# The Staircase Mechanism in Differential Privacy

Quan Geng, Peter Kairouz, Sewoong Oh, and Pramod Viswanath, *Fellow, IEEE*

***Abstract*—Adding Laplacian noise is a standard approach in differential privacy to sanitize numerical data before releasing it. In this paper, we propose an alternative noise adding mechanism: the staircase mechanism, which is a geometric mixture of uniform random variables. The staircase mechanism can replace the Laplace mechanism in each instance in the literature and for the same level of differential privacy, the performance in each instance improves; the improvement is particularly stark in medium-low privacy regimes. We show that the staircase mechanism is the optimal noise adding mechanism in a universal context, subject to a conjectured technical lemma (which we also prove to be true for one and two dimensional data).**

***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. It provides strong privacy guarantees by requiring the indistinguishability of whether an individual is in the dataset or not based on the released information. The key idea of differential privacy is that the presence or absence of any individual data 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. 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 vector real-valued query function

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

where $\mathcal{D}$ is the set of all possible datasets. The vector real-valued query function $q$ will be applied to a dataset, and query output is a real vector. 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 query output $q(D)$, where $D$ is the dataset.

*Definition 1 ($\epsilon$-Differential Privacy [1]): A randomized mechanism $\mathcal{K}$ gives $\epsilon$-differential privacy if for all data sets $D_1$ and $D_2$ differing on at most one element, and all $S \subset \mathrm{Range}(\mathcal{K})$,*

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

*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 $\epsilon$ is , the more private the randomized mechanism is.

The query function $q$ will be applied to a dataset $D \in \mathcal{D}$, and query output $q(D)$ can be written as $q(D) = (q_1(D), q_2(D), \ldots, q_d(D))$, which is a $d$-dimensional vector of real numbers.

The global sensitivity of multidimensional query function is defined as:

*Definition 2 (Query Sensitivity [1]): For a multidimensional real-valued query function $q : \mathcal{D} \to \mathbb{R}^d$, the sensitivity of $q$ is defined as*

$$\Delta := \max_{D_1, D_2 \in \mathcal{D}} \|q(D_1) - q(D_2)\|_1, \tag{3}$$

Manuscript received November 02, 2014; revised February 01, 2015; accepted April 18, 2015. Date of publication April 23, 2015; date of current version September 14, 2015. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Lalitha Sankar.

Q. Geng is with Tower Research Capital LLC, New York, NY 10081 USA (e-mail: gengquanshine@gmail.com).

P. Kairouz and P. Viswanath are with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61820-6903 USA (e-mail: kairouz2@illinois.edu; pramodv@illinois.edu).

S. Oh is with Industrial and Enterprise Systems Engineering Department, University of Illinois at Urbana-Champaign, Urbana, IL 61820-6903 USA (e-mail: swoh@illinois.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2015.2425831

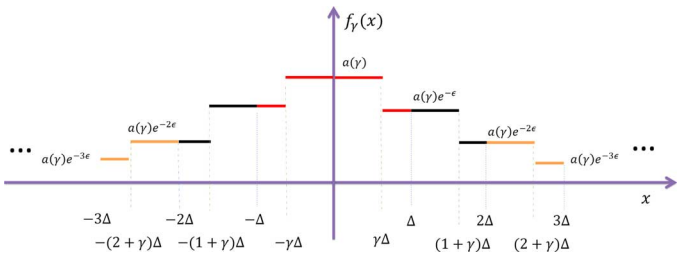


Fig. 1. One dimensional Staircase-Shaped Probability Density Function.

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

The standard approach to preserving the differential privacy is to add noise to the query output[1]. Let $q(D)$ be the value of the query function evaluated at $D \subseteq \mathcal{D}$, the noise-adding mechanism $\mathcal{K}$ will output

$$\mathcal{K}(D) = q(D) + X = (q_1(D) + X_1, \ldots, q_d(D) + X_d),$$

where $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ is the noise added by the mechanism to the output of query function. Specifically, the most popular approach in the literature is to add Laplace noise:

*Definition 3 (Laplacian Mechanism [2]): For a multidimensional real-valued query function $q : \mathcal{D} \to \mathbb{R}^d$ with sensitivity $\Delta$, the Laplacian mechanism will output*

$$\begin{aligned} K(D) &:= q(D) + \mathrm{Lap}\left(\frac{\Delta}{\epsilon}\right)^d \\ &= \left(q_1(D) + \mathrm{Lap}\left(\frac{\Delta}{\epsilon}\right), \ldots, q_d(D) + \mathrm{Lap}\left(\frac{\Delta}{\epsilon}\right)\right) \end{aligned}$$

*where $\mathrm{Lap}(\lambda)$ is a random variable with probability density function*

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

*and all $d$ Laplacian random variables are independent.*

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., [4]–[41]. Despite this near-universal use of the Laplacian mechanism there is no single demonstration of its optimality in any setting.

In this paper we propose an alternative noise distribution, that can *replace* Laplacian noise in each instance in the literature and for the same privacy level add "lesser amount" of noise, in a strong universal sense.

[1]Under the setting that the query function is real-valued and the released query output is also real-valued (either scalar or vector), all privacy preserving mechanisms can be viewed as noise-adding mechanisms, where the noise can be defined as the difference between the true query output and the released query output, and the noise can be either dependent on or independent of the true query output. In this paper we restrict ourselves to query-output independent noise-adding mechanisms, and we conjecture that the optimality of query-output independent noise-adding mechanisms also holds for the multidimensional setting, as for the single dimensional setting in [3].

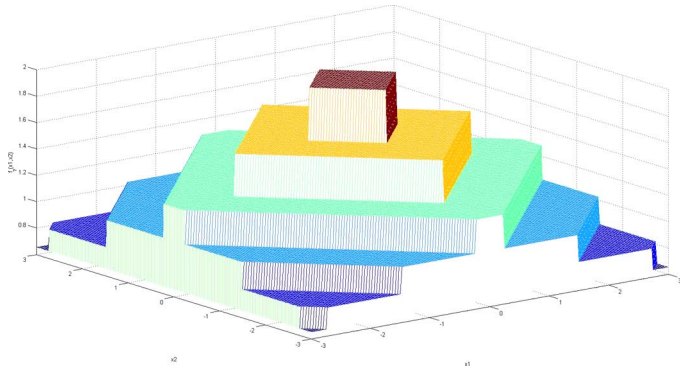

Fig. 2. Two dimensional Staircase-Shaped Probability Density Function.

