall 0 \le x \le y. \tag{35}$$

In addition, we assume $\mathcal{L}(x)$ satisfies a mild technical condition which essentially says that $\mathcal{L}(\cdot)$ does not increase too fast (while still allowing it to be unbounded).

Property 2: There exists a positive integer T such that $\mathcal{L}(T) > 0$ and $\mathcal{L}(x)$ satisfies

$$\sup_{x\ge T} \frac{\mathcal{L}(x+1)}{\mathcal{L}(x)} < +\infty. \tag{36}$$

Consider a staircase-shaped probability distribution with probability density function (p.d.f.) $f_\gamma(\cdot)$ defined as

$$f_\gamma(x) = \begin{cases} a(\gamma) & x \in [0, \gamma\Delta) \\ e^{-\epsilon}a(\gamma) & x \in [\gamma\Delta, \Delta) \\ e^{-k\epsilon}f_\gamma(x - k\Delta) & x \in [k\Delta, (k+1)\Delta) \text{ for } k \in \mathbb{N} \\ f_\gamma(-x) & x < 0 \end{cases} \tag{37}$$

where

$$a(\gamma) \triangleq \frac{1 - e^{-\epsilon}}{2\Delta(\gamma + e^{-\epsilon}(1 - \gamma))} \tag{38}$$

is a normalizing constant to make $\int_{x\in\mathbb{R}} f_\gamma(x)dx = 1$. It is easy to check that for any $\gamma \in [0, 1]$, the probability distribution with p.d.f. $f_\gamma(\cdot)$ satisfies the differential privacy constraint (33). Indeed, the probability density function $f_\gamma(x)$ satisfies

$$f_\gamma(x) \le e^\epsilon f_\gamma(x + d), \quad \forall x \in \mathbb{R}, |d| \le \Delta, \tag{39}$$

which implies (33).

Let $\mathcal{SP}$ denote the set of all probability distributions satisfying (33). Our main result on the optimal noise probability distribution is:

Algorithm 1 Generation of Random Variable With Staircase Distribution

> **Input:** ϵ, Δ, and $\gamma \in [0, 1]$.
> **Output:** X, a random variable (r.v.) with staircase distribution specified by ϵ, Δ and γ.
> Generate a r.v. S with $\Pr[S = 1] = \Pr[S = -1] = \frac{1}{2}$.
> Generate a geometric r.v. G with $\Pr[G = i] = (1 - b)b^i$ for integer $i \geq 0$, where $b = e^{-\epsilon}$.
> Generate a r.v. U uniformly distributed in $[0, 1]$.
> Generate a binary r.v. B with $\Pr[B = 0] = \frac{\gamma}{\gamma + (1-\gamma)b}$ and $\Pr[B = 1] = \frac{(1-\gamma)b}{\gamma + (1-\gamma)b}$.
> $X \leftarrow S((1 - B)((G + \gamma U)\Delta) + B((G + \gamma + (1 - \gamma)U)\Delta))$.
> Output X.

Fig. 4.   Optimal γ^* for cost function $L(x) = x^2$.

Theorem 4: If the cost function $\mathcal{L}(x)$ satisfies Property 1 and Property 2, then

$$\inf_{\mathcal{P} \in \mathcal{SP}} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) = \inf_{\gamma \in [0,1]} \int_{x \in \mathbb{R}} \mathcal{L}(x) f_\gamma(x) dx. \quad (40)$$

Proof: Here we briefly discuss the main proof idea and technique. First, by deriving several properties on the probability distributions satisfying the ϵ-differential privacy constraint, we show that without loss of generality, one can "discretize" any valid probability distribution, even for those which do not have probability density functions. Second, we show that to minimize the cost, the probability density function of the discretized probability distribution should be monotonically and geometrically decaying. Lastly, we show that the optimal probability density function should be staircase-shaped. For the complete proof, see Appendix B. ∎

Therefore, the optimal noise probability distribution to preserve ϵ-differential privacy for any real-valued query function has a staircase-shaped probability density function, which is specified by three parameters ϵ, Δ and $\gamma^* = \arg\min_{\gamma \in [0,1]} \int_{x \in \mathbb{R}} \mathcal{L}(x) f_\gamma(x) dx$.

A natural and simple algorithm to generate random noise with staircase distribution is given in Algorithm 1.

In the formula,

$$X \leftarrow S((1 - B)((G + \gamma U)\Delta) + B((G + \gamma + (1 - \gamma)U)\Delta)), \quad (41)$$

- S determines the sign of the noise,
- G determines which interval $[G\Delta, (G + 1)\Delta)$ the noise lies in,
- B determines which subinterval of $[G\Delta, (G + \gamma)\Delta)$ and $[(G + \gamma)\Delta, (G + 1)\Delta)$ the noise lies in,
- U helps to uniformly sample the subinterval.

IV. APPLICATIONS

In this section, we apply our main result Theorem 4 to derive the parameter γ^* of the staircase mechanism with minimum expectation of noise magnitude and noise second moment, respectively, and then compare the performances with the Laplacian mechanism.
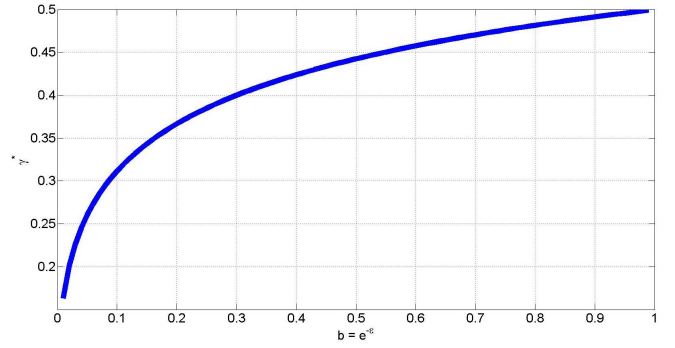
A. Optimal Noise Probability Distribution With Minimum Expectation of Noise Amplitude

To minimize the expectation of amplitude, we have cost function $\mathcal{L}(x) = |x|$, and it is easy to see that it satisfies Property 1 and Property 2.

To simplify notation, define $b \triangleq e^{-\epsilon}$, and define

$$V(\mathcal{P}) \triangleq \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx). \quad (42)$$

for a given probability distribution $\mathcal{P}$.

Theorem 5: To minimize the expectation of the amplitude of noise, the optimal noise probability distribution has probability density function $f_{\gamma^*}(\cdot)$ with

$$\gamma^* = \frac{1}{1 + e^{\frac{\epsilon}{2}}}, \quad (43)$$

and the minimum expectation of noise amplitude is

$$V(\mathcal{P}_{\gamma^*}) = \Delta \frac{e^{\frac{\epsilon}{2}}}{e^\epsilon - 1}. \quad (44)$$

Proof: See Appendix C. ∎

Next, we compare the performances of the optimal noise probability distribution and the Laplacian mechanism. The Laplace distribution has probability density function

$$f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}, \quad (45)$$

where $\lambda = \frac{\Delta}{\epsilon}$. So the expectation of the amplitude of noise with Laplace distribution is

$$V_{Lap} \triangleq \int_{-\infty}^{+\infty} |x| f(x) dx = \frac{\Delta}{\epsilon}. \quad (46)$$

By comparing $V(\mathcal{P}_{\gamma^*})$ and V_{Lap}, it is easy to see that in the high privacy regime (ϵ is small) the Laplacian mechanism is asymptotically optimal, and the additive gap from optimal value goes to 0 as $\epsilon \to 0$; in the low privacy regime (ϵ is large), $V_{Lap} = \frac{\Delta}{\epsilon}$, while $V(\mathcal{P}_{\gamma^*}) = \Theta(\Delta e^{-\frac{\epsilon}{2}})$. Indeed,

Corollary 6: Consider the cost function $\mathcal{L}(x) = |x|$. In the high privacy regime (ϵ is small),

$$V_{Lap} - V(\mathcal{P}_{\gamma^*}) = \Delta \left(\frac{\epsilon}{24} - \frac{7\epsilon^3}{5760} + O(\epsilon^5) \right), \quad (47)$$

as $\epsilon \to 0$.

And in the low privacy regime (ϵ is large),

$$V_{Lap} = \frac{\Delta}{\epsilon}, \tag{48}$$

$$V(\mathcal{P}_{\gamma*}) = \Theta(\Delta e^{-\frac{\epsilon}{2}}), \tag{49}$$

as $\epsilon \to +\infty$.

B. Optimal Noise Probability Distribution With Minimum Power

Given the probability distribution $\mathcal{P}$ of the noise, the power of noise is defined as $\int_{x \in \mathbb{R}} x^2 \mathcal{P}(dx)$. Accordingly, the cost function $\mathcal{L}(x) = x^2$, and it is easy to see it satisfies Property 1 and Property 2.

Recall $b \triangleq e^{-\epsilon}$.

Theorem 7: To minimize the power of noise (accordingly, $\mathcal{L}(x) = x^2$), the optimal noise probability distribution has probability density function $f_{\gamma*}(\cdot)$ with

$$\gamma^* = -\frac{b}{1-b} + \frac{(b - 2b^2 + 2b^4 - b^5)^{1/3}}{2^{1/3}(1-b)^2}, \tag{50}$$

and the minimum power of noise is

$$V(\mathcal{P}_{\gamma*}) = \Delta^2 \frac{2^{-2/3} b^{2/3}(1+b)^{2/3} + b}{(1-b)^2}. \tag{51}$$

Proof: See Appendix D. ∎

Next, we compare the performances of the optimal noise probability distribution and the Laplacian mechanism. The power of noise with Laplace distribution with $\lambda = \frac{\Delta}{\epsilon}$ is

$$V_{Lap} \triangleq \int_{-\infty}^{+\infty} x^2 \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} dx = 2\frac{\Delta^2}{\epsilon^2}. \tag{52}$$

By comparing $V(\mathcal{P}_{\gamma*})$ and V_{Lap}, it is easy to see that in the high privacy regime (ϵ is small) Laplacian mechanism is asymptotically optimal, and the additive gap from optimal value is upper bounded by a constant as $\epsilon \to 0$; in the low privacy regime (ϵ is large), $V_{Lap} = \Theta(\frac{2\Delta^2}{\epsilon^2})$, while $V(\mathcal{P}_{\gamma*}) = \Theta(\Delta^2 e^{-\frac{2\epsilon}{3}})$. Indeed,

Corollary 8: Consider the cost function $\mathcal{L}(x) = x^2$. In the high privacy regime (ϵ is small),

$$V_{Lap} - V(\mathcal{P}_{\gamma*}) = \Delta^2 \left(\frac{1}{12} - \frac{\epsilon^2}{720} + O(\epsilon^4) \right), \tag{53}$$

as $\epsilon \to 0$.

And in the low privacy regime (ϵ is large),

$$V_{Lap} = \frac{2\Delta^2}{\epsilon^2}, \tag{54}$$

$$V(\mathcal{P}_{\gamma*}) = \Theta(\Delta^2 e^{-\frac{2\epsilon}{3}}), \tag{55}$$

as $\epsilon \to +\infty$.

V. PROPERTY OF γ^*

In this section, we derive some asymptotic properties of the optimal γ^* for moment cost functions, and give a heuristic choice of γ which only depends on ϵ.

A. Asymptotic Properties of γ^*

In Section IV, we have seen that for the cost functions $\mathcal{L}(x) = |x|$ and $\mathcal{L}(x) = x^2$, the optimal γ^* lies in the interval $[0, \frac{1}{2}]$ for all ϵ and is a monotonically decreasing function of ϵ; and furthermore, $\gamma^* \to \frac{1}{2}$ as ϵ goes to 0, and $\gamma^* \to 0$ as ϵ goes to $+\infty$.

We generalize these asymptotic properties of γ as a function of ϵ to all moment cost functions. More precisely, given $m \in \mathbb{N}$ and $m \geq 1$,

Theorem 9: Consider the cost function $\mathcal{L}(x) = |x|^m$. Let γ^* be the optimal γ in the staircase mechanism for $\mathcal{L}(x)$, i.e.,

$$\gamma^* = \arg\min_{\gamma \in [0,1]} \int_{x \in \mathbb{R}} |x|^m f_\gamma(x) dx. \tag{56}$$

We have

$$\gamma^* \to \frac{1}{2}, \quad as \ \epsilon \to 0, \tag{57}$$

$$\gamma^* \to 0, \quad as \ \epsilon \to +\infty. \tag{58}$$

Proof: See Appendix E. ∎

Corollary 10: For all the cost functions $\mathcal{L}(\cdot)$ which can be written as

$$\mathcal{L}(x) = \sum_{i=1}^{n} \alpha_i |x|^{d_i}, \tag{59}$$

where $n \geq 1$, $\alpha_i \in \mathbb{R}$, $d_i \in \mathbb{N}$ and α_i, $d_i \geq 0$ for all i, the optimal γ^* in the staircase mechanism has the following asymptotic properties:

$$\gamma^* \to \frac{1}{2}, \quad as \ \epsilon \to 0, \tag{60}$$

$$\gamma^* \to 0, \quad as \ \epsilon \to +\infty. \tag{61}$$

B. A Heuristic Choice of γ

We have shown that in general the optimal γ^* in the staircase mechanism depends on both ϵ and the cost function $\mathcal{L}(\cdot)$. Here we give a heuristic choice of γ which depends only on ϵ, and show that the performance is reasonably good in the low privacy regime.

Consider a particular choice of γ, which is

$$\tilde{\gamma} := \frac{b}{2} = \frac{e^{-\epsilon}}{2}. \tag{62}$$

It is easy to see that $\tilde{\gamma}$ has the same asymptotic properties as the optimal γ^* for moment cost functions, i.e.,

$$\tilde{\gamma} \to 0, \quad as \ b \to 0, \tag{63}$$

$$\tilde{\gamma} \to \frac{1}{2}, \quad as \ b \to 1. \tag{64}$$

Furthermore, the probability that the noise magnitude is less than $\frac{e^{-\epsilon}}{2}\Delta$ is approximately $\frac{1}{3}$ in the low privacy regime ($\epsilon \to +\infty$). Indeed,

$$\Pr[|X| \leq \frac{e^{-\epsilon}}{2}\Delta] = \Pr[|X| \leq \tilde{\gamma}\Delta] \tag{65}$$

$$= 2a(\tilde{\gamma})\tilde{\gamma}\Delta \tag{66}$$

$$= \frac{1-b}{\tilde{\gamma} + b(1-\tilde{\gamma})}\tilde{\gamma} \tag{67}$$

$$= \frac{b-b^2}{3b-b^2}, \tag{68}$$

which goes to $\frac{1}{3}$ as $\epsilon \to +\infty$ (accordingly, $b \to 0$).

On the other hand, for Laplace mechanism,

$$\Pr[|X| \leq \frac{e^{-\epsilon}}{2}\Delta] = \int_{-\frac{e^{-\epsilon}}{2}\Delta}^{\frac{e^{-\epsilon}}{2}\Delta} \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} dx = 1 - e^{-\frac{\epsilon e^{-\epsilon}}{2}}, \quad (69)$$

which goes to zero as $\epsilon \to +\infty$.

We conclude that in the low privacy regime as $\epsilon \to +\infty$, the staircase mechanism with the heuristic parameter $\tilde{\gamma} = \frac{e^{-\epsilon}}{2}$ can guarantee with probability about $\frac{1}{3}$ the additive noise is very close to zero, while the probability given by Laplacian mechanism is approximately zero.

VI. EXTENSION TO THE DISCRETE SETTING

In this section, we extend our main result Theorem 3 and Theorem 4 to the discrete settings, and show that the optimal noise-adding mechanism in the discrete setting is a discrete variant of the staircase mechanism in the continuous setting.

A. Problem Formulation

We first give the problem formulation in the discrete setting. Consider an integer-valued query function[2]

$$q : \mathcal{D} \to \mathbb{Z}, \quad (70)$$

where $\mathcal{D}$ is the domain of the databases. Let Δ denote the sensitivity of the query function q as defined in (12). Clearly, Δ is an integer in this discrete setting.

In the discrete setting, a generic randomized mechanism $\mathcal{K}$ is a family of noise probability distributions over the domain $\mathbb{Z}$ indexed by the query output (denoted by i), i.e.,

$$\mathcal{K} = \{\mathcal{P}_i : i \in \mathbb{Z}\}, \quad (71)$$

and given dataset D, the mechanism $\mathcal{K}$ will release the query output $i = q(D)$ corrupted by additive random noise with probability distribution $\mathcal{P}_i$:

$$\mathcal{K}(D) = i + X_i, \quad (72)$$

where X_i is a discrete random variable with probability distribution $\mathcal{P}_i$.

Then, the ϵ-differential privacy constraint (2) on $\mathcal{K}$ is that for any $i_1, i_2 \in Z$ such that $|i_1 - i_2| \leq \Delta$ (corresponding to the query outputs for two neighboring datasets), and for any subset $S \subset \mathbb{Z}$,

$$\mathcal{P}_{i_1}(j) \leq e^\epsilon \mathcal{P}_{i_2}(j + i_1 - i_2), \quad \forall j \in \mathbb{Z}, \ |i_1 - i_2| \leq \Delta, \quad (73)$$

and the goal is to minimize the worst-case cost

$$\sup_{i \in \mathbb{Z}} \sum_{j=-\infty}^{+\infty} \mathcal{L}(j)\mathcal{P}_i(j) \quad (74)$$

subject to the differential privacy constraint (73).

[2]Without loss of generality, we assume that in the discrete setting the query output is integer-valued. Indeed, any uniformly-spaced discrete setting can be reduced to the integer-valued setting by scaling the query output.

B. Optimality of Query-Output Independent Perturbation

In this section, we show that query-output independent perturbation is optimal in the discrete setting.