## II. Staircase Mechanism

Consider a class of multidimensional probability distributions with symmetric and staircase-shaped probability density function defined as follows. Given $\gamma \in [0, 1]$, define $\mathcal{P}_\gamma$ as the probability distribution with probability density function $f_\gamma(\cdot)$ defined as

$$f_\gamma(\mathbf{x}) = \begin{cases} e^{-k\epsilon} a(\gamma) & \|\mathbf{x}\|_1 \in [k\Delta, (k+\gamma)\Delta) \\ e^{-(k+1)\epsilon} a(\gamma) & \|\mathbf{x}\|_1 \in [(k+\gamma)\Delta, (k+1)\Delta) \end{cases}$$

for $k \in \mathbb{N}$, where $a(\gamma)$ is the normalization factor to make

$$\int \int \cdots \int_{\mathbb{R}^d} f_\gamma(\mathbf{x}) dx_1 dx_2 \ldots dx_d = 1. \qquad (4).$$

Define $b \triangleq e^{-\epsilon}$, and define

$$c_k \triangleq \sum_{i=0}^{+\infty} i^k b^i, \quad \forall k \in \mathbb{N},$$

where by convention $0^0$ is defined as 1. Then the closed-form expression for $a(\gamma)$ is

$$a(\gamma) \triangleq \frac{d!}{2^d \Delta^d \sum_{k=1}^{d} \binom{d}{k} c_{d-k} (b + (1-b)\gamma^k)}.$$

It is straightforward to verify that $f_\gamma(\cdot)$ is a valid probability density function and $\mathcal{P}_\gamma$ satisfies the differential privacy constraint (9). Indeed, the probability density function $f_\gamma(x)$ satisfies

$$f_\gamma(\mathbf{x}) \le e^\epsilon f_\gamma(\mathbf{x} + \mathbf{t}), \ \forall \mathbf{x} \in \mathbb{R}^d, \ \forall \mathbf{t} \in \mathbb{R}^d \text{ s.t. } \|\mathbf{t}\|_1 \le \Delta,$$

which implies (9). We plot the probability density function $f_\gamma(\mathbf{x})$ in Fig. 1 for $d = 1$ and in Fig. 2 for $d = 2$. The nomenclature "staircase-shaped" is the visual structure of the pdf of the noise, as seen in these illustrations. More generally, one can visualize $f_\gamma(\mathbf{x})$ to be multi-dimensional staircase-shaped.

The staircase mechanism can be viewed as a *geometric mixture of uniform* random variables and is very easy to generate algorithmically. For the case of $d = 1$, a simple algorithmic implementation is provided in Algorithm 1.

**Algorithm 1** Generation of Random Variable with Staircase Distribution

**Input:** $\epsilon$, $\Delta$, and $\gamma \in [0,1]$.
**Output:** $X$, a random variable (r.v.) with staircase distribution specified by $\epsilon, \Delta$ and $\gamma$.
Generate a r.v. $S$ with $\Pr[S=1] = \Pr[S=-1] = 1/2$.
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 $\Pr[B = 0] = \gamma/(\gamma + (1-\gamma)b)$ and $\Pr[B = 1] = (1-\gamma)b/(\gamma + (1-\gamma)b)$.
$X \leftarrow S((1-B)((G+\gamma U)\Delta) + B((G+\gamma+(1-\gamma)U)\Delta))$.
Output $X$.

In the formula,

$$X \leftarrow S\left((1-B)\left((G+\gamma U)\Delta\right) + B\left((G+\gamma+(1-\gamma)U)\Delta\right)\right), \tag{5}$$

- $S$ 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.

## III. Comparison With Prior Work

In existing work on studying the tradeoff between accuracy and privacy in differential privacy, the usual metric of accuracy is in terms of the variance, or the expectation of the magnitude of the noise added to the query output. For example, Hardt and Talwar [42] 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 $\ell^2$-norm of the noise among all possible query output. [42] derives lower and upper bounds on the error given the differential privacy constraint. Nikolov, Talwar and Zhang [43] extend the result on the tradeoff between privacy and error to the case of $(\epsilon, \delta)$-differential privacy. Li *et al.*[9] 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, Roughgarden, and Sundararajan [44] 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. [44] 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 [45] show that for general query functions, no universally optimal differential privacy mechanisms exist. Gupte and Sundararajan [46] generalize the result of [44] to a minimax setting. McSherry and Talwar [47] introduce the exponential mechanism which is a generic differentially private mechanism and can apply to general abstract settings. In the multidimensional setting in this paper, the exponential mechanism can be reduced to the Laplacian mechanism.

The staircase mechanism was introduced in [3], for the single dimension case ($d = 1$). There it is proved that given an $\epsilon$-differential privacy constraint, under a general utility-maximization (equivalently, cost-minimization) model:

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

These results are derived under the following settings:

- when the domain of the query output is the entire real line or the set of all integers;
- nothing more about the query function is known beyond its global sensitivity;
- either local sensitivity [48] of the query function is unknown or it is the same as global sensitivity (as in the important case of count queries).

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 [16]), then the optimal privacy mechanism need not be data or query output dependent. The work in [3] has utility model same as the one adopted in [44] and [46], but the real-valued query function can have arbitrary sensitivity.

The contribution of this work is to generalize the results of [3] from the single dimensional setting to the multidimensional setting.

## IV. Optimality of Staircase Mechanisms

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

$$\mathcal{K}(D) = q(D) + X = (q_1(D) + X_1, \ldots, q_d(D) + X_d),$$

where $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ is the *independent* noise added by the mechanism to the output of the query function. We first derive the differential privacy constraint on the probability distribution of $X$ from (2).

$$\begin{aligned} &\Pr[\mathcal{K}(D_1) \in S] \leq e^\epsilon \Pr[\mathcal{K}(D_2) \in S] \\ \Leftrightarrow\ &\Pr[q(D_1) + X \in S] \leq e^\epsilon \Pr[q(D_2) + X \in S] \\ \Leftrightarrow\ &\Pr[X \in S - q(D_1)] \leq e^\epsilon \Pr[X \in S - q(D_2)] \\ \Leftrightarrow\ &\Pr[X \in S'] \leq e^\epsilon \Pr[X \in S' + q(D_1) - q(D_2)], \end{aligned} \tag{6}$$

where $S' \triangleq S - q(D_1) = \{s - q(D_1) | s \in S\}$.

Since (2) holds for all measurable sets $S \subseteq \mathbb{R}^d$, and $\|q(D_1) - q(D_2)\|_1 \leq \Delta$, from (6) we have

$$\Pr[X \in S'] \leq e^\epsilon \Pr[X \in S' + \mathbf{t}], \tag{7}$$

for all measurable sets $S' \subseteq \mathbb{R}$ and for all $\mathbf{t} \in \mathbb{R}^d$ such that $\|\mathbf{t}\|_1 \leq \Delta$. Equation (7) is very similar to the sliding property in [48]. In terms of the probability density function, the differential privacy condition is equivalent to

$$e^{-\epsilon} \leq \frac{f_X(\mathbf{u})}{f_X(\mathbf{u}+\mathbf{t})} \leq e^\epsilon, \quad \forall \mathbf{u} \in \mathbb{R}^d, \mathbf{t} \in \mathbb{R}^d \tag{8}$$

such that $\|\mathbf{t}\|_1 \leq \Delta$. Consider a cost function $\mathcal{L}(\cdot) : \mathbb{R}^d \to \mathbb{R}$ which is a function of the added noise $X$. Our goal is to minimize the expectation of the cost subject to the $\epsilon$-differential privacy constraint (7).

More precisely, let $\mathcal{P}$ denote the probability distribution of $X$ and use $\mathcal{P}(S)$ denote the probability $\Pr[X \in S]$. The optimization problem we study is

$$\begin{aligned}&\underset{\mathcal{P}}{\text{minimize}} \int\int\cdots\int_{\mathbb{R}^d} \mathcal{L}(x_1, x_2, \ldots, x_d)\mathcal{P}(dx_1 dx_2 \ldots dx_d)\\ &\text{subject to } \mathcal{P}(S) \leq e^{\epsilon}\mathcal{P}(S + \mathbf{t}),\\ &\qquad \forall \text{ measurable set } S \subseteq \mathbb{R}^d,\ \forall \|\mathbf{t}\|_1 \leq \Delta. \end{aligned} \tag{9}$$

We solve the above functional optimization problem and derive the optimal noise probability distribution for $\mathcal{L}(x_1, \ldots, x_d) = \sum_{i=1}^{d} |x_i|$. Consider the $\ell^1$ cost function:

$$\mathcal{L}(x_1, x_2, \ldots, x_d) = \sum_{i=1}^{d} |x_i|, \quad \forall (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d. \tag{10}$$

Let $\mathcal{SP}$ be the set of all probability distributions which satisfy the differential privacy constraint (9). Our main result is that the staircase mechanism is optimal for the $\ell^1$ cost function, stated below and proved subject to the validity of two technical lemmas, proved to be true for $d = 2$ and left as a conjecture more generally.

*Theorem 1: For $d = 2$ and the cost function $\mathcal{L}(\mathbf{x}) = \|\mathbf{x}\|_1,\ \forall \mathbf{x} \in \mathbb{R}^2$, then*

$$\begin{aligned}&\inf_{\mathcal{P}\in\mathcal{SP}} \int\int_{\mathbb{R}^2} \mathcal{L}(\mathbf{x})\mathcal{P}(dx_1 dx_2)\\ &= \inf_{\gamma\in[0,1]} \int\int_{\mathbb{R}^2} \mathcal{L}(\mathbf{x}) f_\gamma(\mathbf{x}) dx_1 dx_2.\end{aligned}$$

*Proof:* We briefly discuss the main proof idea and technique. For the full proof, we defer to Section VII.

First, by using a combinatorial argument, we show that given any noise probability distribution satisfying the $\epsilon$-differential privacy constraint, we can discretize the probability distribution by averaging it over each $\ell^1$ layer without increasing the cost. Therefore, we only need to consider those probability distributions with the probability density function being a piecewise constant function of the $\ell^1$-norm of the noise. Second, we show that to minimize the cost, the probability density function as a function of the $\ell^1$-norm of the noise should be monotonically and geometrically decaying. Lastly, we show that the optimal probability density function should be staircase-shaped.

Therefore, the optimal noise probability distribution to preserve $\epsilon$-differential privacy for multidimensional real-valued query function has a staircase-shaped probability density function, which is specified by three parameters $\epsilon$, $\Delta$ and $\gamma^* = \arg\min_{\gamma\in[0,1]} \int\int_{\mathbb{R}^2} \mathcal{L}(x_1, x_2) f_\gamma(\mathbf{x}) dx_1 dx_2$. ■

We conjecture that Theorem 1 holds for arbitrary dimension $d$. To prove this conjecture, one can reuse the whole proof in Section VII and only need to prove that Lemma 4 and Lemma 8 hold for arbitrary $d$, which we believe are true. Lemma 4 shows that when $d = 2$, we can discretize the probability distribution by averaging it over each $\ell^1$ layer without increasing the cost, and the new probability distribution also satisfies the differential privacy constraint. We give a constructive combinatorial argument to prove Lemma 4 for $d = 2$, and believe it holds for arbitrary $d \geq 2$. We prove Lemma 8 for $d = 2$ by studying the monotonicity of the ratio between the cost and volume over each $\ell^1$ layer. Indeed, to prove Lemma 8, one only needs to show that $h_k$, which is defined in Equation (144) in Section V.E of [50], first decreases and then increases as a function of $k$, and $h_0 \leq h_{i-1}$. For fixed $d$, one can derive the explicit formula for $d$ and verify whether $h_k$ satisfies this property (we show it is true for $d = 2$ in our proof). Based on this discussion and the *conjectured* validity of the technical lemmas for $d > 2$, we state the generalization of Theorem 1.