For any integer $n \geq 1$, define

$$\mathcal{K}_n \triangleq \{\{\mathcal{P}_i\}_{i \in \mathbb{Z}} | \{\mathcal{P}_i\}_{i \in \mathbb{Z}} \text{ satisfies (73), and } \mathcal{P}_{i+n} = \mathcal{P}_i, \forall i \in \mathbb{Z}\}. \quad (75)$$

Theorem 11: Given any family of probability distribution $\{\mathcal{P}_i\}_{i \in \mathbb{Z}} \in \cup_{n \geq 1} \mathcal{K}_n$, there exists a probability distribution $\mathcal{P}^*$ such that the family of probability distributions $\{\mathcal{P}_i^*\}_{i \in \mathbb{Z}}$ with $\mathcal{P}_i^* \equiv \mathcal{P}^*$ satisfies the differential privacy constraint (73) and

$$\sup_{i \in \mathbb{Z}} \sum_{j=-\infty}^{+\infty} \mathcal{L}(j)\mathcal{P}_i^*(j) \leq \sup_{i \in \mathbb{Z}} \sum_{j=-\infty}^{+\infty} \mathcal{L}(j)\mathcal{P}_i(j). \quad (76)$$

Proof: The proof is essentially the same as the proof of Theorem 3, and thus is omitted. ∎

Theorem 11 states that if we assume the family of noise probability distributions is periodic in terms of i (the period can be arbitrary), then in the optimal mechanism we can assume $\mathcal{P}_i$ does not depend on i. We conjecture that the technical condition can be done away with.

C. Optimal Noise Probability Distribution

Due to Theorem 11, we restrict to query-output independent perturbation mechanisms.

Let $q(D)$ be the value of the query function evaluated at dataset D. The noise-adding mechanism $\mathcal{K}$ will output

$$\mathcal{K}(D) = q(D) + X, \quad (77)$$

where X is the integer-valued noise added by the mechanism to the output of query function. Let $\mathcal{P}$ be the probability distribution of the noise X. Then the optimization problem we study is

$$\underset{\mathcal{P}}{\text{minimize}} \quad \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}(i) \quad (78)$$

$$\text{subject to } \mathcal{P}(i) \leq e^\epsilon \mathcal{P}(i+d), \quad \forall i \in \mathbb{Z}, \ d \in \mathbb{Z}, \ |d| \leq |\Delta|. \quad (79)$$

It turns out that when the cost function $\mathcal{L}(\cdot)$ is symmetric and monotonically increasing for $i \geq 0$, the solution to the above optimization problem is a discrete variant of the staircase mechanism in the continuous setting.

As in the continuous setting, we also assume that the cost function $\mathcal{L}(\cdot)$ is symmetric and monotonically increasing for $x \geq 0$, i.e.,
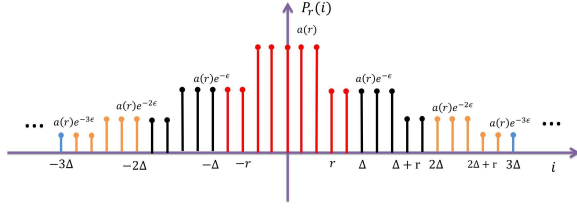
Property 3:

$$\mathcal{L}(i) = \mathcal{L}(-i), \quad \forall i \in \mathbb{Z} \quad (80)$$

$$\mathcal{L}(i) \leq \mathcal{L}(j), \quad \forall i, j \in \mathbb{Z}, \ 0 \leq i \leq j. \quad (81)$$

The easiest case is $\Delta = 1$. In the case that $\Delta = 1$, the solution is the geometric mechanism, which was proposed in [6].

Recall $b \triangleq e^{-\epsilon}$.

Theorem 12: If the cost function $\mathcal{L}(\cdot)$ satisfies Property 3 and $\Delta = 1$, then the geometric mechanism, which has a

Fig. 5.   The Staircase-Shaped Probability Mass Function $\mathcal{P}_r(i)$.

probability mass function $\mathcal{P}$ with $\mathcal{P}(i) = \frac{1-b}{1+b}b^{|i|}$, $\forall i \in \mathbb{Z}$, is the optimal solution to (78).

Proof: See Appendix F.   ■

For fixed general $\Delta \geq 2$, consider a class of symmetric and staircase-shaped probability mass functions defined as follows. Given an integer $1 \leq r \leq \Delta$, denote $\mathcal{P}_r$ as the probability mass function defined by

$$\mathcal{P}_r(i) = \begin{cases} a(r) & 0 \leq i < r \\ e^{-\epsilon}a(r) & r \leq i < \Delta \\ e^{-k\epsilon}\mathcal{P}_r(i - k\Delta) & k\Delta \leq i < (k+1)\Delta \text{ for } k \in \mathbb{N} \\ \mathcal{P}_r(-i) & i < 0 \end{cases}$$

$$(82)$$

for $i \in \mathbb{Z}$, where

$$a(r) \triangleq \frac{1-b}{2r + 2b(\Delta - r) - (1-b)}. \tag{83}$$

It is easy to verify that for any $1 \leq r \leq \Delta$, $\mathcal{P}_r$ is a valid probability mass function and it satisfies the ϵ-differential privacy constraint (79). We plot the staircase-shaped probability mass function $\mathcal{P}_r(i)$ in Figure 5.

Let $\mathcal{SP}$ be the set of all probability mass functions which satisfy the ϵ-differential privacy constraint (79).

Theorem 13: For $\Delta \geq 2$, if the cost function $\mathcal{L}(x)$ satisfies Property 3, then

$$\inf_{\mathcal{P} \in \mathcal{SP}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}(i) = \min_{\{r \in \mathbb{N} | 1 \leq r \leq \Delta\}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_r(i). \tag{84}$$

Proof: See Appendix F.   ■

Therefore, the optimal noise probability distribution to preserve ϵ-differential privacy for integer-valued query function has a staircase-shaped probability mass function, which is specified by three parameters ϵ, Δ and $r^* = \arg\min_{\{r \in \mathbb{N} | 1 \leq r \leq \Delta\}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_r(i)$. In the case $\Delta = 1$, the staircase-shaped probability mass function is reduced to the geometric mechanism.

VII. EXTENSION TO THE ABSTRACT SETTING

In this section, we show how to extend the staircase mechanism to the abstract setting. The approach is essentially the same as the exponential mechanism in [52], except that we replace the exponential function by the staircase function.

Consider a privacy mechanism which maps an input from a domain $\mathcal{D}$ to some output in a range $\mathcal{R}$. Let μ be the base measure of $\mathcal{R}$. In addition, we have a cost function $\mathcal{C} : \mathcal{D} \times \mathcal{R} \to [0, +\infty)$.

Define Δ as

$$\Delta \triangleq \max_{r \in \mathcal{R}, \, D_1, D_2 \subseteq \mathcal{D}:|D_1 - D_2| \leq 1} |\mathcal{C}(D_1, r) - \mathcal{C}(D_2, r)|, \tag{85}$$

i.e., the maximum difference of cost function for any two inputs which differ only on one single value over all $r \in \mathcal{R}$ [52].

A randomized mechanism $\mathcal{K}$ achieves ϵ-differential privacy if for any $D_1, D_2 \subseteq \mathcal{D}$ such that $|D_1 - D_2| \leq 1$, and for any measurable subset $S \subset \mathcal{R}$,

$$\Pr[\mathcal{K}(D_1) \in S] \leq \exp(\epsilon)\Pr[\mathcal{K}(D_2) \in S]. \tag{86}$$

Definition 4 (Staircase Mechanism in the Abstract Setting): For fixed $\gamma \in [0, 1]$, given input $D \in \mathcal{D}$, the staircase mechanism in the abstract setting will output an element in $\mathcal{R}$ with the probability distribution defined as

$$\mathcal{P}_D(S) = \frac{\int_{r \in S} f_\gamma(\mathcal{C}(D, r))\mu(dr)}{\int_{r \in \mathcal{R}} f_\gamma(\mathcal{C}(D, r))\mu(dr)}, \quad \forall \text{ measurable set } S \subset \mathcal{R},$$

$$(87)$$

where f_γ is the staircase-shaped function defined in (37).

Theorem 14: The staircase mechanism in the abstract setting in Definition 4 achieves 2ϵ-differential privacy.

Proof: For any $D_1, D_2 \in \mathcal{D}$ such that $|D_1 - D_2| \leq 1$, and for any measurable set $S \subset \mathcal{R}$,

$$\mathcal{P}_{D_1}(S) = \frac{\int_{r \in S} f_\gamma(\mathcal{C}(D_1, r))\mu(dr)}{\int_{r \in \mathcal{R}} f_\gamma(\mathcal{C}(D_1, r))\mu(dr)} \tag{88}$$

$$\leq e^\epsilon \frac{\int_{r \in S} f_\gamma(\mathcal{C}(D_2, r))\mu(dr)}{\int_{r \in \mathcal{R}} f_\gamma(\mathcal{C}(D_1, r))\mu(dr)} \tag{89}$$

$$\leq e^\epsilon e^\epsilon \frac{\int_{r \in S} f_\gamma(\mathcal{C}(D_2, r))\mu(dr)}{\int_{r \in \mathcal{R}} f_\gamma(\mathcal{C}(D_2, r))\mu(dr)} \tag{90}$$

$$= e^{2\epsilon}\mathcal{P}_{D_2}(S), \tag{91}$$

where we have used the property that $f_\gamma(\mathcal{C}(D_1, r)) \leq e^\epsilon f_\gamma(\mathcal{C}(D_2, r))$ and $f_\gamma(\mathcal{C}(D_2, r)) \leq e^\epsilon f_\gamma(\mathcal{C}(D_1, r))$ for all $r \in \mathcal{R}$.

Therefore, the staircase mechanism in the abstract setting achieves 2ϵ-differential privacy for any $\gamma \in [0, 1]$.   ■

In the case that the output range $\mathcal{R}$ is the set of real numbers $\mathbb{R}$ and the cost function $\mathcal{C}(d, r) = |r - q(d)|$ for some real-valued query function q, the above mechanism is reduced to the staircase mechanism in the continuous setting.

VIII. CONCLUSION

In this work we show that within the classes of mechanisms oblivious of the database and the queries beyond the global sensitivity, adding query-output independent noise with staircase distribution is optimal among all randomized mechanisms (subject to a mild technical condition) that preserve differential privacy. The optimality is for single real-valued query function under a very general utility-maximization (or cost-minimization) framework. The class of optimal noise probability distributions has *staircase-shaped* probability density functions which are symmetric (around the origin), monotonically decreasing and geometrically decaying for $x \geq 0$. The staircase mechanism can be viewed as a *geometric*

mixture of uniform probability distributions, providing a simple algorithmic description for the mechanism. Furthermore, the staircase mechanism naturally generalizes to discrete query output settings as well as more abstract settings.

We explicitly derive the parameter of the staircase mechanism with minimum expectation of noise amplitude and power. Comparing the optimal performances with those of the standard Laplacian mechanism, we show that in the high privacy regime (ϵ is small), Laplacian mechanism is asymptotically optimal as $\epsilon \to 0$; in the low privacy regime (ϵ is large), the minimum expectation of noise amplitude and minimum noise power are $\Theta(\Delta e^{-\frac{\epsilon}{2}})$ and $\Theta(\Delta^2 e^{-\frac{2\epsilon}{3}})$ as $\epsilon \to +\infty$, while the expectation of noise amplitude and power using the Laplacian mechanism are $\frac{\Delta}{\epsilon}$ and $\frac{2\Delta^2}{\epsilon^2}$, where Δ is the sensitivity of the query function. We conclude that the gains are more pronounced in the moderate to low privacy regime.

We would like to point it out that the staircase mechanism has been extended to the multidimensional setting in [53], and the optimal mechanism is also studied in [54] for approximate differential privacy under a similar optimization framework.

APPENDIX A
PROOF OF THEOREM 3

We first give two lemmas on the properties of $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ which satisfies (22).

Lemma 15: Given $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ satisfying (22), and given any scalar $\alpha \in \mathbb{R}$, consider the family of noise probability measures $\{\mathcal{P}_t^{(\alpha)}\}_{t\in\mathbb{R}}$ defined by

$$\mathcal{P}_t^{(\alpha)} \triangleq \mathcal{P}_{t+\alpha}, \quad \forall t \in \mathbb{R}. \tag{92}$$

Then $\{\mathcal{P}_t^{(\alpha)}\}_{t\in\mathbb{R}}$ also satisfies the differential privacy constraint, i.e., $\forall |t_1 - t_2| \le \Delta$,

$$\mathcal{P}_{t_1}^{(\alpha)}(S) \le e^{\epsilon} \mathcal{P}_{t_2}^{(\alpha)}(S + t_1 - t_2). \tag{93}$$

Furthermore, $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ and $\{\mathcal{P}_t^{(\alpha)}\}_{t\in\mathbb{R}}$ have the same cost, i.e.,

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t(dx) = \sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t^{(\alpha)}(dx). \tag{94}$$

Proof: Since by definition the family of probability measures $\{\mathcal{P}_t^{(\alpha)}\}_{t\in\mathbb{R}}$ is a shifted version of $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$, (94) holds.

Next we show that $\{\mathcal{P}_t^{(\alpha)}\}_{t\in\mathbb{R}}$ satisfies (93). Given any t_1, t_2 such that $|t_1 - t_2| \le \Delta$, then for any measurable set $S \subset \mathbb{R}$, we have

$$\mathcal{P}_{t_1}^{(\alpha)} = \mathcal{P}_{t_1+\alpha}(S) \tag{95}$$

$$\le e^{\epsilon} \mathcal{P}_{t_2+\alpha}(S + (t_1 + \alpha) - (t_2 + \alpha)) \tag{96}$$

$$= e^{\epsilon} \mathcal{P}_{t_2+\alpha}(S + t_1 - t_2) \tag{97}$$

$$= e^{\epsilon} \mathcal{P}_{t_2}^{(\alpha)}(S + t_1 - t_2). \tag{98}$$

This completes the proof. ∎

Next we show that given a collection of families of probability measures each of which satisfies the differential privacy constraint (22), we can take a convex combination of them to construct a new family of probability measures satisfying (22) and the new cost is not worse. More precisely,

Lemma 16: Given a collection of finite number of families of probability measures $\{\mathcal{P}_t^{[i]}\}_{t\in\mathbb{R}}$ ($i \in \{1, 2, 3, \ldots, n\}$), such that for each i, $\{\mathcal{P}_t^{[i]}\}_{t\in\mathbb{R}}$ satisfies (22) and

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t^{[i]}(dx) = Q, \quad \forall i, \tag{99}$$

for some real number Q, consider the family of probability measures $\{\tilde{v}_t\}_{t\in\mathbb{R}}$ defined by

$$\tilde{v}_t \triangleq \sum_{i=1}^{n} c_i \mathcal{P}_t^{[i]}, \quad \forall t \in \mathbb{R}, \tag{100}$$

i.e., for any measurable set $S \subset \mathbb{R}$,

$$\tilde{v}_t(S) = \sum_{i=1}^{n} c_i \mathcal{P}_t^{[i]}(S), \tag{101}$$

where $c_i \ge 0$, and $\sum_{i=1}^{n} c_i = 1$.

Then $\{\tilde{v}_t\}_{t\in\mathbb{R}}$ also satisfies the differential privacy constraint (22), and

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\tilde{v}_t(dx) \le Q. \tag{102}$$

Proof: First we show that $\{\tilde{v}_t\}_{t\in\mathbb{R}}$ also satisfies the differential privacy constraint (22). Indeed, $\forall |t_1 - t_2| \le \Delta$, $\forall$ measurable set $S \subset \mathbb{R}$,

$$\tilde{v}_{t_1}(S) = \sum_{i=1}^{n} c_i \mathcal{P}_{t_1}^{[i]}(S) \tag{103}$$

$$\le \sum_{i=1}^{n} c_i e^{\epsilon} \mathcal{P}_{t_2}^{[i]}(S + t_1 - t_2) \tag{104}$$

$$= e^{\epsilon} \tilde{v}_{t_2}(S + t_1 - t_2). \tag{105}$$

Next we show that the cost of $\{\tilde{v}_t\}_{t\in\mathbb{R}}$ is no bigger than Q. Indeed, for any $t \in \mathbb{R}$,

$$\int_{x\in\mathbb{R}} \mathcal{L}(x)\tilde{v}_t(dx) = \sum_{i=1}^{n} c_i \int_{x\in\mathbb{R}} \mathcal{L}(x)\tilde{v}_t^{[i]}(dx) \tag{106}$$

$$\le \sum_{i=1}^{n} c_i Q \tag{107}$$

$$= Q. \tag{108}$$

Therefore,

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\tilde{v}_t(dx) \le Q. \tag{109}$$

∎

Applying Lemma 15 and Lemma 16, we can prove the conjecture under the assumption that the family of probability measures $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ is piecewise constant and periodic over t.

Proof of Theorem 3: We first prove that for any family of probability measures $\{\mathcal{P}_t\}_{t\in\mathbb{R}} \in \mathcal{K}_{T,n}$, there exists a new family of probability measures $\{\tilde{\mathcal{P}}_t\}_{t\in\mathbb{R}} \in \mathcal{K}_{T,n}$ such that $\tilde{\mathcal{P}}_t = \tilde{\mathcal{P}}$ for all $t \in \mathbb{R}$, i.e., the added noise is independent of query output t, and

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\tilde{\mathcal{P}}_t(dx) \le \sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t(dx). \tag{110}$$

Indeed, consider the collection of probability measures $\{\mathcal{P}_t^{(i\frac{T}{n})}\}_{t\in\mathbb{R}}$ for $i \in \{0, 1, 2, \ldots, n-1\}$, where $\{\mathcal{P}_t^{(\alpha)}\}$ is defined in (92). Due to Lemma 15, for all i, $\{\mathcal{P}_t^{(i\frac{T}{n})}\}_{t\in\mathbb{R}}$ satisfies the differential privacy constraint (22), and the cost is the same as the cost of $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$.

Define

$$\tilde{\mathcal{P}}_t = \sum_{i=0}^{n-1} \frac{1}{n}\mathcal{P}_t^{(i\frac{T}{n})}. \tag{111}$$

Then due to Lemma 16, $\{\tilde{\mathcal{P}}_t\}_{t\in\mathbb{R}}$ satisfies (22), and the cost of is not worse, i.e.,

$$\sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\tilde{\mathcal{P}}_t(dx) \leq \sup_{t\in\mathbb{R}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_t(dx). \tag{112}$$

Furthermore, since $\{\mathcal{P}_t\}_{t\in\mathbb{R}} \in \mathcal{K}_{T,n}$, for any $t \in \mathbb{R}$,

$$\tilde{\mathcal{P}}_t = \sum_{i=0}^{n-1} \frac{1}{n}\mathcal{P}_t^{(i\frac{T}{n})} = \sum_{i=0}^{n-1} \frac{1}{n}\mathcal{P}_{i\frac{T}{n}}. \tag{113}$$

Hence, $\tilde{\mathcal{P}}_t$ is independent of t.