*Theorem 2: For $d > 2$ dimensional query, under the conjecture that Lemma 4 and Lemma 8 hold for arbitrary d, the staircase mechanism has the least cost function $\mathcal{L}(\mathbf{x}) = \|\mathbf{x}\|_1,\ \forall \mathbf{x} \in \mathbb{R}^d$ among all query-output independent noise-adding $\epsilon$-differentially private mechanisms, i.e.,*

$$\begin{aligned}&\inf_{\mathcal{P}\in\mathcal{SP}} \int\cdots\int_{\mathbb{R}^d} \mathcal{L}(\mathbf{x})\mathcal{P}(dx_1 dx_2 \cdots dx_d)\\ &\qquad = \inf_{\gamma\in[0,1]} \int\cdots\int_{\mathbb{R}^d} \mathcal{L}(\mathbf{x}) f_\gamma(\mathbf{x}) dx_1 dx_2 \cdots dx_d.\end{aligned}$$

## V. Implications

There are three parameters in the staircase mechanism: $\epsilon$, $\Delta$ and $\gamma^*$. The parameter $\epsilon$ is set by the differential privacy constraint, the parameter $\Delta$ is set by the global sensitivity of the query functions considered. The final parameter $\gamma \in [0, 1]$ is a free parameter that can be tuned to the specific cost function being considered. For instance, [3] studies the setting of $\gamma$ for the $d = 1$ setting in general. We recall these results briefly here.

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

$$\gamma^* = \frac{1}{1 + e^{\epsilon/2}}, \tag{11}$$

and the minimum expectation of noise amplitude is

$$V_1^* = \Delta \frac{e^{\epsilon/2}}{e^{\epsilon} - 1}. \tag{12}$$

On the other hand, the expectation of the amplitude of noise with Laplace distribution is

$$V_{Lap} \triangleq \int_{-\infty}^{+\infty} |x| f(x) dx = \frac{\Delta}{\epsilon}. \tag{13}$$

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

$$V_{Lap} - V(\mathcal{P}_{\gamma^*}) = \Delta\left(\frac{\epsilon}{24} - \frac{7\epsilon^3}{5760} + O(\epsilon^5)\right), \tag{14}$$

as $\epsilon \to 0$. In the low privacy regime ($\epsilon$ is large),

$$V_{Lap} = \frac{\Delta}{\epsilon}, \tag{15}$$

$$V(\mathcal{P}_{\gamma^*}) = \Theta(\Delta e^{-\epsilon/2}), \tag{16}$$

as $\epsilon \to +\infty$. Thus the gains of the staircase mechanism are particularly significant when $\epsilon$ is large, i.e., in the medium-low privacy regimes.

We now generalize this result to the multidimensional setting. Specifically, consider the setting of $d = 2$. Let $V^*$ denote the optimal cost for the $\ell^1$ cost function with $d = 2$. We have

*Corollary 3: In the high privacy regime,*

$$V^* = \frac{2\Delta}{\epsilon} - \frac{\Delta\epsilon^2}{36\sqrt{3}} + O(\epsilon^3), \epsilon \to 0,$$

*and in the low privacy regime,*

$$V^* = \sqrt[3]{2}\Delta e^{-\epsilon/3} + \frac{\Delta e^{-2\epsilon/3}}{\sqrt[3]{2}} + o\left(e^{-2\epsilon/3}\right), \epsilon \to +\infty.$$

The Laplacian mechanism adds independent Laplacian noise to each component of the query output, and the cost is $2\Delta/\epsilon$. Therefore, in the high privacy regime, the gap between optimal cost and the cost achieved by Laplacian mechanism goes to zero, as $\epsilon \to 0$, and we conclude Laplacian mechanism is approximately optimal in the high privacy regime. However, in the low privacy regime (as $\epsilon \to +\infty$), the optimal cost is proportional to $e^{-\epsilon/3}$, while the cost of Laplacian mechanism is proportional to $1/\epsilon$. We conclude the gap is significant in the low privacy regime.

## VI. Discussion

The differential privacy constraint on the pdf of the noise, from (8), implies that the ratio of the pdf evaluated at two different instances that are "neighbors" of each other is in the range $[e^{-\epsilon}, e^{\epsilon}]$. A closer look at the staircase mechanism reveals that its pdf satisfies the condition that the ratio of the pdf evaluated at two different instances that are "neighbors" of each other is exactly one of *three discrete values*: $\{e^{-\epsilon}, 1, e^{\epsilon}\}$. Motivated by the results here, any generic family of such differentially private mechanisms are denoted as *abstract staircase mechanisms* in [49]. In that work, it is also shown that staircase mechanisms are *extremal points* of the (convex) space of differentially private mechanisms and optimality of a large class of utility maximization problems is achieved by one of these staircase mechanisms.

## VII. Proof of Main Result

In this section we provide details of the proof of Theorem 1, and, due to space limitations, occasionally defer to [50, Section V] for the full details. The proof consists of 4 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 distributions which have symmetric piecewise constant probability density functions.
- Step 2 proves that we only need to consider those symmetric piecewise constant probability density functions which are monotonically decreasing.
- Step 3 proves that optimal probability density function should periodically decay.
- Step 4 proves that the optimal probability density function is staircase-shaped in the multidimensional setting, and it concludes the proof.

### A. Step 1

Given $\mathcal{P} \in \mathcal{SP}$, define

$$V(\mathcal{P}) \triangleq \int\int\cdots\int_{\mathbb{R}^d} \mathcal{L}(\mathbf{x})\mathcal{P}(dx_1 dx_2 \ldots dx_d). \quad (17).$$

Define

$$V^* \triangleq \inf_{\mathcal{P}\in\mathcal{SP}} V(\mathcal{P}). \quad (18).$$

Our goal is to prove that $V^* = \inf_{\gamma\in[0,1]} \int\int\cdots\int_{\mathbb{R}^d} \mathcal{L}(\mathbf{x}) f_\gamma(\mathbf{x}) dx_1 dx_2 \ldots dx_d$.

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

$$\inf_{\gamma\in[0,1]} \int\int\cdots\int_{\mathbb{R}^d} \mathcal{L}(\mathbf{x}) f_\gamma(\mathbf{x}) dx_1 dx_2 \ldots dx_d \geq V^* = +\infty, \quad (19)$$

and thus $\inf_{\gamma\in[0,1]} \int\int\cdots\int_{\mathbb{R}^d} \mathcal{L}(\mathbf{x}) f_\gamma(\mathbf{x}) dx_1 dx_2 \ldots dx_d = V^* = +\infty$. 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 show that given any probability measure $\mathcal{P} \in \mathcal{SP}$, we can use a sequence of probability measures with multidimensionally piecewise constant probability density functions to approximate $\mathcal{P}$.