Therefore, among the collection of probability measures in $\cup_{T>0} \cup_{n\geq 1} \mathcal{K}_{T,n}$, to minimize the cost we only need to consider the families of noise probability measures which are independent of the query output t. Then due to Theorem 4, the staircase mechanism is optimal among all query-output-independent noise-adding mechanisms. This completes the proof of Theorem 3. ∎

APPENDIX B
PROOF OF THEOREM 4

In this section, we give detailed and rigorous proof of Theorem 4.

A. Outline of Proof

The key idea of the proof is to use a sequence of probability distributions with piecewise constant probability density functions to approximate any probability distribution satisfying the differential privacy constraint (33). The proof consists of 8 steps in total, and in each step we narrow down the set of probability distributions where the optimal probability distribution should lie in:

- Step 1 proves that we only need to consider symmetric probability distributions.
- Step 2 and Step 3 prove that we only need to consider probability distributions which have symmetric piecewise constant probability density functions.
- Step 4 proves that we only need to consider those symmetric piecewise constant probability density functions which are monotonically decreasing for $x \geq 0$.
- Step 5 proves that optimal probability density function should periodically decay.
- Step 6, Step 7 and Step 8 prove that the optimal probability density function over the interval $[0, \Delta)$ is a step function, and they conclude the proof of Theorem 4.

B. Step 1

Define

$$V^* \triangleq \inf_{\mathcal{P}\in\mathcal{SP}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx). \tag{114}$$

Our goal is to prove that $V^* = \inf_{\gamma\in[0,1]} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_\gamma(dx)$.

If $V^* = +\infty$, then due to the definition of V^*, we have

$$\inf_{\gamma\in[0,1]} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_\gamma(dx) \geq V^* = +\infty, \tag{115}$$

and thus $\inf_{\gamma\in[0,1]} \int_{x\in\mathbb{R}} \mathcal{L}(x) = V^* = +\infty$. So we only need to consider the case $V^* < +\infty$, i.e., V^* is finite. Therefore, in the rest of the proof, we assume V^* is finite.

First we prove that we only need to consider symmetric probability measures.

Lemma 17: Given $\mathcal{P} \in \mathcal{SP}$, define a symmetric probability distribution $\mathcal{P}_{sym}$ as

$$\mathcal{P}_{sym}(S) \triangleq \frac{\mathcal{P}(S) + \mathcal{P}(-S)}{2}, \quad \forall \text{ measurable set } S \subseteq \mathbb{R}, \tag{116}$$

where the set $-S \triangleq \{-x \mid x \in S\}$. Then $\mathcal{P}_{sym} \in \mathcal{SP}$, i.e., $\mathcal{P}_{sym}$ satisfies the differential privacy constraint (33), and

$$\int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_{sym}(dx) = \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx). \tag{117}$$

Proof: It is easy to verify that $\mathcal{P}_{sym}$ is a valid probability distribution. Due to the definition of $\mathcal{P}_{sym}$ in (116), we have

$$\mathcal{P}_{sym}(S) = \frac{\mathcal{P}(S) + \mathcal{P}(-S)}{2} = \mathcal{P}_{sym}(-S), \tag{118}$$

for any measurable set $S \subseteq \mathbb{R}$. Thus, $\mathcal{P}_{sym}$ is a symmetric probability distribution.

Next, we show that $\mathcal{P}_{sym}$ satisfies (33). Indeed, $\forall$ measurable set $S \subseteq \mathbb{R}$ and $\forall |d| \leq \Delta$,

$$\mathcal{P}_{sym}(S) = \frac{\mathcal{P}(S) + \mathcal{P}(-S)}{2} \tag{119}$$

$$\leq \frac{e^\epsilon \mathcal{P}(S+d) + e^\epsilon \mathcal{P}(-S-d)}{2} \tag{120}$$

$$= \frac{e^\epsilon \mathcal{P}(S+d) + e^\epsilon \mathcal{P}(-(S+d))}{2} \tag{121}$$

$$= e^\epsilon \mathcal{P}_{sym}(S+d), \tag{122}$$

where in (120) we use the facts $\mathcal{P}(S) \leq e^\epsilon \mathcal{P}(S+d)$ and $\mathcal{P}(-S) \leq e^\epsilon \mathcal{P}(-S-d)$.

Lastly, since $\mathcal{L}(x)$ is symmetric,

$$\int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) = \int_{x\in\mathbb{R}} \frac{\mathcal{L}(x)+\mathcal{L}(-x)}{2}\mathcal{P}(dx) \tag{123}$$

$$= \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}_{sym}(dx). \tag{124}$$

∎

Therefore, if we define

$$\mathcal{SP}_{sym} \triangleq \{\mathcal{P}_{sym} | \mathcal{P} \in \mathcal{SP}\}, \tag{125}$$

due to Lemma 17,

Lemma 18:

$$V^* = \inf_{\mathcal{P}\in\mathcal{SP}_{sym}} \int_{x\in\mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx). \tag{126}$$

C. Step 2

Next we prove that for any probability distribution $\mathcal{P}$ satisfying differential privacy constraint (33), the probability $\Pr(X = x) = 0$, $\forall x \in \mathbb{R}$, and $\mathcal{P}([y, z]) \neq 0$ for any $y < z \in \mathbb{R}$.

Lemma 19: $\forall \mathcal{P} \in \mathcal{SP}$, $\forall x \in \mathbb{R}$, $\mathcal{P}(\{x\}) = 0$. *And, for any* $y < z \in \mathbb{R}$, $\mathcal{P}([y, z]) \neq 0$.

Proof: Given $\mathcal{P} \in \mathcal{SP}$, suppose $\mathcal{P}(\{x_0\}) = p_0 > 0$, for some $x_0 \in \mathbb{R}$. Then for any $x \in [x_0, x_0 + \Delta]$,

$$\mathcal{P}(\{x\}) \geq e^{-\epsilon}, \tag{127}$$

due to (33).

So $\mathcal{P}(\{x\})$ is strictly lower bounded by a positive constant for uncountable number of x, and thus $\mathcal{P}([x_0, x_0+\Delta]) = +\infty$, which contradicts with the fact $\mathcal{P}$ is a probability distribution.

Therefore, $\forall \mathcal{P} \in \mathcal{SP}$, $\forall x \in \mathbb{R}$, $\mathcal{P}(\{x\}) = 0$.

Suppose $\mathcal{P}([y, z]) = 0$ for some $y < z \in \mathbb{R}$. Then from (33) we have for any $|d| \leq \Delta$,

$$\mathcal{P}([y + d, z + d]) \leq e^{\epsilon} \mathcal{P}([y, z]) = 0, \tag{128}$$

and thus $\mathcal{P}([y + d, z + d]) = 0$. By induction, for any $k \in \mathbb{Z}$, $\mathcal{P}([y + kd, z + kd]) = 0$, which implies that $\mathcal{P}((-\infty, +\infty)) = 0$. Contradiction. So for any $y < z \in \mathbb{R}$, $\mathcal{P}([y, z]) \neq 0$. ∎

D. Step 3

In this subsection, we show that for any $\mathcal{P} \in \mathcal{SP}_{\mathrm{sym}}$ with

$$V(\mathcal{P}) \triangleq \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) < +\infty, \tag{129}$$

we can use a sequence of probability measures $\{\mathcal{P}_i \in \mathcal{SP}_{\mathrm{sym}}\}_{i \geq 1}$ with symmetric piecewise constant probability density functions to approximate $\mathcal{P}$ with $\lim_{i \to +\infty} V(\mathcal{P}_i) = V(\mathcal{P})$.

Lemma 20: Given $\mathcal{P} \in \mathcal{SP}_{\mathrm{sym}}$ with $V(\mathcal{P}) < +\infty$, any positive integer $i \in N$, define $\mathcal{P}_i$ as the probability distribution with a symmetric probability density function $f_i(x)$ defined as

$$f_i(x) = \begin{cases} a_k \triangleq \dfrac{\mathcal{P}([k\frac{D}{i}, (k+1)\frac{D}{i}))}{\frac{D}{i}} & x \in [k\frac{D}{i}, (k+1)\frac{D}{i}) \text{ for } k \in \mathbb{N} \\ f_i(-x) & x < 0 \end{cases} \tag{130}$$

Then $\mathcal{P}_i \in \mathcal{SP}_{\mathrm{sym}}$ and

$$\lim_{i \to +\infty} V(\mathcal{P}_i) = V(\mathcal{P}). \tag{131}$$

Proof: First we prove that $\mathcal{P}_i \in \mathcal{SP}_{\mathrm{sym}}$, i.e., $\mathcal{P}_i$ is symmetric and satisfies the differential privacy constraint (33).

By definition $f_i(x)$ is a symmetric and nonnegative function, and

$$\int_{-\infty}^{+\infty} f_i(x)dx = 2 \int_0^{+\infty} f_i(x)dx \tag{132}$$

$$= 2 \int_{x \in [0, +\infty)} \mathcal{P}(dx) \tag{133}$$

$$= 2 \int_{x \in (0, +\infty)} \mathcal{P}(dx) \tag{134}$$

$$= 1, \tag{135}$$

where in (134) we used the fact $\mathcal{P}(\{0\}) = 0$ due to Lemma 19. In addition, due to Lemma 19, $a_k > 0$, $\forall k \in \mathbb{N}$.

So $f_i(x)$ is a valid symmetric probability density function, and thus $\mathcal{P}_i$ is a valid symmetric probability distribution.

Define the density sequence of $\mathcal{P}_i$ as the sequence $\{a_0, a_1, a_2, \ldots, a_n, \ldots\}$. Since $\mathcal{P}$ satisfies (33), it is easy to see that

$$a_j \leq e^{\epsilon} a_{j+k} \text{ and } a_{j+k} \leq e^{\epsilon} a_j, \quad \forall j \geq 0, \ 0 \leq k \leq i. \tag{136}$$

Therefore, for any x, y such that $|x - y| \leq \Delta$, we have

$$f_i(x) \leq e^{\epsilon} f_i(y) \text{ and } f_i(y) \leq e^{\epsilon} f_i(x), \tag{137}$$

which implies that $\mathcal{P}_i$ satisfies (33). Hence, $\mathcal{P}_i \in \mathcal{SP}_{\mathrm{sym}}$.

Next we show that

$$\lim_{i \to +\infty} V(\mathcal{P}_i) = V(\mathcal{P}). \tag{138}$$

Since $\mathcal{L}(x)$ satisfies Property 2, we can assume there exists a constant $B > 0$ such that

$$\mathcal{L}(x + 1) \leq B\mathcal{L}(x), \quad \forall x \geq T. \tag{139}$$

Given $\delta > 0$, since $V(\mathcal{P})$ is finite, there exists integer $T^* > T$ such that

$$\int_{x \geq T^*} \mathcal{L}(x)\mathcal{P}(dx) < \frac{\delta}{B}. \tag{140}$$

For any integers $i \geq 1$, $N \geq T^*$,

$$\int_{x \in [N, N+1)} \mathcal{L}(x)\mathcal{P}_i(dx) \leq \mathcal{P}_i([N, N+1))\mathcal{L}(N+1) \tag{141}$$

$$= \mathcal{P}([N, N+1))\mathcal{L}(N+1) \tag{142}$$

$$\leq \int_{x \in [N, N+1)} B\mathcal{L}(x)\mathcal{P}(dx). \tag{143}$$

Therefore,

$$\int_{x \in [T^*, +\infty)} \mathcal{L}(x)\mathcal{P}_i(dx) \leq \int_{x \in [T^*, +\infty)} B\mathcal{L}(x)\mathcal{P}(dx) \tag{144}$$

$$\leq B\frac{\delta}{B} \tag{145}$$

$$= \delta. \tag{146}$$

For $x \in [0, T^*)$, $\mathcal{L}(x)$ is a bounded function, and thus by the definition of Riemann-Stieltjes integral, we have

$$\lim_{i \to \infty} \int_{x \in [0, T^*)} \mathcal{L}(x)\mathcal{P}_i(dx) = \int_{x \in [0, T^*)} \mathcal{L}(x)\mathcal{P}(dx). \tag{147}$$

So there exists a sufficiently large integer i^* such that for all $i \geq i^*$

$$\left| \int_{x \in [0, T^*)} \mathcal{L}(x)\mathcal{P}_i(dx) - \int_{x \in [0, T^*)} \mathcal{L}(x)\mathcal{P}(dx) \right| \leq \delta. \tag{148}$$

Hence, for all $i \geq i^*$

$$|V(\mathcal{P}_i) - V(\mathcal{P})|$$
$$= \left| \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}_i(dx) - \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) \right|$$
$$= 2\left| \int_{x \in [0,T^*)} \mathcal{L}(x)\mathcal{P}_i(dx) - \int_{x \in [0,T^*)} \mathcal{L}(x)\mathcal{P}(dx) \right.$$
$$+ \int_{x \in [T^*,+\infty)} \mathcal{L}(x)\mathcal{P}_i(dx) - \int_{x \in [T^*,+\infty)} \mathcal{L}(x)\mathcal{P}(dx) \Big|$$
$$\leq 2 \left| \int_{x \in [0,T^*)} \mathcal{L}(x)\mathcal{P}_i(dx) - \int_{x \in [0,T^*)} \mathcal{L}(x)\mathcal{P}(dx) \right|$$
$$+ 2 \int_{x \in [T^*,+\infty)} \mathcal{L}(x)\mathcal{P}_i(dx) + 2 \int_{x \in [T^*,+\infty)} \mathcal{L}(x)\mathcal{P}(dx)$$
$$\leq 2\left(\delta + \delta + \frac{\delta}{B}\right)$$
$$\leq \left(4 + \frac{2}{B}\right)\delta.$$

Therefore,

$$\lim_{i \to +\infty} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}_i(dx) = \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx). \quad (149)$$

∎

Define $\mathcal{SP}_{i,\mathrm{sym}} \triangleq \{\mathcal{P}_i | \mathcal{P} \in \mathcal{SP}_{\mathrm{sym}}\}$ for $i \geq 1$, i.e., $\mathcal{SP}_{i,\mathrm{sym}}$ is the set of probability distributions satisfying differential privacy constraint (33) and having symmetric piecewise constant (over intervals $[k\frac{\Delta}{i}, (k+1)\frac{\Delta}{i}) \ \forall k \in \mathbb{N}$) probability density functions.

Due to Lemma 20,

Lemma 21:

$$V^* = \inf_{\mathcal{P} \in \cup_{i=1}^{\infty} \mathcal{SP}_{i,\mathrm{sym}}} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx). \quad (150)$$

Therefore, to characterize V^*, we only need to study probability distributions with symmetric and piecewise constant probability density functions.

E. Step 4

Next we show that indeed we only need to consider those probability distributions with symmetric piecewise constant probability density functions which are *monotonically decreasing* when $x \geq 0$.

Lemma 22: Given $\mathcal{P}_a \in \mathcal{SP}_{i,\mathrm{sym}}$ *with symmetric piecewise constant probability density function* $f(\cdot)$, *let* $\{a_0, a_1, \ldots, a_n, \ldots\}$ *be the density sequence of* $f(\cdot)$, *i.e,*

$$f(x) = a_k, \quad x \in [k\frac{\Delta}{i}, (k+1)\frac{\Delta}{i}) \ \forall k \in \mathbb{N}. \quad (151)$$

Then we can construct a new probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{sym}}$ *the probability density function of which is monotonically decreasing when* $x \geq 0$, *and*

$$\int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}_b(dx) \leq \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}_a(dx). \quad (152)$$

Proof: Since $a_k > 0, \forall k \in \mathbb{N}$, and

$$\sum_{k=0}^{+\infty} a_k \frac{\Delta}{i} = \frac{1}{2}, \quad (153)$$

we have $\lim_{k \to +\infty} a_k = 0$.

Given the density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$, construct a new monotonically decreasing density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ and a bijective mapping $\pi : \mathbb{N} \to \mathbb{N}$ as follows

$$I_0 = \arg\max_{k \in \mathbb{N}} a_k, \quad (154)$$
$$\pi(0) = \min_{n \in I_0} n, \text{ i.e., the smallest element in } I_0, \quad (155)$$
$$b_0 = a_{\pi(0)}, \quad (156)$$
$$\forall m \in \mathbb{N} \text{ and } m \geq 1, \quad (157)$$
$$I_m = \arg\max_{k \in \mathbb{N}\setminus\{\pi(j)|j<m\}} a_k, \quad (158)$$
$$\pi(m) = \min_{n \in I_m} n, \text{ i.e., the smallest element in } I_m, \quad (159)$$
$$b_m = a_{\pi(m)}. \quad (160)$$

Since the sequence $\{a_k\}$ converges to 0, the maximum of $\{a_k\}$ always exists in (154) and (158). Therefore, I_m is well defined for all $m \in \mathbb{N}$.

Note that since $\sum_{k=0}^{\infty} a_k \frac{\Delta}{i} = \frac{1}{2}$ and the sequence $\{b_k\}_{k \in \mathbb{N}}$ is simply a permutation of $\{a_k\}_{k \in \mathbb{N}}$, $\sum_{k=1}^{\infty} b_k \frac{\Delta}{i} = \frac{1}{2}$.

Therefore, if we define a function $g(\cdot)$ as

$$g(x) = \begin{cases} b_k & x \in [k\frac{D}{i}, (k+1)\frac{D}{i}) \text{ for } k \in \mathbb{N} \\ g(-x) & x < 0 \end{cases} \quad (161)$$

then $g(\cdot)$ is a valid symmetric probability density function, and

$$\int_{x \in \mathbb{R}} \mathcal{L}(x)g(x)dx \leq \int_{x \in \mathbb{R}} \mathcal{L}(x)f(x)dx. \quad (162)$$

Next, we prove that the probability distribution $\mathcal{P}_b$ with probability density function $g(\cdot)$ satisfies the differential privacy constraint (33). Since $\{b_k\}_{k \in \mathbb{N}}$ is a monotonically decreasing sequence, it is sufficient and necessary to prove that for all $k \in \mathbb{N}$,

$$\frac{b_k}{b_{k+i}} \leq e^\epsilon. \quad (163)$$

To simplify notation, given k, we define

$$a^*(k) = \min_{k \leq j \leq k+i} a_k, \quad (164)$$

i.e., $a^*(k)$ denotes the smallest number of $\{a_k, a_{k+1}, \ldots, a_{k+i}\}$.

First, when $k = 0$, it is easy to prove that $\frac{b_0}{b_i} \leq e^\epsilon$. Indeed, recall that $b_0 = a_{\pi(0)}$ and consider the $i + 1$ consecutive numbers $\{a_{\pi(0)}, a_{\pi(0)+1}, \ldots, a_{\pi(0)+i}\}$ in the original sequence $\{a_k\}_{k \in \mathbb{N}}$. Then $a^*(0) \leq b_i$, since b_i is the $(i+1)$th largest number in the sequence $\{a_k\}_{k \in \mathbb{N}}$. Therefore,

$$\frac{b_0}{b_i} = \frac{a_{\pi(0)}}{b_i} \leq \frac{a_{\pi(0)}}{a^*(0)} \leq e^\epsilon. \quad (165)$$

For $k = 1$, $b_1 = a_{\pi(1)}$ and consider the $i + 1$ consecutive numbers $\{a_{\pi(1)}, a_{\pi(1)+1}, \ldots, a_{\pi(1)+i}\}$. If $\pi(0) \notin [\pi(1), \pi(1) + i]$, then $a^*(\pi(1)) \leq b_{i+1}$, and thus

$$\frac{b_1}{b_{1+i}} = \frac{a_{\pi(1)}}{b_{1+i}} \leq \frac{a_{\pi(1)}}{a^*(\pi(1))} \leq e^\epsilon. \quad (166)$$

Algorithm 2 Identifying j^* such that $j* < k$ and $a^*(\pi(j^*)) \leq b_{k+i}$

   $j^* \leftarrow k$
   while there exists some $j < k$ and $\pi_j \in [\pi(j^*)+1, \pi(j^*)+ i]$ **do**
      $j^* \leftarrow j$
   end while
   Output j^*

If $\pi(0) \in [\pi(1), \pi(1) + i]$, then $a^*(\pi(0)) \leq b_{i+1}$ and $\frac{a_{\pi(0)}}{a^*(\pi(0))} \leq e^\epsilon$. Therefore,

$$\frac{b_1}{b_{i+1}} \leq \frac{b_0}{b_{1+i}} \leq \frac{b_0}{a^*(\pi(0))} \leq e^\epsilon. \tag{167}$$

Hence, $\frac{b_k}{b_{k+i}} \leq e^\epsilon$ holds for $k = 1$.

In general, given k, we prove $\frac{b_k}{b_{k+i}} \leq e^\epsilon$ as follows. First, if $\pi_j \notin [\pi(k), \pi(k) + i]$, $\forall j < k$, then $a^*(\pi(k)) \leq b_{k+i}$, and hence

$$\frac{b_k}{b_{i+k}} = \frac{a_{\pi(k)}}{b_{i+k}} \leq \frac{a_{\pi(k)}}{a^*(\pi(k))} \leq e^\epsilon. \tag{168}$$

If there exists $j < k$ and $\pi_j \in [\pi(k) + 1, \pi(k) + i]$, we use Algorithm 2 to compute a number j^* such that $j^* < k$ and $\pi_j \notin [\pi(j^*) + 1, \pi(j^*) + i]$, $\forall j < k$.

It is easy to show that the loop in Algorithm 2 will terminate after at most k steps.

After finding j^*, we have $j^* < k$, and $a^*(\pi(j^*)) \leq b_{k+i}$. Therefore

$$\frac{b_k}{b_{i+k}} \leq \frac{a_{\pi(j^*)}}{b_{i+k}} \leq \frac{a_{\pi(j^*)}}{a^*(\pi(j^*))} \leq e^\epsilon. \tag{169}$$

So $\frac{b_k}{b_{k+i}} \leq e^\epsilon$ holds for all $k \in \mathbb{N}$. Therefore, $\mathcal{P}_b \in \mathcal{SP}_{i,\text{sym}}$. This completes the proof of Lemma 22. ∎

Therefore, if we define

$\mathcal{SP}_{i,\text{md}} \triangleq \{\mathcal{P}|\mathcal{P} \in \mathcal{SP}_{i,\text{sym}}, \text{ and the density sequence of } \mathcal{P}$
$\qquad\qquad\qquad\qquad\qquad \text{is monotonically decreasing}\},$

then due to Lemma 22,

Lemma 23:

$$V^* = \inf_{\mathcal{P} \in \cup_{i=1}^{\infty} \mathcal{SP}_{i,\text{md}}} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx). \tag{170}$$

F. Step 5

Next we show that among all symmetric piecewise constant probability density functions, we only need to consider those which are periodically decaying.

More precisely, given positive integer i,

$\mathcal{SP}_{i,\text{pd}} \triangleq \{\mathcal{P}|\mathcal{P} \in \mathcal{SP}_{i,\text{md}}, \text{ and } \mathcal{P} \text{ has density sequence}$
$\{a_0, a_1, \ldots, a_n, \ldots, \} \text{ satisfying } \frac{a_k}{a_{k+i}} = e^\epsilon, \forall k \in \mathbb{N}\},$

then

Lemma 24:

$$V^* = \inf_{\mathcal{P} \in \cup_{i=1}^{\infty} \mathcal{SP}_{i,\text{pd}}} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx). \tag{171}$$

Proof: Due to Lemma 23, we only need to consider probability distributions with symmetric and piecewise constant probability density functions which are monotonically decreasing for $x \geq 0$.

We first show that given $\mathcal{P}_a \in \mathcal{SP}_{i,\text{md}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots, \}$, if $\frac{a_0}{a_i} < e^\epsilon$, then we can construct a probability distributions $\mathcal{P}_b \in \mathcal{SP}_{i,\text{md}}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots, \}$ such that $\frac{b_0}{b_i} = e^\epsilon$ and

$$V(\mathcal{P}_a) \geq V(\mathcal{P}_b). \tag{172}$$

Define a new sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ by scaling up a_0 and scaling down $\{a_1, a_2, \ldots\}$. More precisely, let $\delta = \frac{i}{2D((\frac{i}{2D}-a_0)e^{-\epsilon}\frac{a_0}{a_i}+a_0)} - 1 > 0$, and set

$$b_0 = a_0(1 + \delta), \tag{173}$$
$$b_k = a_k(1 - \delta'), \quad \forall k \geq 1, \tag{174}$$

where $\delta' \triangleq \frac{a_0\delta}{\frac{i}{2D}-a_0} > 0$, and we have chosen δ such that $\frac{b_0}{b_i} = \frac{a_0}{a_k}\frac{\frac{i}{2D}-a_0}{\frac{i}{2D(1+\delta)}-a_0} = e^\epsilon$.

It is easy to see the sequence $\{b_0, b_1, \ldots, b_n, \ldots, \}$ corresponds to a valid probability density function and it also satisfies the differential privacy constraint (33), i.e.,

$$\frac{b_k}{b_{k+i}} \leq e^\epsilon, \quad \forall k \geq 0. \tag{175}$$

Let $\mathcal{P}_b$ be the probability distribution with $\{b_0, b_1, \ldots, b_n, \ldots, \}$ as the density sequence of its probability density function. Next we show $V(\mathcal{P}_b) \leq V(\mathcal{P}_a)$.

It is easy to compute $V(\mathcal{P}_a)$, which is

$$V(\mathcal{P}_a) = 2\frac{\Delta}{i}\left(a_0 \int_0^{\frac{\Delta}{i}} \mathcal{L}(x)dx + \sum_{k=1}^{\infty} a_k \int_{k\frac{\Delta}{i}}^{(k+1)\frac{\Delta}{i}} \mathcal{L}(x)dx\right). \tag{176}$$

Similarly, we can compute $V(\mathcal{P}_b)$ by

$V(\mathcal{P}_b)$
$$= 2\frac{\Delta}{i}\left(b_0 \int_0^{\frac{\Delta}{i}} \mathcal{L}(x)dx + \sum_{k=1}^{\infty} b_k \int_{k\frac{\Delta}{i}}^{(k+1)\frac{\Delta}{i}} \mathcal{L}(x)dx\right)$$
$$= V(\mathcal{P}_a)$$
$$+ 2\frac{\Delta}{i}\left(a_0\delta \int_0^{\frac{D}{i}} \mathcal{L}(x)dx - \delta' \sum_{k=1}^{\infty} a_k \int_{k\frac{D}{i}}^{(k+1)\frac{D}{i}} \mathcal{L}(x)dx\right)$$
$$= V(\mathcal{P}_a) + 2\frac{\Delta}{i}\frac{a_0\delta}{\frac{i}{2\Delta}-a_0}$$
$$\times \left(\sum_{k=1}^{\infty} a_k \int_0^{\frac{\Delta}{i}} \mathcal{L}(x)dx - \sum_{k=1}^{\infty} a_k \int_{k\frac{\Delta}{i}}^{(k+1)\frac{\Delta}{i}} \mathcal{L}(x)dx\right)$$
$$= V(\mathcal{P}_a) + 2\frac{\Delta}{i}\frac{a_0\delta}{\frac{i}{2\Delta}-a_0}$$
$$\times \sum_{k=1}^{\infty} a_k \left(\int_0^{\frac{\Delta}{i}} \mathcal{L}(x)dx - \int_{k\frac{\Delta}{i}}^{(k+1)\frac{\Delta}{i}} \mathcal{L}(x)dx\right)$$
$$\leq V(\mathcal{P}_a),$$

where in the last step we used the fact that $\left(\int_0^{\frac{\Delta}{i}} \mathcal{L}(x)dx - \int_{k\frac{\Delta}{i}}^{(k+1)\frac{\Delta}{i}} \mathcal{L}(x)dx\right) \leq 0$, since $\mathcal{L}(\cdot)$ is a monotonically increasing function for $x \geq 0$.

Therefore, for given $i \in \mathbb{N}$, we only need to consider $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying $\frac{a_0}{a_i} = e^\epsilon$.

Next, we argue that among all probability distributions $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots, \}$ satisfying $\frac{a_0}{a_i} = e^\epsilon$, we only need to consider those probability distributions with density sequence also satisfying $\frac{a_1}{a_{i+1}} = e^\epsilon$.

Given $\mathcal{P}_a \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying $\frac{a_0}{a_i} = e^\epsilon$ and $\frac{a_1}{a_{i+1}} < e^\epsilon$, we can construct a new probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ satisfying

$$\frac{b_0}{b_i} = e^\epsilon, \tag{177}$$

$$\frac{b_1}{b_{i+1}} = e^\epsilon, \tag{178}$$

and $V(\mathcal{P}_a) \geq V(\mathcal{P}_b)$.

First, it is easy to see a_1 is strictly less than a_0, since if $a_0 = a_1$, then $\frac{a_1}{a_{i+1}} = \frac{a_0}{a_{i+1}} \geq \frac{a_0}{a_i} = e^\epsilon$. Then we construct a new density sequence by increasing a_1 and decreasing a_{i+1}. More precisely, we define a new sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ as

$$b_k = a_k, \quad \forall k \neq 1, \; k \neq i+1, \tag{179}$$

$$b_1 = a_1 + \delta, \tag{180}$$

$$b_{i+1} = a_{i+1} - \delta, \tag{181}$$

where $\delta = \frac{e^\epsilon a_{i+1} - a_1}{1 + e^\epsilon}$ and thus $\frac{b_1}{b_{i+1}} = e^\epsilon$.

It is easy to verify that $\{b_0, b_1, \ldots, b_n, \ldots\}$ is a valid probability density sequence and the corresponding probability distribution $\mathcal{P}_b$ satisfies the differential privacy constraint (33). Moreover, $V(\mathcal{P}_a) \geq V(\mathcal{P}_b)$. Therefore, we only need to consider $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequences $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying $\frac{a_0}{a_i} = e^\epsilon$ and $\frac{a_1}{a_{i+1}} = e^\epsilon$.

Use the same argument, we can show that we only need to consider $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{md}}$ with density sequences $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying

$$\frac{a_k}{a_{i+k}} = e^\epsilon, \quad \forall k \geq 0. \tag{182}$$

Therefore,

$$V^* = \inf_{\mathcal{P} \in \cup_{i=1}^\infty \mathcal{SP}_{i,\mathrm{pd}}} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx). \tag{183}$$

∎

Due to Lemma 24, we only need to consider probability distribution with symmetric, monotonically decreasing (for $x \geq 0$), and periodically decaying piecewise constant probability density function. Because of the properties of symmetry and periodically decaying, for this class of probability distributions, the probability density function over $\mathbb{R}$ is completely determined by the probability density function over the interval $[0, \Delta)$.

Next, we study what the optimal probability density function should be over the interval $[0, \Delta)$. It turns out that the optimal probability density function over the interval $[0, \Delta)$ is a step function. We use the following three steps to prove this result.

G. Step 6

Lemma 25: Consider a probability distribution $\mathcal{P}_a \in \mathcal{SP}_{i,pd}$ ($i \geq 2$) with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$, and $\frac{a_0}{a_{i-1}} < e^\epsilon$. Then there exists a probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,pd}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ such that $\frac{b_0}{b_{i-1}} = e^\epsilon$, and

$$V(\mathcal{P}_b) \leq V(\mathcal{P}_a). \tag{184}$$

Proof: For each $0 \leq k \leq (i-1)$, define

$$w_k \triangleq \sum_{j=0}^{+\infty} e^{-j\epsilon} \int_{(j+\frac{k}{i})\Delta}^{(j+\frac{k+1}{i})\Delta} \mathcal{L}(x)dx. \tag{185}$$

Since $\mathcal{L}(\cdot)$ satisfies Property 2 and $V^* < \infty$, it is easy to show that the sum of series in (185) exists and is finite, and thus w_k is well defined for all $0 \leq k \leq (i-1)$. In addition, it is easy to see

$$w_0 \leq w_1 \leq w_2 \leq \cdots \leq w_{i-1}, \tag{186}$$

since $\mathcal{L}(x)$ is a monotonically increasing function when $x \geq 0$. Then

$$V(\mathcal{P}_a) = \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}_a(dx) = 2\sum_{k=0}^{i-1} w_k a_k. \tag{187}$$

Since $\frac{a_0}{a_{i-1}} < e^\epsilon$, we can scale a_0 up and scale $\{a_1, \ldots, a_{i-1}\}$ down to derive a new valid probability density function with smaller cost. More precisely, define a new probability measure $\mathcal{P}_b \in \mathcal{SP}_{i,pd}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ via

$$b_0 \triangleq \gamma a_0, \tag{188}$$

$$b_k \triangleq \gamma' a_k, \quad \forall 1 \leq k \leq i-1, \tag{189}$$

for some $\gamma > 1$ and $\gamma' < 1$ such that

$$\frac{b_0}{b_{i-1}} = e^\epsilon. \tag{190}$$

To make $\{b_0, b_1, \ldots, b_n, \ldots\}$ be a valid density sequence, i.e., to make the integral of the corresponding probability density function over $\mathbb{R}$ be 1, we have

$$\sum_{k=0}^{i-1} b_k = \sum_{k=0}^{i-1} a_k = \frac{1 - e^{-\epsilon}}{2} \frac{i}{\Delta}. \tag{191}$$

Define $t \triangleq \frac{1 - e^{-\epsilon}}{2} \frac{i}{\Delta}$, then we have two linear equations on γ and γ':

$$\gamma a_0 = e^\epsilon \gamma' \tag{192}$$

$$\gamma a_0 + \gamma'(t - a_0) = t. \tag{193}$$

From (192) and (193), we can easily get

$$\gamma = \frac{e^\epsilon t a_{i-1}}{a_0(t - a_0 + e^\epsilon a_{i-1})} > 1 \tag{194}$$

$$\gamma' = \frac{t}{t - a_0 + e^\epsilon a_{i-1}} < 1. \tag{195}$$

Then we can verify that the $V(\mathcal{P}_a) \geq V(\mathcal{P}_a)$. Indeed,

$$V(\mathcal{P}_a) - V(\mathcal{P}_b) \tag{196}$$

$$= \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_a(dx) - \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_b(dx) \tag{197}$$

$$= 2 \sum_{k=0}^{i-1} w_k a_k - 2 \sum_{k=0}^{i-1} w_k b_k \tag{198}$$

$$= 2 \left((1-\gamma) w_0 a_0 + (1-\gamma') \sum_{k=1}^{i-1} w_k a_k \right) \tag{199}$$

$$\geq 2 \left((1-\gamma) w_0 a_0 + (1-\gamma') \sum_{k=1}^{i-1} w_0 a_k \right) \tag{200}$$

$$= 2 \left((1-\gamma) w_0 a_0 + (1-\gamma') w_0 (t - a_0) \right) \tag{201}$$

$$= 2 w_0 \left(a_0 - \frac{a_{i-1} e^\epsilon t}{t - a_0 + e^\epsilon a_{i-1}} + (t - a_0) \frac{-a_0 + e^\epsilon a_{i-1}}{t - a_0 + e^\epsilon a_{i-1}} \right) \tag{202}$$

$$= 0. \tag{203}$$

This completes the proof. ∎

Therefore, due to Lemma 25, for all $i \geq 2$, we only need to consider probability distributions $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{pd}}$ with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying $\frac{a_0}{a_{i-1}} = e^\epsilon$.