Given $i \in \mathbb{N}$ and $k \in \mathbb{N}$, define

$$A_i(k) = \left\{\mathbf{x} \in \mathbb{R}^d | k\frac{\Delta}{i} \leq \|\mathbf{x}\|_1 < (k+1)\frac{\Delta}{i}\right\} \subset \mathbb{R}^d. \quad (20)$$

It is easy to calculate the volume of $A_i(k)$, which is

$$\mathrm{Vol}(A_i(k)) = \frac{2^d}{d!}\left((k+1)^d - k^d\right)\frac{\Delta^d}{i^d}. \quad (21)$$

*Lemma 4: Given $\mathcal{P} \in \mathcal{SP}$ with $V(\mathcal{P}) < +\infty$, any positive integer $i \in \mathbb{N}$, define $\mathcal{P}_i$ as the probability distribution with probability density function $f_i(\mathbf{x})$ defined as*

$$f_i(\mathbf{x}) = a_i(k) \triangleq \frac{\mathcal{P}(A_i(k))}{\mathrm{Vol}(A_i(k))}\mathbf{x} \in A_i(k) \text{ for } k \in \mathbb{N}. \quad (22)$$

*Then $\mathcal{P}_i \in \mathcal{SP}$ and $\lim_{i\to+\infty} V(\mathcal{P}_i) = V(\mathcal{P})$.*

We conjecture that Lemma 4 holds for arbitrary dimension $d$, and prove it for the case $d = 2$. Before proving Lemma 4 for $d = 2$, we prove an auxiliary Lemma which shows that for probability mass function over $\mathbb{Z}^2$ satisfying $\epsilon$-differential privacy constraint, we can construct a new probability mass function by averaging the old probability mass function over each $\ell^1$ ball and the new probability mass function still satisfies the $\epsilon$-differential privacy constraint.

*Lemma 5: For any given probability mass function $\mathcal{P}$ defined over the set $\mathbb{Z}^2$ satisfying that*

$$\mathcal{P}(i_1, j_1) \leq e^{\epsilon}\mathcal{P}(i_2, j_2), \quad \forall |i_1 - i_2| + |j_1 - j_2| \leq \Delta, \quad (23)$$

*define the probability mass function $\tilde{\mathcal{P}}$ via*

$$\tilde{\mathcal{P}}(i,j) = \begin{cases} \mathcal{P}(0,0) & (i,j) = (0,0) \\ p_{|i|+|j|} & (i,j) \neq (0,0) \end{cases} \quad (24)$$

*where $p_k \triangleq \sum_{(i',j')\in\mathbb{Z}^2:|i'|+|j'|=k} \mathcal{P}(i',j')/4k$, $\forall k \geq 1$.*

*Then $\tilde{\mathcal{P}}$ is also a probability mass function satisfying the differential privacy constraint, i.e.,*

$$\tilde{\mathcal{P}}(i_1,j_1) \le e^{\epsilon}\tilde{\mathcal{P}}(i_2,j_2), \quad \forall |i_1-i_2|+|j_1-j_2| \le \Delta. \quad (25)$$

*Proof:* Due to the way how we define $\tilde{\mathcal{P}}$, we have

$$\sum_{(i,j)\in\mathbb{Z}^2} \tilde{\mathcal{P}}(i,j) = \sum_{(i,j)\in\mathbb{Z}^2} \mathcal{P}(i,j) = 1, \quad (26)$$

and thus $\tilde{\mathcal{P}}$ is a valid probability mass function defined over $\mathbb{Z}^2$.

Next we prove that $\tilde{\mathcal{P}}$ satisfies (25). To simplify notation, define $p_0 \triangleq \mathcal{P}(0,0)$. Then we only need to prove that for any $k_1,k_2\in\mathbb{N}$ such that $|k_1-k_2|\le\Delta$, we have $p_{k_1}\le e^{\epsilon}p_{k_2}$. Due to the symmetry property, without loss of generality, we can assume $k_1<k_2$. The easiest case is $k_1=0$. When $k_1=0$, we have $k_2\le\Delta$ and

$$\mathcal{P}(0,0) \le e^{\epsilon}\mathcal{P}(i,j), \quad \forall |i|+|j|=k_2. \quad (27)$$

The number of distinct pairs $(i,j)$ satisfying $|i|+|j|=k$ is $4k$ for $k\ge 1$. Sum up all inequalities in (27), and we get $p_0\le e^{\epsilon}p_{k_2}$. For general $0<k_1<k_2$, let $\Delta' \triangleq k_2-k_1\le\Delta$. Define $B_k$ via

$$B_k \triangleq \{(i,j)\in\mathbb{Z}^2 \| i|+|j|=k\}, \quad \forall k\in\mathbb{N}. \quad (28)$$

Then the differential privacy constraint (23) implies that

$$\mathcal{P}(i_1,j_1) \le e^{\epsilon}\mathcal{P}(i_2,j_2), \quad \forall (i_1,j_1)\in B_{k_1}, (i_2,j_2)\in B_{k_2}, \quad (29)$$

and $|i_1-i_2|+|j_1-j_2|=\Delta'$. The set of points in $B_k$ forms a rectangle, which has 4 corner points and $4(k-1)$ interior points on the edges. For each corner point in $B_{k_1}$, which appears in the left side of (29), there are $(2\Delta'+1)$ points in $B_{k_2}$ close to it with an $\ell^1$ distance of $\Delta'$. And for each interior point in $B_{k_1}$, there are $(\Delta'+1)$ points in $B_{k_2}$ close to it with an $\ell^1$ distance of $\Delta'$. Therefore, there are in total $4(2\Delta'+1)+4(k_1-1)(\Delta'+1)$ distinct inequalities in (29).

If we can find certain nonnegative coefficients such that multiplying each inequality in (29) by these nonnegative coefficients and summing them up gives us

$$\frac{\sum_{(i',j')\in\mathbb{Z}^2:|i'|+|j'|=k_1}\mathcal{P}(i',j')}{4k_1} \le e^{\epsilon}\frac{\sum_{(i',j')\in\mathbb{Z}^2:|i'|+|j'|=k_2}\mathcal{P}(i',j')}{4k_2}, \quad (30)$$

then (25) holds. Therefore, our goal is to find the "right" coefficients associated with each inequality in (29). We formulate it as a matrix filling-in problem in which we need to choose nonnegative coefficients for certain entries in a matrix such that the sum of each row is $(k_1+\Delta')/k_1$, and the sum of each column is 1.

More precisely, label the $4k_1$ points in $B_{k_1}$ by $\{I_1,I_2,I_3,\ldots,I_{4k_1}\}$, where we label the topmost point by 1 and sequentially label other points clockwise. Similarly, we label the $4k_2$ points in $B_{k_2}$ by $\{O_1,O_2,O_3,\ldots,O_{4k_2}\}$, where we label the topmost point by 1 and sequentially label other points clockwise.

Consider the following $4k_1$ by $4k_2$ matrix $M$, where each row corresponds to the point in $B_{k_1}$ and each column corresponds to the point in $B_{k_2}$, and the entry $M_{ij}$ in the $i$th row and $j$th column is the coefficient corresponds to inequality involved with the points $I_i$ and $O_j$. If there is no inequality associated with the points $I_i$ and $O_j$, then $M_{ij}=0$. In the case $k_1=2$ and $\Delta'=3$, the zeros/nonzeros pattern of $M$ has the following form:

$$\begin{pmatrix}
x & x & x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x & x \\
0 & x & x & x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & x & x & x & x & x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & x & x & x & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & x & x & x & x & x & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x & x & x & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x & x & x & x & x \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x & x
\end{pmatrix}, \quad (31)$$

where $x$ denotes an entry which can take any nonnegative coefficient.

For general $k_1$ and $k_2$, the pattern of $M$ is that the first, $(k_1+1)$th, $(2k_1+1)$th and $(3k_1+1)$th rows can have $2\Delta'+1$ nonzero entries, and all other rows can have $\Delta'+1$ nonzero entries.

We want to show that

$$\frac{\sum_{(i',j')\in\mathbb{Z}^2:|i'|+|j'|=k_1}\mathcal{P}(i',j')}{4k_1} \le e^{\epsilon}\frac{\sum_{(i',j')\in\mathbb{Z}^2:|i'|+|j'|=k_2}\mathcal{P}(i',j')}{4k_2},$$

or equivalently,

$$\left(1+\frac{\Delta'}{k_1}\right)\sum_{(i',j')\in\mathbb{Z}^2:|i'|+|j'|=k_1}\mathcal{P}(i',j') \le e^{\epsilon}\sum_{(i',j')\in\mathbb{Z}^2:|i'|+|j'|=k_2}\mathcal{P}(i',j').$$

Therefore, our goal is to find nonnegative coefficients to substitute each $x$ in the matrix such that the sum of each column is 1 and the sum of each column is $(1+\Delta'/k_1)$. We will give explicit formulas on how to choose the coefficients. The case $k_1=1$ is trivial. Indeed, one can set all diagonal entries to be 1, and set all other nonzero entries to be 1/2. Therefore, we can assume $k_1>1$. Consider two different cases: $k_1\le\Delta'$ and $k_1\ge\Delta'+1$. We first consider the case $k_1\le\Delta'$. Due to the periodic patterns in $M$, we only need to consider rows from 1 to $k_1+1$. Set all entries to be zero except that we set

$$M_{11}=M_{22}=\cdots=M_{k_1k_1}=1, \quad (32)$$

$$M_{2,\Delta'+2}=M_{3,\Delta'+3}=\cdots=M_{k_1+1,k_1+\Delta'+1}=1 \quad (33)$$

$$M_{i,j}=\frac{1-\frac{\Delta'}{k_1(\Delta'-k_1+1)}}{k_1-1}. \quad (34)$$

Further set $M_{1,j}=\Delta'/2k_1(\Delta'-k_1+1)$ for $j\in[k_1+1,\Delta'+1]\cup[4k_1-\Delta'+1,4k_1]$ and $M_{k_1+1,j}=\Delta'/2k_1(\Delta'-k_1+1)$ for $j\in[k_1+1,\Delta'+1]\cup[2k_1+1+\Delta',k_1+1+2\Delta']$. It is straightforward to verify that the above matrix $M$ satisfies the properties that the sum of each column is 1 and the sum of each row is $(1+\Delta'/k_1)$. Therefore, we have $p_{k_1}\le e^{\epsilon}p_{k_2}, \forall 0<k_1<k_2, k_1\le k_2-k_1\le\Delta$.

Next we solve the case $k_1\ge\Delta'+1$. Again due to the periodic patterns in $M$, we only need to consider the nonzero entries in rows from 1 to $k_1+1$. We use the following procedures to construct $M$:

1) For the first row, set $M_{11}=1$ and set all other $2\Delta'$ nonzero entries to be $1/2k_1$.
2) For the second row, $M_{22}$ is uniquely determined to be $1-1/2k_1$. Set the next $\Delta'-1$ nonzero entries in the second row to be $1/k_1$, i.e., $M_{2j}=1/k_1$ for $j\in[3,\Delta'+1]$. The last nonzero entry $M_{2,\Delta'+2}$ is uniquely determined to be

$$\left(1+\frac{\Delta'}{k_1}\right)-\left(1-\frac{1}{2k_1}\right)-\frac{\Delta'-1}{k_1}=\frac{3}{2k_1}. \tag{35}$$

3) For the third row, the first nonzero entry $M_{33}$ is uniquely determined to be $1-1/2k_1-1/k_1=1-3/2k_1$. Set the next $\Delta'-1$ nonzero entries to be $1/k_1$, i.e., $M_{3j}=1/k_1$ for $j\in[4,\Delta'+2]$. The last nonzero entry $M_{3,\Delta'+3}$ is uniquely determined to be

$$\left(1+\frac{\Delta'}{k_1}\right)-\left(1-\frac{3}{2k_1}\right)-\frac{\Delta'-1}{k_1}=\frac{5}{2k_1}. \tag{36}$$

4) In general, for the $i$th row ($i\in[2,k_1-1]$), the first nonzero entry $M_{ii}$ is set to be $M_{ii}=1-(2i-3)/2k_1$, and the next $\Delta'-1$ nonzero entries are $1/k_1$, and the last nonzero entry $M_{i,i+\Delta'}=(2i-1)/2k_1$.
5) For $(k_1+1)$th row, by symmetry, we set $M_{k_1+1,k_1+1}=1$ and set other $2\Delta'$ nonzero entries to be $1/2k_1$.
6) The nonzero entries in the $k_1$th row are uniquely determined. Indeed, we have

$$M_{k_1,k_1}=1-\frac{2k_1-3}{2k_1}, \tag{37}$$

$$M_{k_1,k_1+\Delta'}=1-\frac{1}{2k_1}, \tag{38}$$

$$M_{k_1,k_1+j}=\frac{1}{k_1},\ j\in[2,\Delta'-1]. \tag{39}$$

It is straightforward to verify that each entry in $M$ is nonnegative and $M$ satisfies the properties that the sum of each column is 1 and the sum of each row is $(1+\Delta'/k_1)$. Therefore, we have $p_{k_1}\le e^{\epsilon}p_{k_2},\forall 0<k_1<k_2,k_1\ge\Delta'+1=k_2-k_1+1$. Therefore, for all $k_1,k_2\in\mathbb{N}$ such that $k_2-k_1|\le\Delta$, we have $p_{k_1}\le e^{\epsilon}p_{k_2}$. This completes the proof of Lemma 5. ■