More precisely, define

$$\mathcal{SP}_{i,\mathrm{fr}} = \{\mathcal{P} \in \mathcal{SP}_{i,\mathrm{pd}} | \mathcal{P} \text{ has density sequence}$$
$$\{a_0, a_1, \ldots, a_n, \ldots\} \text{ satisfying } \frac{a_0}{a_{i-1}} = e^\epsilon\}.$$

Then due to Lemma 25,

Lemma 26:

$$V^* = \inf_{\mathcal{P} \in \cup_{i=3}^\infty \mathcal{SP}_{i,\mathrm{fr}}} \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}(dx). \tag{204}$$

H. Step 7

Next, we argue that for each probability distribution $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{fr}}$ ($i \geq 3$) with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$, we can assume that there exists an integer $1 \leq k \leq (i-2)$, such that

$$a_j = a_0, \quad \forall 0 \leq j < k, \tag{205}$$
$$a_j = a_{i-1}, \quad \forall k < j < i. \tag{206}$$

More precisely,

Lemma 27: Consider a probability distribution $\mathcal{P}_a \in \mathcal{SP}_{i,\mathrm{fr}}$ ($i \geq 3$) with density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$. Then there exists a probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{fr}}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ such that there exists an integer $1 \leq k \leq (i-2)$ with

$$b_j = a_0, \quad \forall 0 \leq j < k, \tag{207}$$
$$b_j = a_{i-1}, \quad \forall k < j < i, \tag{208}$$

and

$$V(\mathcal{P}_b) \leq V(\mathcal{P}_a). \tag{209}$$

Proof: If there exists integer $1 \leq k \leq (i-2)$ such that

$$a_j = a_0, \quad \forall 0 \leq j < k, \tag{210}$$
$$a_j = a_{i-1}, \quad \forall k < j < i, \tag{211}$$

then we can set $\mathcal{P}_b = \mathcal{P}_a$.

Otherwise, let k_1 be the smallest integer in $\{0, 1, 2, \ldots, i-1\}$ such that

$$a_{k_1} \neq a_0, \tag{212}$$

and let k_2 be the biggest integer in $\{0, 1, 2, \ldots, i-1\}$ such that

$$a_{k_2} \neq a_{i-1}. \tag{213}$$

It is easy to see that $k_1 \neq k_2$. Then we can increase a_{k_1} and decrease a_{k_2} simultaneously by the same amount to derive a new probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{fr}}$ with smaller cost. Indeed, if

$$a_0 - a_{k_1} \leq a_{k_2} - a_{i-1}, \tag{214}$$

then consider a probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{fr}}$ with density sequence $\{b_0, b_1, \ldots, b_{i-1}, \ldots\}$ defined as

$$b_j = a_0, \quad \forall 0 \leq j \leq k_1, \tag{215}$$
$$b_j = a_j, \quad \forall k_1 < j \leq k_2 - 1, \tag{216}$$
$$b_{k_2} = a_{k_2} - (a_0 - a_{k_1}), \tag{217}$$
$$b_j = a_j, \quad \forall k_2 < j \leq i - 1. \tag{218}$$

We can verify that $V(\mathcal{P}_a) \geq V(\mathcal{P}_b)$ via

$$V(\mathcal{P}_a) - V(\mathcal{P}_b) \tag{219}$$

$$= \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_a(dx) - \int_{x \in \mathbb{R}} \mathcal{L}(x) \mathcal{P}_b(dx) \tag{220}$$

$$= 2(w_{k_1} b_{k_1} + w_{k_2} b_{k_2}) - 2(w_{k_1} a_{k_1} + w_{k_2} a_{k_2}) \tag{221}$$

$$= 2 w_{k_1}(a_0 - a_{k_1}) + 2 w_{k_2}(a_{k_2} - (a_0 - a_{k_1}) - a_{k_2}) \tag{222}$$

$$= 2(a_0 - a_{k_1})(w_{k_1} - w_{k_2}) \tag{223}$$

$$\leq 0, \tag{224}$$

where w_i is defined in (185).

If $a_0 - a_{k_1} \geq a_{k_2} - a_{i-1}$, then accordingly we can construct $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{fr}}$ by setting

$$b_j = a_0, \quad \forall 0 \leq j < k_1, \tag{225}$$
$$b_{k_1} = a_{k_1} + (a_{k_2} - a_{i-1}), \tag{226}$$
$$b_j = a_j, \quad \forall k_1 < j \leq k_2 - 1, \tag{227}$$
$$b_j = a_{i-1}, \quad \forall k_2 \leq j \leq i - 1. \tag{228}$$

And similarly, it is easy to verify that $V(\mathcal{P}_a) \geq V(\mathcal{P}_b)$.

Therefore, continue in this way, and finally we will obtain a probability distribution $\mathcal{P}_b \in \mathcal{SP}_{i,\mathrm{fr}}$ with density sequence $\{b_0, b_1, \ldots, b_n, \ldots\}$ such that (207), (208) and (209) hold. This completes the proof. ∎

Define

$$\mathcal{SP}_{i,\mathrm{step}} = \{\mathcal{P} \in \mathcal{SP}_{i,\mathrm{fr}} \mid \mathcal{P} \text{ has density sequence}$$
$$\{a_0, a_1, \ldots, a_n, \ldots\} \text{ satisfying (207) and (208)}$$
$$\text{for some } 1 \leq k \leq (i-2)\}.$$

Then due to Lemma 27,

Lemma 28:

$$V^* = \inf_{\mathcal{P} \in \cup_{i=3}^{\infty} \mathcal{SP}_{i,step}} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx). \tag{229}$$

I. Step 8

Proof of Theorem 4: Since $\{\mathcal{P}_\gamma | \gamma \in [0, 1]\} \subseteq \mathcal{SP}$, we have

$$V^* = \inf_{\mathcal{P} \in \mathcal{SP}} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) \le \inf_{\gamma \in [0,1]} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}_\gamma(dx). \tag{230}$$

We prove the reverse direction in the following.

We first prove that for any $\mathcal{P} \in \mathcal{SP}_{i,step}$ ($i \ge 3$), there exists $\gamma \in [0, 1]$ such that

$$\int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}_\gamma(dx) \le \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx). \tag{231}$$

Consider the density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$ of $\mathcal{P}$. Since $\mathcal{P} \in \mathcal{SP}_{i,step}$, there exists an integer $0 \le k \le i-2$ such that

$$a_j = a_0, \quad \forall 0 \le j < k, \tag{232}$$
$$a_j = a_0 e^{-\epsilon}, \quad \forall k < j \le i-1. \tag{233}$$

Let

$$\gamma' \triangleq \frac{\frac{1-e^{-\epsilon}}{2\Delta} - a_0 e^{-\epsilon}}{a_0(1-e^{-\epsilon})} \in [0, 1]. \tag{234}$$

Then $a(\gamma') = a_0$.

It is easy to verify that

$$k\frac{\Delta}{i} \le \gamma'\Delta \le (k+1)\frac{\Delta}{i}. \tag{235}$$

The probability density functions of $\mathcal{P}$ and $\mathcal{P}_{\gamma'}$ are the same when $x \in [0, \frac{k}{i}\Delta) \cup [\frac{k+1}{i}\Delta, \Delta)$. Since the integral of probability density functions over $[0, \Delta)$ is $\frac{1-e^{-\epsilon}}{2}$ due to the periodically decaying property, we have

$$a_k\frac{\Delta}{i} = a_0(\gamma' - \frac{k}{i})\Delta + e^{-\epsilon}a_0(\frac{k+1}{i} - \gamma')\Delta. \tag{236}$$

Define $\beta \triangleq i(\gamma' - \frac{k}{i}) \in [0, 1]$. Then

$$a_k = \beta a_0 + (1-\beta)e^{-\epsilon}a_0. \tag{237}$$

Define

$$w_k^{(1)} \triangleq \sum_{j=0}^{+\infty} e^{-j\epsilon} \int_{(j+\frac{k}{i})\Delta}^{(j+\gamma')\Delta} \mathcal{L}(x)dx, \tag{238}$$

$$w_k^{(2)} \triangleq \sum_{j=0}^{+\infty} e^{-j\epsilon} \int_{(j+\gamma')\Delta}^{(j+\frac{k+1}{i})\Delta} \mathcal{L}(x)dx. \tag{239}$$

Note that $w_k = w_k^{(1)} + w_k^{(2)}$. Since $\mathcal{L}(x)$ is a monotonically increasing function when $x \ge 0$, we have

$$\frac{w_k^{(2)}}{w_k^{(1)}} \ge \frac{(j+\frac{k+1}{i})\Delta - (j+\gamma')\Delta}{(j+\gamma')\Delta - (j+\frac{k}{i})\Delta} = \frac{\frac{k+1}{i} - \gamma'}{\gamma' - \frac{k}{i}}. \tag{240}$$

Therefore,

$$\int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) - \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}_{\gamma'}(dx) \tag{241}$$

$$= 2w_k a_k - 2\left(w_k^{(1)}a_0 + w_k^{(2)}a_0 e^{-\epsilon}\right) \tag{242}$$

$$= 2\left(w_k^{(1)} + w_k^{(2)}\right)a_k - 2\left(w_k^{(1)}a_0 + w_k^{(2)}a_0 e^{-\epsilon}\right) \tag{243}$$

$$= 2(a_k - a_0 e^{-\epsilon})w_k^{(2)} - 2(a_0 - a_k)w_k^{(1)}. \tag{244}$$

Since

$$\frac{a_k - a_0 e^{-\epsilon}}{a_0 - a_k} = \frac{\beta(a_0 - a_0 e^{-\epsilon})}{(1-\beta)(a_0 - a_0 e^{-\epsilon})} \tag{245}$$

$$= \frac{\beta}{1-\beta} \tag{246}$$

$$= \frac{\gamma' - \frac{k}{i}}{\frac{k+1}{i} - \gamma'} \tag{247}$$

$$\ge \frac{w_k^{(1)}}{w_k^{(2)}}, \tag{248}$$

we have

$$\int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) - \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}_{\gamma'}(dx) \tag{249}$$

$$= 2(a_k - a_0 e^{-\epsilon})w_k^{(2)} - 2(a_0 - a_k)w_k^{(1)} \tag{250}$$

$$\ge 0. \tag{251}$$

Therefore,

$$V^* = \inf_{\mathcal{P} \in \cup_{i=3}^{\infty} \mathcal{SP}_{i,step}} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) \tag{252}$$

$$\ge \inf_{\gamma \in [0,1]} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}_\gamma(dx). \tag{253}$$

We conclude

$$V^* = \inf_{\mathcal{P} \in \mathcal{SP}} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(dx) \tag{254}$$

$$= \inf_{\gamma \in [0,1]} \int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}_\gamma(dx) \tag{255}$$

$$= \inf_{\gamma \in [0,1]} \int_{x \in \mathbb{R}} \mathcal{L}(x)f_\gamma(x)dx. \tag{256}$$

This completes the proof of Theorem 4. ∎

APPENDIX C
PROOF OF THEOREM 5

Proof of Theorem 5: Recall $b \triangleq e^{-\epsilon}$, and $\mathcal{L}(x) = |x|$. We can compute $V(\mathcal{P}_\gamma)$ via

$$V(\mathcal{P}_\gamma) = \int_{x \in \mathbb{R}} |x| f^\gamma(x)dx \tag{257}$$

$$= 2\int_0^{+\infty} x f^\gamma(x)dx \tag{258}$$

$$= 2\sum_{k=0}^{+\infty}\left(\int_0^{\gamma\Delta}(x+k\Delta)a(\gamma)e^{-k\epsilon}dx\right.$$

$$\left. + \int_{\gamma\Delta}^{\Delta}(x+k\Delta)a(\gamma)e^{-\epsilon}e^{-k\epsilon}dx\right) \tag{259}$$

$$= 2\Delta^2 a(\gamma)$$

$$\times \sum_{k=0}^{+\infty} \left(e^{-k\epsilon} \frac{(k+\gamma)^2 - k^2}{2} + e^{-(k+1)\epsilon} \frac{(k+1)^2 - (k+\gamma)^2}{2} \right) \tag{260}$$

$$= 2\Delta^2 a(\gamma)$$

$$\times \sum_{k=0}^{+\infty} \left(e^{-k\epsilon} \frac{\gamma^2 + 2k\gamma}{2} + e^{-(k+1)\epsilon} \frac{2k+1-2k\gamma-\gamma^2}{2} \right) \tag{261}$$

$$= 2\Delta^2 a(\gamma) \sum_{k=0}^{+\infty} \left((b+(1-b)\gamma)k e^{-k\epsilon} + \frac{b+(1-b)\gamma^2}{2} e^{-k\epsilon} \right) \tag{262}$$

$$= 2\Delta^2 a(\gamma) \left((b+(1-b)\gamma) \frac{b}{(1-b)^2} + \frac{b+(1-b)\gamma^2}{2} \frac{1}{1-b} \right) \tag{263}$$

$$= 2\Delta^2 \frac{1-b}{2\Delta(b+(1-b)\gamma)}$$

$$\times \left((b+(1-b)\gamma) \frac{b}{(1-b)^2} + \frac{b+(1-b)\gamma^2}{2} \frac{1}{1-b} \right) \tag{264}$$

$$= \Delta \left(\frac{b}{1-b} + \frac{1}{2} \frac{b+(1-b)\gamma^2}{b+(1-b)\gamma} \right), \tag{265}$$

where in (263) we use the formulas

$$\sum_{k=1}^{+\infty} b^k = \frac{1}{1-b}, \tag{266}$$

$$\sum_{k=1}^{+\infty} k b^k = \frac{b}{(1-b)^2}. \tag{267}$$

Note that the first term $\frac{b}{1-b}$ is independent of γ. Define

$$g(\gamma) \triangleq \frac{b+(1-b)\gamma^2}{b+(1-b)\gamma}, \tag{268}$$

and thus to minimize $V(\mathcal{P}_\gamma)$ over $\gamma \in [0, 1]$, we only need to minimize $g(\gamma)$ over $\gamma \in [0, 1]$.

Since $\gamma \in [0, 1]$, $g(\gamma) \le 1$. Also note that $g(0) = g(1) = 1$. So the optimal γ^* which minimize $g(\gamma)$ lies in $(0, 1)$.

Compute the derivative of $g(\gamma)$ via

$$g'(\gamma) = \frac{2\gamma(1-b)(b+(1-b)\gamma) - (b+(1-b)\gamma^2)(1-b)}{(b+(1-b)\gamma)^2} \tag{269}$$

$$= (1-b) \frac{(1-b)\gamma^2 + 2b\gamma - b}{(b+(1-b)\gamma)^2}. \tag{270}$$

Set $g'(\gamma^*) = 0$ and we get

$$\gamma^* = \frac{\sqrt{b} - b}{1 - b} \tag{271}$$

$$= \frac{e^{-\frac{1}{2}\epsilon} - e^{-\epsilon}}{1 - e^{-\epsilon}} \tag{272}$$

$$= \frac{1}{1 + e^{\frac{\epsilon}{2}}}. \tag{273}$$

Therefore,

$$V(\mathcal{P}_{\gamma^*}) = \Delta \left(\frac{b}{1-b} + \frac{1}{2} \frac{b+(1-b)\gamma^{*2}}{b+(1-b)\gamma^*} \right) \tag{274}$$

$$= \Delta \frac{e^{\frac{\epsilon}{2}}}{e^\epsilon - 1}. \tag{275}$$

Due to Theorem 4, the minimum expectation of noise amplitude is $V(\mathcal{P}_{\gamma^*}) = \Delta \frac{e^{\frac{\epsilon}{2}}}{e^\epsilon - 1}$. ∎

APPENDIX D
PROOF OF THEOREM 7

Proof of Theorem 7: Recall $b \triangleq e^{-\epsilon}$. Then we compute $V(\mathcal{P}_\gamma)$ for the cost function $\mathcal{L}(x) = x^2$ via

$$V(\mathcal{P}_\gamma) = \int_{x \in \mathbb{R}} x^2 f^\gamma(x) dx \tag{276}$$

$$= \Delta^2 \left(\frac{b^2+b}{(1-b)^2} + \frac{b+(1-b)\gamma^2}{b+(1-b)\gamma} \frac{b}{1-b} + \frac{1}{3} \frac{b+(1-b)\gamma^3}{b+(1-b)\gamma} \right). \tag{277}$$

Note that the first term $\frac{b^2+b}{(1-b)^2}$ is independent of γ. Define

$$h(\gamma) \triangleq \frac{b+(1-b)\gamma^2}{b+(1-b)\gamma} \frac{b}{1-b} + \frac{1}{3} \frac{b+(1-b)\gamma^3}{b+(1-b)\gamma} \tag{278}$$

$$= \frac{\frac{(1-b)\gamma^3}{3} + b\gamma^2 + \frac{b^2}{1-b} + \frac{b}{3}}{b+(1-b)\gamma}, \tag{279}$$

and thus to minimize $V(\mathcal{P}_\gamma)$ over $\gamma \in [0, 1]$, we only need to minimize $h(\gamma)$ over $\gamma \in [0, 1]$.

Since $\gamma \in [0, 1]$, $h(\gamma) \le \frac{b}{1-b} + \frac{1}{3}$. Also note that $h(0) = h(1) = \frac{b}{1-b} + \frac{1}{3}$. So the optimal γ^* which minimize $h(\gamma)$ lies in $(0, 1)$.

The derivative of $h(\gamma)$ is

$$h'(\gamma) = \frac{\frac{2}{3}(1-b)^2\gamma^3 + 2b(1-b)\gamma^2 + 2b^2\gamma - \frac{2b^2+b}{3}}{(b+(1-b)\gamma)^2}. \tag{280}$$

Set $h'(\gamma^*) = 0$ and we get

$$\frac{2}{3}(1-b)^2\gamma^{*3} + 2b(1-b)\gamma^{*2} + 2b^2\gamma^* - \frac{2b^2+b}{3} = 0. \tag{281}$$

Therefore, the optimal γ^* is the real-valued root of the cubic equation (281), which is

$$\gamma^* = -\frac{b}{1-b} + \frac{(b-2b^2+2b^4-b^5)^{1/3}}{2^{1/3}(1-b)^2}. \tag{282}$$

We plot γ^* as a function of b in Figure 4, and we can see $\gamma^* \to \frac{1}{2}$ as $\epsilon \to 0$, and $\gamma^* \to 0$ as $\epsilon \to +\infty$. This also holds in the case $\mathcal{L}(x) = |x|$.