We defer the (conceptually straightforward) proof of Lemma 4 to [50, Section V], due to space limitations. Define $\mathcal{SP}_{i,\mathrm{sym}}\triangleq\{\mathcal{P}_i|\mathcal{P}\in\mathcal{SP}\}$ for $i\ge1$, i.e., $\mathcal{SP}_{i,\mathrm{sym}}$ is the set of probability distributions satisfying differential privacy constraint (9) and having symmetric piecewise constant (over $A_i(k)\ \forall k\in\mathbb{N}$) probability density functions.

Due to Lemma 4, we have that $V^*=\inf_{\mathcal{P}\in\cup_{i=1}^{\infty}\mathcal{SP}_{i,\mathrm{sym}}}V(\mathcal{P})$. Therefore, to characterize $V^*$, we only need to study probability distributions with symmetric and piecewise constant probability density functions.

### B. Step 2

Given $\mathcal{P}\in\mathcal{P}_{\mathrm{sym}}$, we call $\{a_i(0),a_i(1),a_i(2),\ldots\}$ the density sequence of $\mathcal{P}_i\in\mathcal{SP}_{i,\mathrm{sym}}$, where $a_i(k)$ is defined in (22) $\forall k\in\mathbb{N}$.

Next we show that indeed we only need to consider those probability distributions with symmetric piecewise constant probability density functions the density sequences of which are *monotonically decreasing*. Define $\mathcal{SP}_{i,\mathrm{md}}\triangleq\{\mathcal{P}|\mathcal{P}\in\mathcal{SP}_{i,\mathrm{sym}}$, and the density sequence of $\mathcal{P}$ is monotonically decreasing. Then

*Lemma 6:*

$$V^*=\inf_{\mathcal{P}\in\cup_{i=1}^{\infty}\mathcal{SP}_{i,\mathrm{md}}}V(\mathcal{P}). \tag{40}$$

*Proof:* Due to the space limit, we refer the readers to Section V.C of [50] for the proof. ■

### C. Step 3

Next we show that among all symmetric piecewise constant probability density functions, we only need to consider those which are geometrically decaying. More precisely, given positive integer $i$, we have $\mathcal{SP}_{i,\mathrm{pd}}\triangleq\{\mathcal{P}|\mathcal{P}\in\mathcal{SP}_{i,\mathrm{md}}$ and $\mathcal{P}$ has density sequence $\{a_0,a_1,\ldots,a_n,\ldots,\}$ satisfying $a_k/a_{k+i}=e^{\epsilon},\forall k\in\mathbb{N}\}$, then

*Lemma 7:*

$$V^*=\inf_{\mathcal{P}\in\cup_{i=1}^{\infty}\mathcal{SP}_{i,\mathrm{pd}}}V(\mathcal{P}). \tag{41}$$

*Proof:* Due to the space limit, we refer the readers to Section V.D of [50] for the proof. ■

Due to Lemma 7, we only need to consider probability distribution with symmetric, monotonically decreasing, and geometrically decaying piecewise constant probability density function. Because of the properties of symmetry and periodically (geometrically) decaying, for this class of probability distributions, the probability density function over $\mathbb{R}^d$ is completely determined by the probability density function over the set $\{\mathbf{x}\in\mathbb{R}^d|\|\mathbf{x}\|_1<\Delta\}$. Next, we study what the optimal probability density function should be over the set $\{\mathbf{x}\in\mathbb{R}^d|\|\mathbf{x}\|_1<\Delta\}$. It turns out that the optimal probability density function over the set $\{\mathbf{x}\in\mathbb{R}^d|\|\mathbf{x}\|_1<\Delta\}$ is a step function. We use the following three steps to prove this result.

### D. Step 4

*Lemma 8: Consider a probability distribution $\mathcal{P}_a\in\mathcal{SP}_{i,\mathrm{pd}}$ ($i\ge2$) with density sequence $\{a_0,a_1,\ldots,a_n,\ldots\}$. Then there exists an integer $k(i)$ and a probability distribution $\mathcal{P}_b\in\mathcal{SP}_{i,\mathrm{pd}}$ with density sequence $\{b_0,b_1,\ldots,b_n,\ldots\}$ such that*

$$b_0=b_1=b_2=\cdots=b_{k(i)}, \tag{42}$$

$$\frac{b_0}{b_{i-1}}=e^{\epsilon}, \tag{43}$$

*and*

$$V(\mathcal{P}_b)\le V(\mathcal{P}_a). \tag{44}$$

*Proof:* Due to the space limit, we refer the readers to Section V.E of [50] for the proof. ■

Therefore, due to Lemma 8, for sufficiently large $i$, 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

$$a_0=a_1=a_2=\cdots=a_{k(i)}, \tag{45}$$

$$\frac{b_0}{b_{i-1}}=e^{\epsilon}. \tag{46}$$

More precisely, define $\mathcal{SP}_{i,\mathrm{fr}} = \{\mathcal{P} \in \mathcal{SP}_{i,\mathrm{pd}} | \mathcal{P}$ has density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying (45) and (46). Then due to Lemma 8,

*Lemma 9:*

$$V^* = \inf_{\mathcal{P} \in \cup_{i=3}^{\infty} \mathcal{SP}_{i,\mathrm{fr}}} V(\mathcal{P}). \tag{47}$$