Plug (282) into (277), and we get the minimum noise power

$$V(\mathcal{P}_{\gamma^*}) = \Delta^2 \frac{2^{-2/3} b^{2/3} (1+b)^{2/3} + b}{(1-b)^2}. \tag{283}$$

Due to Theorem 4, the minimum expectation of noise power is $V(\mathcal{P}_{\gamma^*}) = \Delta^2 \frac{2^{-2/3} b^{2/3} (1+b)^{2/3} + b}{(1-b)^2}$. ∎

APPENDIX E
PROOF OF THEOREM 9

Proof of Theorem 9: Let $n = m + 1$, and define

$$c_i \triangleq \sum_{k=0}^{+\infty} b^k k^i, \qquad (284)$$

for nonnegative integer i.

First we compute $V(\mathcal{P}_\gamma)$ via

$$V(\mathcal{P}_\gamma) = 2\sum_{k=0}^{+\infty} \left(\int_0^{\gamma\Delta} (x + k\Delta)^m a(\gamma) e^{-k\epsilon} dx \right.$$

$$\left. + \int_{\gamma\Delta}^{\Delta} (x + k\Delta)^m a(\gamma) e^{-(k+1)\epsilon} dx \right)$$

$$= 2a(\gamma)\Delta^{m+1} \sum_{k=0}^{+\infty} \left(b^k \frac{(k+\gamma)^{m+1} - k^{m+1}}{m+1} \right.$$

$$\left. + b^{k+1} \frac{(k+1)^{m+1} - (k+\gamma)^{m+1}}{m+1} \right)$$

$$= 2\Delta^n a(\gamma) \sum_{k=0}^{+\infty} \left(b^k \frac{\sum_{i=1}^n \binom{n}{i}\gamma^i k^{n-i}}{n} \right.$$

$$\left. + bb^k \frac{\sum_{i=1}^n \binom{n}{i}(1-\gamma^i)k^{n-i}}{n} \right)$$

$$= 2\Delta^n a(\gamma) \left(\sum_{i=1}^n \frac{\binom{n}{i}\gamma^i c_{n-i}}{n} + b\sum_{i=1}^n \frac{\binom{n}{i}(1-\gamma^i)c_{n-i}}{n} \right)$$

$$= 2\Delta^n a(\gamma) \sum_{i=1}^n \frac{\binom{n}{i}c_{n-i}(\gamma^i(1-b)+b)}{n}$$

$$= \frac{2\Delta^n(1-b)}{2\Delta n} \frac{\sum_{i=1}^n \binom{n}{i}c_{n-i}(\gamma^i(1-b)+b)}{\gamma(1-b)+b}.$$

Let $h_i(\gamma) \triangleq \frac{\gamma^i(1-b)+b}{\gamma(1-b)+b}$ for $i \geq 2$. Since $h_i(0) = h_i(1) = 1$ and $h_i(\gamma) < 1$ for $\gamma \in (0, 1)$, $h_i(\gamma)$ achieves the minimum value in the open interval $(0, 1)$.

Therefore, if we define $h(\gamma) \triangleq \frac{\sum_{i=1}^n \binom{n}{i}c_{n-i}(\gamma^i(1-b)+b)}{\gamma(1-b)+b}$, the optimal $\gamma^* \in [0, 1]$, which minimizes $V(\mathcal{P}_\gamma)$, should satisfy

$$h'(\gamma^*) = 0, \qquad (285)$$

where $h'(\cdot)$ denotes the first order derivative of $h(\cdot)$.

It is straightforward to derive the expression for $h'(\cdot)$:

$$h'(\gamma) = \frac{A}{B}, \qquad (286)$$

where

$$A := \sum_{i=1}^n \binom{n}{i}c_{n-i}(i-1)\gamma^i(1-b)^2$$

$$+ \sum_{i=1}^n \binom{n}{i}c_{n-i}i\gamma^{i-1}(1-b)b - \sum_{i=1}^n \binom{n}{i}c_{n-i}b(1-b),$$

and

$$B := (\gamma(1-b)+b)^2.$$

Therefore, γ^* should make the numerator of (286), i.e., A be zero, i.e., γ^* satisfies

$$A = 0. \qquad (287)$$

Since

$$\sum_{i=1}^n \binom{n}{i}c_{n-i}(i-1)\gamma^i(1-b)^2$$

$$+ \sum_{i=1}^n \binom{n}{i}c_{n-i}i\gamma^{i-1}(1-b)b - \sum_{i=1}^n \binom{n}{i}c_{n-i}b(1-b)$$

$$= \sum_{i=1}^n \binom{n}{i}c_{n-i}(i-1)\gamma^i(1-b)^2$$

$$+ \sum_{i=0}^{n-1} \binom{n}{i+1}c_{n-(i+1)}(i+1)\gamma^i(1-b)b$$

$$- \sum_{i=1}^n \binom{n}{i}c_{n-i}b(1-b)$$

$$= c_0(n-1)\gamma^n(1-b)^2 + \sum_{i=1}^{n-1} \left(\binom{n}{i}c_{n-i}(i-1)(1-b)^2 \right.$$

$$\left. + \binom{n}{i+1}c_{n-(i+1)}(i+1)(1-b)b \right)\gamma^i$$

$$+ nc_{n-1}(1-b)b - \sum_{i=1}^n \binom{n}{i}c_{n-i}b(1-b)$$

$$= c_0(n-1)\gamma^n(1-b)^2 + \gamma^i \sum_{i=1}^{n-1} \binom{n}{i}c_{n-i}(i-1)(1-b)^2$$

$$+ \gamma^i \sum_{i=1}^{n-1} \binom{n}{i+1}c_{n-(i+1)}(i+1)(1-b)b$$

$$- \sum_{i=2}^n \binom{n}{i}c_{n-i}b(1-b),$$

γ^* satisfies

$$c_0(n-1)\gamma^{*n}(1-b)^2 + \gamma^{*i} \sum_{i=1}^{n-1} \binom{n}{i}c_{n-i}(i-1)(1-b)^2$$

$$+ \gamma^{*i} \sum_{i=1}^{n-1} \binom{n}{i+1}c_{n-(i+1)}(i+1)(1-b)b$$

$$- \sum_{i=2}^n \binom{n}{i}c_{n-i}b(1-b)$$

$$= 0 \qquad (288)$$

We can derive the asymptotic properties of γ^* from (288). Before deriving the properties of γ^*, we first study the asymptotic properties of c_i, which are functions of b.

There are closed-form formulas for c_i $(i = 0, 1, 2, 3)$:

$$c_0 = \sum_{k=0}^{+\infty} b^k = \frac{1}{1-b}, \qquad (289)$$

$$c_1 = \sum_{k=0}^{+\infty} b^k k = \frac{b}{(1-b)^2}, \qquad (290)$$

$$c_2 = \sum_{k=0}^{+\infty} b^k k^2 = \frac{b^2 + b}{(1-b)^3}, \tag{291}$$

$$c_3 = \sum_{k=0}^{+\infty} b^k k^3 = \frac{b^3 + 4b^2 + b}{(1-b)^4}. \tag{292}$$

In general, for $i \geq 1$,

$$c_{i+1} = \sum_{k=0}^{+\infty} b^k k^{i+1} = \sum_{k=1}^{+\infty} b^k k^{i+1} = b + \sum_{k=1}^{+\infty} b^{k+1}(k+1)^{i+1}, \tag{293}$$

$$bc_{i+1} = \sum_{k=0}^{+\infty} b^{k+1} k^{i+1} = \sum_{k=1}^{+\infty} b^{k+1} k^{i+1}. \tag{294}$$

Therefore,

$$c_{i+1} - bc_{i+1} = b + \sum_{k=1}^{+\infty} b^{k+1}((k+1)^{i+1} - k^{i+1}) \tag{295}$$

$$= b + \sum_{k=1}^{+\infty} b^{k+1} \sum_{j=0}^{i} \binom{i+1}{j} k^j \tag{296}$$

$$= b + b \sum_{j=0}^{i} \binom{i+1}{j} \sum_{k=1}^{+\infty} k^j b^k \tag{297}$$

$$= b + b\left(\frac{b}{1-b} + \sum_{j=1}^{i} \binom{i+1}{j} c_j\right) \tag{298}$$

$$= \frac{b}{1-b} + b \sum_{j=1}^{i} \binom{i+1}{j} c_j, \tag{299}$$

and thus

$$c_{i+1} = \frac{b}{(1-b)^2} + \frac{b}{1-b} \sum_{j=1}^{i} \binom{i+1}{j} c_j. \tag{300}$$

From (300), by induction we can easily prove that
- as $b \to 0$, $c_i \to 0$, $\forall i \geq 1$;
- as $b \to 1$, $\forall i \geq 0$, $c_i \to +\infty$, $c_i = \Omega(\frac{i!}{(1-b)^{i+1}})$ and

$$\lim_{b \to 1} \frac{c_{i+1}}{c_i}(1-b) = i+1. \tag{301}$$

As $b \to 0$, since $c_i \to 0$ for $i \geq 1$ and $c_0 = 1$, the last two terms of (288) go to zero, and thus from (288) we can see that γ^* goes to zero as well.

As $b \to 1$, since $c_i = \Omega(\frac{1}{(1-b)^{i+1}})$ and γ^* is bounded by 1, the first term of (288) goes to zero, and the dominated terms in (288) are

$$\binom{n}{2} c_{n-2} 2(1-b)b\gamma^* - \binom{n}{2} c_{n-2} b(1-b) = 0. \tag{302}$$

Thus, in the limit we have $\gamma^* = \frac{1}{2}$. Therefore, as $b \to 1$, $\gamma^* \to \frac{1}{2}$.

This completes the proof. ∎

APPENDIX F
PROOF OF THEOREM 12 AND THEOREM 13

In this section, we prove Theorem 12 and Theorem 13, which give the optimal noise-adding mechanisms in the discrete setting.

A. Outline of Proof

The proof technique is very similar to the proof in the continuous settings in Appendix B. The proof consists of 5 steps in total, and in each step we narrow down the set of probability distributions where the optimal probability distribution should lie in:

- Step 1 proves that we only need to consider probability mass functions which are monotonically increasing for $i \leq 0$ and monotonically decreasing for $i \geq 0$.
- Step 2 proves that we only need to consider symmetric probability mass functions.
- Step 3 proves that we only need to consider symmetric probability mass functions which have periodic and geometric decay for $i \geq 0$, and this proves Theorem 12.
- Step 4 and Step 5 prove that the optimal probability mass function over the interval $[0, \Delta)$ is a discrete step function, and they conclude the proof of Theorem 13.

B. Step 1

Recall $\mathcal{SP}$ denotes the set of all probability mass functions which satisfy the ϵ-differential privacy constraint (79). Define

$$V^* \triangleq \inf_{\mathcal{P} \in \mathcal{SP}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}(i). \tag{303}$$

First we prove that we only need to consider probability mass functions which are monotonically increasing for $i \leq 0$ and monotonically decreasing for $i \geq 0$.

Define

$$\mathcal{SP}_{mono} \triangleq \{\mathcal{P} \in \mathcal{SP} | \mathcal{P}(i) \leq \mathcal{P}(j), \mathcal{P}(m) \geq \mathcal{P}(n),$$
$$\forall i \leq j \leq 0, 0 \leq m \leq n\}.$$

Lemma 29:

$$V^* = \inf_{\mathcal{P} \in \mathcal{SP}_{mono}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}(i). \tag{304}$$

Proof: We will prove that given a probability mass function $\mathcal{P}_a \in \mathcal{SP}$, we can construct a new probability mass function $\mathcal{P}_b \in \mathcal{SP}_{mono}$ such that

$$\sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_a(i) \geq \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_b(i). \tag{305}$$

Given $\mathcal{P}_a \in \mathcal{SP}$, consider the sequence $sa = \{\mathcal{P}_a(0), \mathcal{P}_a(1), \mathcal{P}_a(-1), \mathcal{P}_a(2), \mathcal{P}_a(-2), \ldots\}$. Use the same argument in Lemma 19 and we can show $\mathcal{P}_a(i) > 0$, $\forall i \in \mathbb{Z}$. Let the sequence $sb = \{b_0, b_1, b_{-1}, b_2, b_{-2}, \ldots\}$ be a permutation of the sequence sa in descending order. Since $\sum_{i=-\infty}^{+\infty} \mathcal{P}_a(i) = 1$, $\lim_{i \to -\infty} \mathcal{P}_a(i) = \lim_{i \to +\infty} \mathcal{P}_a(i) = 0$, and thus sb is well defined. Let π be the corresponding permutation mapping, i.e., $\pi : \mathbb{Z} \to \mathbb{Z}$, and

$$b_i = \mathcal{P}_a(\pi(i)). \tag{306}$$

Since $\mathcal{L}(\cdot)$ is a symmetric function and monotonically decreasing for $i \geq 0$, we have

$$\mathcal{L}(0) \leq \mathcal{L}(1) \leq \mathcal{L}(-1) \leq \mathcal{L}(2) \leq \mathcal{L}(-2) \leq \cdots \leq \mathcal{L}(i)$$
$$\leq \mathcal{L}(-i) \leq \mathcal{L}(i+1) \leq \mathcal{L}(-(i+1)) \leq \cdots.$$

Algorithm 3 Identifying at least 2Δ elements in the sequence sb which are in the range $[b_i e^{-\epsilon}, b_i]$

$j_R^* \leftarrow i$
while there exists some j which appears before i in the sequence $\{0, 1, -1, 2, -2, \ldots\}$ and $\pi(j) \in [\pi(j_R^*) + 1, \pi(j_R^*) + \Delta]$ **do**
    $j_R^* \leftarrow j$
end while
$j_L^* \leftarrow i$
while there exists some j which appears before i in the sequence $\{0, 1, -1, 2, -2, \ldots\}$ and $\pi(j) \in [\pi(j_L^*) - \Delta, \pi(j_L^*) - 1]$ **do**
    $j_L^* \leftarrow j$
end while
Output j_R^* and j_L^*.

Therefore, if we define a probability mass function $\mathcal{P}_b$ with

$$\mathcal{P}_b(i) = b_i, \quad \forall i \in \mathbb{Z}, \tag{307}$$

then

$$\sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_a(i) \geq \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_b(i). \tag{308}$$

Next, we only need to prove $\mathcal{P}_b \in \mathcal{SP}_{\text{mono}}$, i.e., we need to show that $\mathcal{P}_b$ satisfies the differential privacy constraint (79).

Due to the way how we construct the sequence sb, we have

$$b_0 \geq b_1 \geq b_2 \geq b_3 \geq \cdots, \tag{309}$$
$$b_0 \geq b_{-1} \geq b_{-2} \geq b_{-3} \geq \cdots. \tag{310}$$

Therefore, it is both sufficient and necessary to prove that

$$\frac{b_i}{b_{i+\Delta}} \leq e^\epsilon, \quad \forall i \geq 0, \tag{311}$$
$$\frac{b_i}{b_{i-\Delta}} \leq e^\epsilon, \quad \forall i \leq 0. \tag{312}$$

Since $\mathcal{P}_a \in \mathcal{SP}$, $\forall i \in \{\pi(0) - \Delta, \pi(0) - \Delta + 1, \pi(0) - \Delta + 2, \ldots, \pi(0) + \Delta\}$,

$$\frac{\mathcal{P}_a(\pi(0))}{\mathcal{P}_a(i)} \leq e^\epsilon. \tag{313}$$

Therefore, in the sequence sb there exist at least 2Δ elements which are no smaller than $b_0 e^{-\epsilon}$. Since $b_{-\Delta}$ and b_Δ are the 2Δth and $(2\Delta - 1)$th largest elements in the sequence sb other than b_0, we have $\frac{b_0}{b_{-\Delta}} \leq e^\epsilon$ and $\frac{b_0}{b_\Delta} \leq e^\epsilon$.

In general, given $i \in \mathbb{Z}$, we can use Algorithm 3 to find at least 2Δ elements in the sequence sb which are no bigger than b_i and no smaller than $b_i e^{-\epsilon}$.

More precisely, given $i \in \mathbb{Z}$, let j_R^* and j_L^* be the output of Algorithm 3. Note that since the while loops in Algorithm 3 can take only at most $2(|i|+1)$ steps, the algorithm will always terminate. For all integers $j \in [\pi(j_L^*) - \Delta, \pi(j_L^*) - 1]$, $\mathcal{P}_a(j)$ is no bigger than b_i and is no smaller than $\mathcal{P}_a(j_L^*)e^{-\epsilon}$; and for all integers $j \in [\pi(j_R^*) + 1, \pi(j_R^*) + \Delta]$, $\mathcal{P}_a(j)$ is no bigger than b_i and is no smaller than $\mathcal{P}_a(j_R^*)e^{-\epsilon}$. Since $\mathcal{P}_a(j_R^*), \mathcal{P}_a(j_L^*) \geq b_i$, for all $j \in [\pi(j_L^*) - \Delta, \pi(j_L^*) - 1] \cup [\pi(j_R^*) + 1, \pi(j_R^*) + \Delta]$, $\mathcal{P}_a(j)$ is no bigger than b_i and is no smaller than $b_i e^{-\epsilon}$.

Therefore, there exist at least 2Δ elements in the sequence sb which are no bigger than b_i and no smaller than $b_i e^{-\epsilon}$.

If $i \leq 0$, then $b_{i-\Delta}$ is the 2Δth largest element in the sequence sb which is no bigger than b_i and no smaller than $b_i e^{-\epsilon}$; and if $i \geq 0$, then $b_{i+\Delta}$ is the $(2\Delta - 1)$th largest element in the sequence sb which is no bigger than b_i and no smaller than $b_i e^{-\epsilon}$. Therefore, we have

$$\frac{b_i}{b_{i+\Delta}} \leq e^\epsilon, \quad \forall i \geq 0, \tag{314}$$
$$\frac{b_i}{b_{i-\Delta}} \leq e^\epsilon, \quad \forall i \leq 0. \tag{315}$$

This completes the proof of Lemma 29. ∎

C. Step 2

Next we prove that we only need to consider symmetric probability mass functions which are monotonically decreasing when $i \geq 0$.

Define

$$\mathcal{SP}_{\text{sym}} \triangleq \{\mathcal{P} \in \mathcal{SP}_{\text{mono}} | \mathcal{P}(i) = \mathcal{P}(-i), \forall i \in \mathbb{Z}\}. \tag{316}$$

Lemma 30:

$$V^* = \inf_{\mathcal{P} \in \mathcal{SP}_{\text{sym}}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}(i). \tag{317}$$

Proof: The proof is essentially the same as the proof of Lemma 17.

Given $\mathcal{P}_a \in \mathcal{SP}_{\text{mono}}$, define a new probability mass function $\mathcal{P}_b$ with

$$\mathcal{P}_b(i) \triangleq \frac{\mathcal{P}_a(i) + \mathcal{P}_a(-i)}{2}, \quad \forall i \in \mathbb{Z}. \tag{318}$$

It is easy to see $\mathcal{P}_b$ is a valid probability mass function and symmetric. Since the cost function $\mathcal{L}(\cdot)$ is symmetric,

$$\sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_a(i) = \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_b(i). \tag{319}$$

Next we show that $\mathcal{P}_b$ also satisfies the differential privacy constraint (79). For any $i \in \mathbb{Z}$ and $|d| \leq \Delta$, since $\mathcal{P}_a(i) \leq e^\epsilon \mathcal{P}_a(i + d)$ and $\mathcal{P}_a(-i) \leq e^\epsilon \mathcal{P}_a(-i - d)$, we have

$$\mathcal{P}_b(i) = \frac{\mathcal{P}_a(i) + \mathcal{P}_a(-i)}{2} \tag{320}$$
$$\leq \frac{e^\epsilon \mathcal{P}_a(i + d) + e^\epsilon \mathcal{P}_a(-i - d)}{2} \tag{321}$$
$$= e^\epsilon \mathcal{P}_b(i + d). \tag{322}$$

Therefore, $\mathcal{P}_b$ satisfies (79).

Finally, for any $0 \leq i \leq j$,

$$\mathcal{P}_b(i) = \frac{\mathcal{P}_a(i) + \mathcal{P}_a(-i)}{2} \tag{323}$$
$$\geq \frac{\mathcal{P}_a(j) + \mathcal{P}_a(-j)}{2} \tag{324}$$
$$= \mathcal{P}_b(j). \tag{325}$$

So $\mathcal{P}_b \in \mathcal{SP}_{\text{mono}}$, and thus $\mathcal{P}_b \in \mathcal{SP}_{\text{sym}}$. We conclude

$$V^* = \inf_{\mathcal{P} \in \mathcal{SP}_{\text{sym}}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}(i). \tag{326}$$

∎

D. Step 3

Next we show that among all symmetric and monotonically decreasing (for $i \geq 0$) probability mass function, we only need to consider those which are periodically and geometrically decaying.

More precisely, define

$$\mathcal{SP}_{\mathrm{pd}} \triangleq \{\mathcal{P} \in \mathcal{SP}_{\mathrm{sym}} | \frac{\mathcal{P}(i)}{\mathcal{P}(i+\Delta)} = e^\epsilon, \forall i \in \mathbb{N}\}. \quad (327)$$

Then

Lemma 31:

$$V^* = \inf_{\mathcal{P} \in \mathcal{SP}_{pd}} V(\mathcal{P}). \quad (328)$$

Proof: Due to Lemma 30, we only need to consider probability mass functions which are symmetric and monotonically decreasing for $i \geq 0$.

We first show that given $\mathcal{P}_a \in \mathcal{SP}_{\mathrm{sym}}$, if $\frac{\mathcal{P}_a 0}{\mathcal{P}_a \Delta} < e^\epsilon$, then we can construct a probability mass function $\mathcal{P}_b \in \mathcal{SP}_{\mathrm{sym}}$ such that $\frac{\mathcal{P}_b 0}{\mathcal{P}_b \Delta} = e^\epsilon$ and

$$V(\mathcal{P}_a) \geq V(\mathcal{P}_b). \quad (329)$$

Since $\mathcal{P}_a$ is symmetric,

$$V(\mathcal{P}_a) = \mathcal{L}(0)\mathcal{P}_a(0) + 2\sum_{i=1}^{+\infty} \mathcal{L}(i)\mathcal{P}_a(i). \quad (330)$$

Suppose $\frac{\mathcal{P}_a 0}{\mathcal{P}_a \Delta} < e^\epsilon$, then define a new symmetric probability mass function $\mathcal{P}_b$ with

$$\mathcal{P}_b(0) \triangleq (1+\delta)\mathcal{P}_a(0), \quad (331)$$
$$\mathcal{P}_b(i) \triangleq (1-\delta')\mathcal{P}_a(i), \quad \forall i \in \mathbb{Z}\backslash\{0\}, \quad (332)$$

where

$$\delta = \frac{e^\epsilon \frac{\mathcal{P}_a(\Delta)}{\mathcal{P}_a(0)} - 1}{1 + e^\epsilon \frac{\mathcal{P}_a(\Delta)}{1-\mathcal{P}_a(0)}} > 0, \quad (333)$$

$$\delta' = \frac{e^\epsilon \frac{\mathcal{P}_a(\Delta)}{\mathcal{P}_a(0)} - 1}{\frac{1}{\mathcal{P}_a(0)} + e^\epsilon \frac{\mathcal{P}_a(\Delta)}{\mathcal{P}_a(0)} - 1} > 0, \quad (334)$$

so that $\frac{\mathcal{P}_b(0)}{\mathcal{P}_b(\Delta)} = e^\epsilon$.

It is easy to see $\mathcal{P}_b \in \mathcal{SP}_{\mathrm{sym}}$, and

$$V(\mathcal{P}_b) - V(\mathcal{P}_a) \quad (335)$$

$$= \delta\mathcal{L}(0)\mathcal{P}_a(0) - 2\delta'\sum_{i=1}^{+\infty} \mathcal{L}(i)\mathcal{P}_a(i) \quad (336)$$

$$\leq \delta\mathcal{L}(0)\mathcal{P}_a(0) - 2\delta'\sum_{i=1}^{+\infty} \mathcal{L}(0)\mathcal{P}_a(i) \quad (337)$$

$$\leq \delta\mathcal{L}(0)\mathcal{P}_a(0) - \delta'\mathcal{L}(0)(1 - \mathcal{P}_a(0)) \quad (338)$$

$$= 0. \quad (339)$$

Therefore, we only need to consider $\mathcal{P} \in \mathcal{SP}_{\mathrm{sym}}$ satisfying $\frac{\mathcal{P}(0)}{\mathcal{P}(\Delta)} = e^\epsilon$.

By using the same argument as in the proof of Lemma 24, one can conclude that we only need to consider $\mathcal{P} \in \mathcal{SP}_{\mathrm{sym}}$ satisfying

$$\frac{\mathcal{P}(i)}{\mathcal{P}(i+\Delta)} = e^\epsilon, \quad \forall i \in \mathbb{N}. \quad (340)$$

Therefore, $V^* = \inf_{\mathcal{P} \in \mathcal{SP}_{\mathrm{pd}}} V(\mathcal{P})$. ∎

Proof of Theorem 12: In the case that $\Delta = 1$, due to Lemma 31, the symmetry property and (340) completely characterize the optimal noise probability mass function, which is the geometric mechanism. ∎

E. Step 4

Due to Lemma 31, the optimal probability mass function $\mathcal{P}$ is completely characterized by $\mathcal{P}(0), \mathcal{P}(1), \ldots, \mathcal{P}(\Delta-1)$. Next we derive the properties of optimal probability mass function in the domain $\{0, 1, 2, \ldots, \Delta-1\}$.

Since Lemma 31 solves the case $\Delta = 1$, in the remaining of this section, we assume $\Delta \geq 2$.

Define

$\mathcal{SP}_{\mathrm{step}_\lambda} \triangleq \{\mathcal{P} \in \mathcal{SP}_{\mathrm{pd}} \mid \exists k \in \{0, 1, \ldots, \Delta-2\}$, such that

$$\mathcal{P}(i) = \mathcal{P}(0), \quad \forall i \in \{0, 1, \ldots, k\},$$

$$\mathcal{P}(j) = \lambda\mathcal{P}(0), \forall j \in \{k+1, k+2, \ldots, \Delta-1\}\}.$$

Lemma 32:

$$V^* = \inf_{\mathcal{P} \in \cup_{\lambda \in [e^{-\epsilon}, 1]} \mathcal{SP}_{\mathrm{step}_\lambda}} V(\mathcal{P}). \quad (341)$$

Proof: If $\Delta = 2$, then for any $\mathcal{P} \in \mathcal{SP}_{\mathrm{pd}}$, we can set $k = 0$, and $\mathcal{P} \in \mathcal{SP}_{\mathrm{step}\frac{\mathcal{P}(\Delta-1)}{\mathcal{P}(0)}}$. Therefore, Lemma 32 holds for $\Delta = 2$.

Assume $\Delta \geq 3$. First, we prove that we only need to consider probability mass function $\mathcal{P} \in \mathcal{SP}_{\mathrm{pd}}$ such that there exists $k \in \{1, 2, \ldots, \Delta-2\}$ with

$$\mathcal{P}(i) = \mathcal{P}(0), \quad \forall i \in \{0, 1, \ldots, k-1\} \quad (342)$$
$$\mathcal{P}(j) = \mathcal{P}(\Delta-1), \quad \forall i \in \{k+1, k+2, \ldots, \Delta-1\}. \quad (343)$$

More precisely, let $\mathcal{P}_a \in \mathcal{SP}_{\mathrm{pd}}$, we can construct a probability mass function $\mathcal{P}_b \in \mathcal{SP}_{\mathrm{pd}}$ such that there exists k satisfying (342) and (343), and $V(\mathcal{P}_b) \geq V(\mathcal{P}_a)$.

The proof technique is very similar to proof of Lemma 27. Suppose there does not exists such k for $\mathcal{P}_a$, then let k_1 be the smallest integer in $\{1, 2, \ldots, \Delta-1\}$ such that

$$\mathcal{P}_a(k_1) \neq \mathcal{P}_a(0), \quad (344)$$

and let k_2 be the biggest integer in $\{0, 1, \ldots, \Delta-2\}$ such that

$$\mathcal{P}_a(k_2) \neq \mathcal{P}_a(\Delta-1). \quad (345)$$

It is easy to see that $k_1 < k_2$, and $k_1 \neq 0$. Then we can increase $\mathcal{P}_a(k_1)$ and decrease $\mathcal{P}_a(k_2)$ simultaneously by the same amount to derive a new probability mass function $\mathcal{P}_b \in \mathcal{SP}_{\mathrm{pd}}$ with smaller cost. Indeed, if

$$\mathcal{P}_a(0) - \mathcal{P}_a(k_1) \leq \mathcal{P}_a(k_2) - \mathcal{P}_a(\Delta-1), \quad (346)$$

then consider a probability mass function $\mathcal{P}_b \in \mathcal{SP}_{\mathrm{pd}}$ with

$$\mathcal{P}_b(i) = \mathcal{P}_a(0), \quad \forall 0 \leq i \leq k_1, \quad (347)$$

$$\mathcal{P}_b(i) = \mathcal{P}_a(i), \quad \forall k_1 < i < k_2, \quad (348)$$

$$\mathcal{P}_b(k_2) = \mathcal{P}_a(k_2) - (\mathcal{P}_a(0) - \mathcal{P}_a(k_1)), \quad (349)$$

$$\mathcal{P}_b(i) = \mathcal{P}_a(i), \quad \forall k_2 < i \leq \Delta-1. \quad (350)$$

Define

$$w_0 \triangleq \mathcal{L}(0) + 2 \sum_{k=1}^{\infty} \mathcal{L}(k\Delta) e^{-k\epsilon}, \tag{351}$$

$$w_i \triangleq 2 \sum_{k=0}^{\infty} \mathcal{L}(i + k\Delta) e^{-k\epsilon}, \quad \forall i \in \{1, 2, \ldots, \Delta - 1\}. \tag{352}$$

Note that since $\mathcal{L}(\cdot)$ is a monotonically decreasing function when $i \geq 0$, we have $w_0 \leq w_1 \leq \cdots \leq w_{\Delta-1}$.

Then we can verify that $V(\mathcal{P}_b) \leq V(\mathcal{P}_a)$ via

$$V(\mathcal{P}_b) - V(\mathcal{P}_a) \tag{353}$$

$$= \sum_{i=0}^{\Delta-1} \mathcal{P}_b(i) w_i - \sum_{i=0}^{\Delta-1} \mathcal{P}_a(i) w_i \tag{354}$$

$$= (\mathcal{P}_a(0) - \mathcal{P}_a(k_1))(w_{k_1} - w_{k_2}) \tag{355}$$

$$\leq 0. \tag{356}$$

If

$$\mathcal{P}_a(0) - \mathcal{P}_a(k_1) \geq \mathcal{P}_a(k_2) - \mathcal{P}_a(\Delta - 1), \tag{357}$$

then we can define $\mathcal{P}_b \in \mathcal{SP}_{\text{pd}}$ by setting

$$\mathcal{P}_b(i) = \mathcal{P}_a(0), \quad \forall 0 \leq i < k_1, \tag{358}$$

$$\mathcal{P}_b(k_1) = \mathcal{P}_a(k_1) + (\mathcal{P}_a(k_2) - \mathcal{P}_a(\Delta - 1)), \tag{359}$$

$$\mathcal{P}_b(i) = \mathcal{P}_a(i), \quad \forall k_1 < i < k_2, \tag{360}$$

$$\mathcal{P}_b(i) = \mathcal{P}_a(\Delta - 1), \quad \forall k_2 \leq i \leq \Delta - 1. \tag{361}$$

And similarly, we have

$$V(\mathcal{P}_b) - V(\mathcal{P}_a) = (\mathcal{P}_a(k_2) - \mathcal{P}_a(\Delta - 1))(w_{k_1} - w_{k_2}) \leq 0. \tag{362}$$

Therefore, continue in this way, and finally we will obtain a probability mass function $\mathcal{P}_b \in \mathcal{SP}_{\text{pd}}$ such that there exists k to satisfy (342) and (343) and $V(\mathcal{P}_b) \leq V(\mathcal{P}_a)$.

From the above argument, we can see that in the optimal solution $\mathcal{P}^* \in \mathcal{SP}_{\text{pd}}$, the probability mass function can only take at most three distinct values for all $i \in \{0, 1, \ldots, \Delta - 1\}$, which are $\mathcal{P}^*(0), \mathcal{P}^*(k)$ and $\mathcal{P}^*(\Delta - 1)$. Next we show that indeed either $\mathcal{P}^*(k) = \mathcal{P}^*(0)$ and $\mathcal{P}^*(k) = \mathcal{P}^*(\Delta - 1)$, and this will complete the proof of Lemma 32.

The optimal probability mass function $\mathcal{P} \in \mathcal{SP}_{\text{pd}}$ can be specified by three parameters $\mathcal{P}(0), \lambda \in [e^{-\epsilon}, 1], k \in \{1, 2, \ldots, \Delta - 2\}$ and $\mathcal{P}(k)$. We will show that when k and λ are fixed, to minimize the cost, we have either $\mathcal{P}(k) = \mathcal{P}(0)$ or $\mathcal{P}(k) = \mathcal{P}(\Delta - 1) = \lambda \mathcal{P}(0)$.

Since $\sum_{i=-\infty}^{+\infty} \mathcal{P}(i) = 1$,

$$2 \frac{k\mathcal{P}(0) + \mathcal{P}(k) + (\Delta - k - 1)\lambda\mathcal{P}(0)}{1 - b} - \mathcal{P}(0) = 1, \tag{363}$$

and thus $\mathcal{P}(k) = \frac{(1+\mathcal{P}(0))(1-b)-2\mathcal{P}(0)k-2\lambda\mathcal{P}(0)(\Delta-k-1)}{2}$.

The cost for $\mathcal{P}$ is

$$V(\mathcal{P}) = \mathcal{P}(0) \sum_{i=0}^{k-1} w_i + \mathcal{P}(\Delta - 1) \sum_{i=k+1}^{\Delta-1} w_i + \mathcal{P}(k) w_k \tag{364}$$

$$= \mathcal{P}(0) \sum_{i=0}^{k-1} w_i + \lambda \mathcal{P}(0) \sum_{i=k+1}^{\Delta-1} w_i \tag{365}$$

$$+ \left(\frac{(1+\mathcal{P}(0))(1-b)-2\mathcal{P}(0)k-2\lambda\mathcal{P}(0)(\Delta-k-1)}{2} \right) w_k, \tag{366}$$

which is a linear function of the parameter $\mathcal{P}(0)$.

Since $\mathcal{P}(k) \geq \lambda\mathcal{P}(0)$ and $\mathcal{P}(k) \leq \mathcal{P}(0)$, we have

$$2 \frac{k\mathcal{P}(0) + \mathcal{P}(k) + (\Delta - k - 1)\lambda\mathcal{P}(0)}{1 - b} - \mathcal{P}(0) = 1$$

$$\leq 2 \frac{k\mathcal{P}(0) + \mathcal{P}(0) + (\Delta - k - 1)\lambda\mathcal{P}(0)}{1 - b} - \mathcal{P}(0),$$

$$2 \frac{k\mathcal{P}(0) + \mathcal{P}(k) + (\Delta - k - 1)\lambda\mathcal{P}(0)}{1 - b} - \mathcal{P}(0) = 1$$

$$\geq 2 \frac{k\mathcal{P}(0) + \lambda\mathcal{P}(0) + (\Delta - k - 1)\lambda\mathcal{P}(0)}{1 - b} - \mathcal{P}(0),$$

and thus the constraints on $\mathcal{P}(0)$ are

$$\frac{1 - b}{2k + 2 + 2\lambda(\Delta - k - 1) - 1 + b} \leq \mathcal{P}(0)$$

$$\leq \frac{1 - b}{2k + 2\lambda(\Delta - k) - 1 + b}.$$

Since $V(\mathcal{P})$ is a linear function of $\mathcal{P}(0)$, to minimize the cost $V(\mathcal{P})$, either $\mathcal{P}(0) = \frac{1-b}{2k+2+2\lambda(\Delta-k-1)-1+b}$ or $\mathcal{P}(0) = \frac{1-b}{2k+2\lambda(\Delta-k)-1+b}$, i.e., $\mathcal{P}(0)$ should take one of the two extreme points of (367). To get these two extreme points, we have either $\mathcal{P}(k) = \mathcal{P}(0)$ or $\mathcal{P}(k) = \lambda\mathcal{P}(0) = \mathcal{P}(\Delta - 1)$.

Therefore, in the optimal probability mass function $\mathcal{P} \in \mathcal{SP}_{\text{pd}}$, there exists $k \in \{0, 1, \ldots, \Delta - 2\}$ such that

$$\mathcal{P}(i) = \mathcal{P}(0), \quad \forall i \in \{0, 1, \ldots, k\} \tag{367}$$

$$\mathcal{P}(i) = \mathcal{P}(\Delta - 1), \quad \forall i \in \{k + 1, k + 2, \ldots, \Delta - 1\}. \tag{368}$$

This completes the proof of Lemma 32. ∎

F. Step 5

In the last step, we prove that although $\lambda \in [e^{-\epsilon}, 1]$, in the optimal probability mass function, λ is either $e^{-\epsilon}$ or 1, and this will complete the proof of Theorem 13.

Proof: For fixed $k \in \{0, 1, \ldots, \Delta - 2\}$, consider $\mathcal{P} \in \mathcal{SP}_{\text{pd}}$ with

$$\mathcal{P}(i) = \mathcal{P}(0), \quad \forall i \in \{0, 1, \ldots, k\}, \tag{369}$$

$$\mathcal{P}(i) = \lambda\mathcal{P}(0), \quad \forall i \in \{k + 1, k + 2, \ldots, \Delta - 1\}. \tag{370}$$

Since $\sum_{i=-\infty}^{+\infty} \mathcal{P}(i) = 1$,

$$2 \frac{(k + 1)\mathcal{P}(0) + (\Delta - k - 1)\lambda\mathcal{P}(0)}{1 - b} - \mathcal{P}(0) = 1, \tag{371}$$

and thus

$$\mathcal{P}(0) = \frac{1 - b}{2(k + 1) + 2(\Delta - k - 1)\lambda - 1 + b}. \tag{372}$$

Hence, $\mathcal{P}$ is specified by only one parameter λ.

The cost of $\mathcal{P}$ is

$$V(\mathcal{P}) = \sum_{i=0}^{\Delta-1} \mathcal{P}(i)w_i \tag{373}$$

$$= \mathcal{P}(0)\sum_{i=0}^{k} w_i + \lambda\mathcal{P}(0)\sum_{k+1}^{\Delta-1} w_i \tag{374}$$

$$= \frac{(1-b)(\sum_{i=0}^{k} w_i + \lambda\sum_{i=k+1}^{\Delta-1} w_i)}{2(k+1) + 2(\Delta-k-1)\lambda - 1 + b} \tag{375}$$

$$= (1-b)(C_1 + \frac{C_2}{2(k+1) + 2(\Delta-k-1)\lambda - 1 + b}), \tag{376}$$

where C_1 and C_2 are constant terms independent of λ. Therefore, to minimize $V(\mathcal{P})$ over $\lambda \in [e^{-\epsilon}, 1]$, λ should take the extreme points, either $e^{-\epsilon}$ or 1, depending on whether C_2 is negative or positive.

When $\lambda = 1$, then the probability mass function is uniquely determined, which is $\mathcal{P} \in \mathcal{SP}_{\mathrm{pd}}$ with

$$\mathcal{P}(i) = \frac{1-b}{2\Delta - 1 + b}, \quad \forall i \in \{0, 1, \ldots, \Delta - 1\}, \tag{377}$$

which is exactly $\mathcal{P}_r$ defined in (82) with $r = \Delta$.

When $\lambda = e^{-\epsilon}$, the probability mass function is exactly $\mathcal{P}_r$ with $r = k + 1$.

Therefore, we conclude that

$$V^* = \min_{\{r \in \mathbb{N} \mid 1 \le r \le \Delta\}} \sum_{i=-\infty}^{+\infty} \mathcal{L}(i)\mathcal{P}_r(i). \tag{378}$$

∎

Acknowledgment

References

[1] C. Dwork, "Differential privacy: A survey of results," in *Proc. 5th Int. Conf. Theory Appl. Models Comput.*, vol. 4978. 2008, pp. 1–19.

[2] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography* (Lecture Notes in Computer Science), vol. 3876, S. Halevi and T. Rabin, Eds. Berlin, Germany: Springer, 2006, pp. 265–284.

[3] M. Hardt and K. Talwar, "On the geometry of differential privacy," in *Proc. 42nd ACM Symp. Theory Comput. (STOC)*, New York, NY, USA, 2010, pp. 705–714. [Online]. Available: http://doi.acm.org/10.1145/1806689.1806786

[4] A. Nikolov, K. Talwar, and L. Zhang, "The geometry of differential privacy: The sparse and approximate cases," in *Proc. 45th Annu. ACM Symp. Theory Comput. (STOC)*, 2012, pp. 351–360.

[5] C. Li, M. Hay, V. Rastogi, G. Miklau, and A. McGregor, "Optimizing linear counting queries under differential privacy," in *Proc. 29th ACM SIGMOD-SIGACT-SIGART Symp. Principles Database Syst. (PODS)*, New York, NY, USA, 2010, pp. 123–134. [Online]. Available: http://doi.acm.org/10.1145/1807085.1807104

[6] A. Ghosh, T. Roughgarden, and M. Sundararajan, "Universally utility-maximizing privacy mechanisms," in *Proc. 41st Annu. ACM Symp. Theory Comput. (STOC)*, New York, NY, USA, 2009, pp. 351–360. [Online]. Available: http://doi.acm.org/10.1145/1536414.1536464

[7] H. Brenner and K. Nissim, "Impossibility of differentially private universally optimal mechanisms," in *Proc. 51st Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, Oct. 2010, pp. 71–80.

[8] M. Gupte and M. Sundararajan, "Universally optimal privacy mechanisms for minimax agents," in *Proc. Symp. Principles Database Syst.*, 2010, pp. 135–146.

[9] Q. Geng and P. Viswanath, "The optimal mechanism in differential privacy," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2014, pp. 2371–2375.

[10] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 1376–1385.

[11] L. Wasserman and S. Zhou, "A statistical framework for differential privacy," *J. Amer. Statist. Assoc.*, vol. 105, no. 489, pp. 375–389, 2010. [Online]. Available: http://amstat.tandfonline.com/doi/abs/10.1198/jasa.2009.tm08651

[12] M. Hardt, K. Ligett, and F. McSherry, "A simple and practical algorithm for differentially private data release," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 2348–2356. [Online]. Available: http://books.nips.cc/papers/files/nips25/NIPS2012_1143.pdf

[13] F. McSherry and I. Mironov, "Differentially private recommender systems: Building privacy into the net," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2009, pp. 627–636. [Online]. Available: http://doi.acm.org/10.1145/1557019.1557090

[14] X. Xiao, G. Wang, and J. Gehrke, "Differential privacy via wavelet transforms," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 8, pp. 1200–1214, Aug. 2011.

[15] Z. Huang, S. Mitra, and G. Dullerud, "Differentially private iterative synchronous consensus," in *Proc. ACM Workshop Privacy Electron. Soc. (WPES)*, New York, NY, USA, 2012, pp. 81–90. [Online]. Available: http://doi.acm.org/10.1145/2381966.2381978

[16] F. McSherry, "Privacy integrated queries: An extensible platform for privacy-preserving data analysis," *Commun. ACM*, vol. 53, no. 9, pp. 89–97, Sep. 2010. [Online]. Available: http://doi.acm.org/10.1145/1810891.1810916

[17] B. Barak, K. Chaudhuri, C. Dwork, S. Kale, F. McSherry, and K. Talwar, "Privacy, accuracy, and consistency too: A holistic solution to contingency table release," in *Proc. 26th ACM SIGMOD-SIGACT-SIGART Symp. Principles Database Syst. (PODS)*, New York, NY, USA, 2007, pp. 273–282. [Online]. Available: http://doi.acm.org/10.1145/1265530.1265569

[18] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Proc. 24th Annu. Int. Conf. Theory Appl. Cryptogr. Techn. (EUROCRYPT)*, Berlin, Germany, 2006, pp. 486–503. [Online]. Available: http://dx.doi.org/10.1007/11761679_29

[19] C. Dwork and J. Lei, "Differential privacy and robust statistics," in *Proc. 41st Annu. ACM Symp. Theory Comput. (STOC)*, New York, NY, USA, 2009, pp. 371–380. [Online]. Available: http://doi.acm.org/10.1145/1536414.1536466

[20] A. Roth and T. Roughgarden, "Interactive privacy via the median mechanism," in *Proc. 42nd ACM Symp. Theory Comput. (STOC)*, New York, NY, USA, 2010, pp. 765–774. [Online]. Available: http://doi.acm.org/10.1145/1806689.1806794

[21] Y. Lindell and E. Omri, "A practical application of differential privacy to personalized online advertising," in *Proc. IACR Cryptol. ePrint Arch.*, 2011, p. 152. [Online]. Available: http://dblp.uni-trier.de/db/journals/iacr/iacr2011.html#LindellO11

[22] A. Smith, "Privacy-preserving statistical estimation with optimal convergence rates," in *Proc. 43rd Annu. ACM Symp. Theory Comput. (STOC)*, New York, NY, USA, 2011, pp. 813–822. [Online]. Available: http://doi.acm.org/10.1145/1993636.1993743

[23] K. Chaudhuri and C. Monteleoni, "Privacy-preserving logistic regression," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 289–296.

[24] C. Dwork, M. Naor, T. Pitassi, and G. N. Rothblum, "Differential privacy under continual observation," in *Proc. 42nd ACM Symp. Theory Comput. (STOC)*, New York, NY, USA, 2010, pp. 715–724. [Online]. Available: http://doi.acm.org/10.1145/1806689.1806787

[25] B. Ding, M. Winslett, J. Han, and Z. Li, "Differentially private data cubes: Optimizing noise sources and consistency," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, New York, NY, USA, 2011, pp. 217–228. [Online]. Available: http://doi.acm.org/10.1145/1989323.1989347

[26] M. Hardt and G. N. Rothblum, "A multiplicative weights mechanism for privacy-preserving data analysis," in *Proc. IEEE 51st Annu. Symp. Found. Comput. Sci. (FOCS)*, Washington, DC, USA, Oct. 2010, pp. 61–70. [Online]. Available: http://dx.doi.org/10.1109/FOCS.2010.85

[27] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. (Dec. 2012). "Geo-indistinguishability: Differential privacy for location-based systems." [Online]. Available: http://arxiv.org/abs/1212.1984

[28] 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. [Online]. Available: http://dx.doi.org/10.1137/090756090

[29] I. Mironov, "On significance of the least significant bits for differential privacy," in *Proc. ACM Conf. Comput. Commun. Security (CCS)*, New York, NY, USA, 2012, pp. 650–661. [Online]. Available: http://doi.acm.org/10.1145/2382196.2382264

[30] R. Sarathy and K. Muralidhar, "Evaluating laplace noise addition to satisfy differential privacy for numeric data," *Trans. Data Privacy*, vol. 4, no. 1, pp. 1–17, Apr. 2011. [Online]. Available: http://dl.acm.org/citation.cfm?id=2019312.2019313

[31] X. Xiao, G. Bender, M. Hay, and J. Gehrke, "iReduct: Differential privacy with reduced relative errors," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, New York, NY, USA, 2011, pp. 229–240. [Online]. Available: http://doi.acm.org/10.1145/1989323.1989348

[32] F. K. Dankar and K. El Emam, "The application of differential privacy to health data," in *Proc. Joint EDBT/ICDT Workshops (EDBT-ICDT)*, New York, NY, USA, 2012, pp. 158–166. [Online]. Available: http://doi.acm.org/10.1145/2320765.2320816

[33] A. Friedman and A. Schuster, "Data mining with differential privacy," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2010, pp. 493–502. [Online]. Available: http://doi.acm.org/10.1145/1835804.1835868

[34] J. Zhang, Z. Zhang, X. Xiao, Y. Yang, and M. Winslett, "Functional mechanism: Regression analysis under differential privacy," *Proc. VLDB Endowment*, vol. 5, no. 11, pp. 1364–1375, 2012.

[35] J. Lei, "Differentially private m-estimators," in *Proc. 23rd Annu. Conf. Neural Inf. Process. Syst.*, 2011, pp. 361–369.

[36] C. Dwork, M. Naor, T. Pitassi, G. N. Rothblum, and S. Yekhanin, "Pan-private streaming algorithms," in *Proc. ICS*, 2010, pp. 1–32.

[37] A. Gupta, K. Ligett, F. McSherry, A. Roth, and K. Talwar, "Differentially private combinatorial optimization," in *Proc. 21st Annu. ACM-SIAM Symp. Discrete Algorithms (SODA)*, Philadelphia, PA, USA, 2010, pp. 1106–1125. [Online]. Available: http://dl.acm.org/citation.cfm?id=1873601.1873691

[38] A. Blum and A. Roth. (2011). "Fast private data release algorithms for sparse queries." [Online]. Available: http://arxiv.org/abs/1111.6842

[39] J. Hsu, S. Khanna, and A. Roth, "Distributed private heavy hitters," in *Proc. 39th Int. Colloq. Conf. Automata, Lang., Program.*, 2012, pp. 461–472.

[40] J. Hsu, A. Roth, and J. Ullman. (2012). "Differential privacy for the analyst via private equilibrium computation." [Online]. Available: http://arxiv.org/abs/1211.0877

[41] J. Blocki, A. Blum, A. Datta, and O. Sheffet. (2012). "The Johnson–Lindenstrauss transform itself preserves differential privacy." [Online]. Available: http://arxiv.org/abs/1204.2136

[42] M. Hardt and A. Roth. (2012). "Beyond worst-case analysis in private singular vector computation." [Online]. Available: http://arxiv.org/abs/1211.0975

[43] M. Hardt, G. N. Rothblum, and R. A. Servedio, "Private data release via learning thresholds," in *Proc. 23rd Annu. ACM-SIAM Symp. Discrete Algorithms*, 2012, pp. 168–187.

[44] A. Gupta, A. Roth, and J. Ullman, "Iterative constructions and private data release," in *Proc. 9th Theory Cryptogr. Conf.*, 2012, pp. 339–356.

[45] S. P. Kasiviswanathan, K. Nissim, S. Raskhodnikova, and A. Smith, "Analyzing graphs with node differential privacy," in *Theory of Cryptography*. Tokyo, Japan: Springer, 2013, pp. 457–476.

[46] V. Karwa, S. Raskhodnikova, A. Smith, and G. Yaroslavtsev, "Private analysis of graph structure," *Proc. VLDB Endowment*, vol. 4, no. 11, pp. 1146–1157, 2011.

[47] G. Cormode, C. Procopiuc, D. Srivastava, E. Shen, and T. Yu, "Differentially private spatial decompositions," in *Proc. IEEE 28th Int. Conf. Data Eng. (ICDE)*, Apr. 2012, pp. 20–31.

[48] K. Nissim, S. Raskhodnikova, and A. Smith, "Smooth sensitivity and sampling in private data analysis," in *Proc. 29th Annu. ACM Symp. Theory Comput. (STOC)*, New York, NY, USA, 2007, pp. 75–84. [Online]. Available: http://doi.acm.org/10.1145/1250790.1250803

[49] A. Beimel, S. Kasiviswanathan, and K. Nissim, "Bounds on the sample complexity for private learning and private data release," in *Theory of Cryptography* (Lecture Notes in Computer Science), vol. 5978, D. Micciancio, Ed. Berlin, Germany: Springer, 2010, pp. 437–454. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-11799-2_26

[50] S. Shamai (Shitz) and S. Verdu, "Worst-case power-constrained noise for binary-input channels," *IEEE Trans. Inf. Theory*, vol. 38, no. 5, pp. 1494–1511, Sep. 1992.

[51] M. Hay, V. Rastogi, G. Miklau, and D. Suciu, "Boosting the accuracy of differentially private histograms through consistency," *Proc. VLDB Endowment*, vol. 3, nos. 1–2, pp. 1021–1032, Sep. 2010. [Online]. Available: http://dl.acm.org/citation.cfm?id=1920841.1920970

[52] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proc. 48th Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, Washington, DC, USA, Oct. 2007, pp. 94–103. [Online]. Available: http://dx.doi.org/10.1109/FOCS.2007.41

[53] Q. Geng, P. Kairouz, S. Oh, and P. Viswanath, "The staircase mechanism in differential privacy," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 7, pp. 1176–1184, Oct. 2015.

[54] Q. Geng and P. Viswanath, "Optimal noise adding mechanisms for approximate differential privacy," *IEEE Trans. Inf. Theory*, to be published.

[55] J. Soria-Comas and J. Domingo-Ferrer, "Work Session on Statistical Data Confidentiality," United Nations Economic Commission for Europe (UNECE)/Eurostat, Tarragona, Spain, Tech. Rep., Oct. 2011. [Online]. Available: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2011/presentations/Topic_4__24__Soria-Comas_Domingo.pdf

[56] J. Soria-Comas and J. Domingo-Ferrer, "Optimal data-independent noise for differential privacy," *Inf. Sci.*, vol. 250, pp. 200–214, Nov. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025513005094

Quan Geng is a software engineer at Google Inc., New York. He received his B.S. in Electronic Engineering from Tsinghua Univeristy in 2009, M.S. in Electrical and Computer Engineering in 2011, M.S. in Mathematics in 2012, and Ph.D. in Electrical and Computer Engineering in 2013 from University of Illinois at Urbana Champaign. He has interned at Microsoft Research Asia, Qualcomm Flarion Technologies and Tower Research Capital LLC. He worked at Tower Research Capital LLC as a high-frequency quantitative analyst from January 2014 to September 2015. His research interests include information theory, wireless communication, machine learning and differential privacy.

Pramod Viswanath (S'98–M'03–SM'10–F'13) received the Ph.D. degree in electrical engineering and computer science from the University of California at Berkeley, Berkeley, in 2000. He was a member of technical staff at Flarion Technologies until August 2001 before joining the Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign (UIUC), Urbana. Dr. Viswanath is a recipient of the Xerox Award for Faculty Research from the College of Engineering at UIUC (2010), the Eliahu Jury Award from the Electrical Engineering and Computer Science Department of the University of California at Berkeley (2000), the Bernard Friedman Award from the Mathematics Department of the University of California at Berkeley (2000), and the National Science Foundation (NSF) CAREER Award (2003). He was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY for the period 2006-2008.