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 $k(i) + 1 \leq k \leq (i-2)$, such that

$$a_j = a_0, \quad \forall 0 \leq j < k, \tag{48}$$

$$a_j = a_{i-1}, \quad \forall k < j < i. \tag{49}$$

More precisely,

*Lemma 10: 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 $k(i) + 1 \leq k \leq (i-2)$ with*

$$b_j = a_0, \quad \forall\, 0 \leq j < k, \tag{50}$$

$$b_j = a_{i-1}, \quad \forall\, k < j < i, \tag{51}$$

*and*

$$V(\mathcal{P}_b) \leq V(\mathcal{P}_a). \tag{52}$$

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

$$a_j = a_0, \quad \forall\, 0 \leq j < k, \tag{53}$$

$$a_j = a_{i-1}, \quad \forall\, k < j < i, \tag{54}$$

then we can set $\mathcal{P}_b = \mathcal{P}_a$. Otherwise, let $k_1$ be the smallest integer in $\{k(i)+1, k(i)+2, \ldots, i-1\}$ such that $a_{k_1} \neq a_0$, and let $k_2$ be the biggest integer in $\{k(i)+1, k(i)+2, \ldots, i-1\}$ such that $a_{k_2} \neq a_{i-1}$. It is easy to see that $k_1 \neq k_2$. Then we can scale up $a_{k_1}$ and scale down $a_{k_2}$ simultaneously until either $a_{k_1} = a_0$ or $a_{k_2} = a_{i-1}$. Since $h_k \triangleq w_k/u_k$ is an increasing function of $k$ when $k > k(i)$, and $k(i) < k_1 < k_2$, this scaling operation will not increase the cost. After this scaling operation we can update $k_1$ and $k_2$, and either $k_1$ is increased by one or $k_2$ is decreased by one. 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 (50), (51) and (52) hold. This completes the proof. ■

Define $\mathcal{SP}_{i,\mathrm{step}} = \{\mathcal{P} \in \mathcal{SP}_{i,\mathrm{fr}} \mid \mathcal{P}$ has density sequence $\{a_0, a_1, \ldots, a_n, \ldots\}$ satisfying (50) and (51) for some $k(i) < k \leq (i-2)$ Then due to Lemma 10,

*Lemma 11:* $V^* = \inf_{\mathcal{P} \in \cup_{i=3}^{\infty} \mathcal{SP}_{i,\mathrm{step}}} V(\mathcal{P})$.

As $i \to \infty$, the probability density function of $\mathcal{P} \in \mathcal{SP}_{i,\mathrm{fr}}$ will converge to a multidimensional staircase function. Therefore, for $d = 2$ and the cost function $\mathcal{L}(\mathbf{x}) = \|\mathbf{x}\|_1, \forall \mathbf{x} \in \mathbb{R}^2$, then

$$\inf_{\mathcal{P} \in \mathcal{SP}} \int\int_{\mathbb{R}^2} \mathcal{L}(\mathbf{x}) \mathcal{P}(dx_1 dx_2) = \inf_{\gamma \in [0,1]} \int\int_{\mathbb{R}^2} \mathcal{L}(\mathbf{x}) f_\gamma(\mathbf{x}) dx_1 dx_2. \tag{55}$$

This completes the proof of Theorem 1.

## ACKNOWLEDGMENT

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

## REFERENCES

[1] C. Dwork, "Differential privacy: A survey of results," in *Proc. 5th Int. Conf. Theory Applicat. Models Comput.*, Berlin/Heidelberg, Germany, 2008, TAMC'08, 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*, S. Halevi and T. Rabin, Eds. Berlin/Heidelberg, Germany: Springer, 2006, vol. 3876, Lecture Notes in Computer Science, pp. 265–284.

[3] Q. Geng and P. Viswanath, "The optimal mechanism in differential privacy," *ArXiv E-Prints*, Dec. 2012.

[4] M. Hardt, K. Ligett, and F. McSherry, P. Bartlett, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., "A simple and practical Algorithm for differentially private data release," in *Adv. Neural Inf. Process. Syst.*, 2012, pp. 2348–2356.

[5] F. McSherry and I. Mironov, "Differentially private recommender systems: Building privacy into the net," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Disc. Data Mining*, New York, NY, USA, 2009, KDD '09, pp. 627–636.

[6] 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.

[7] Z. Huang, S. Mitra, and G. Dullerud, "Differentially private iterative synchronous consensus," in *Proc. ACM Workshop Privacy Electron. Soc.*, New York, NY, USA, 2012, WPES '12, pp. 81–90.

[8] F. McSherry, "Privacy integrated queries: An extensible platform for privacy-preserving data analysis," *Commun. ACM*, vol. 53, no. 9, pp. 89–97, Sep. 2010.

[9] 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 '10)*, New York, NY, USA, 2010, pp. 123–134.

[10] 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 '07)*, New York, NY, USA, 2007, pp. 273–282.

[11] 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 Applicat. Cryptographic Tech. (EUROCRYPT '06)*, Berlin/Heidelberg, Germany, 2006, pp. 486–503.

[12] C. Dwork and J. Lei, "Differential privacy and robust statistics," in *Proc. 41st Annu. ACM Symp. Theory Comput.*, New York, NY, USA, 2009, STOC '09, pp. 371–380.

[13] A. Roth and T. Roughgarden, "Interactive privacy via the median mechanism," in *Proc. 42nd ACM Symp. Theory Comput. (STOC '10)*, New York, NY, USA, 2010, pp. 765–774.

[14] Y. Lindell and E. Omri, "A practical application of differential privacy to personalized online advertising," *IACR Cryptology ePrint Archive*, vol. 2011, p. 152, 2011.

[15] A. Smith, "Privacy-preserving statistical estimation with optimal convergence rates," in *Proc. 43rd Annu. ACM Symp. Theory Comput. (STOC '11)*, New York, NY, USA, 2011, pp. 813–822.

[16] K. Chaudhuri and C. Monteleoni, "Privacy-preserving logistic regression," in *Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2008, pp. 289–296.

[17] C. Dwork, M. Naor, T. Pitassi, and G. N. Rothblum, "Differential privacy under continual observation," in *Proc. 42nd ACM Symp. Theory Comput. (STOC '10)*, New York, NY, USA, 2010, pp. 715–724.

[18] 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 '11)*, New York, NY, USA, 2011, pp. 217–228.

[19] M. Hardt and G. N. Rothblum, "A multiplicative weights mechanism for privacy-preserving data analysis," in *Proc. IEEE 51st Annu. Symp. Foundat. Comput. Sci.*, Washington, DC, USA, 2010, FOCS '10, pp. 61–70.

[20] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," *ArXiv E-Prints*, Dec. 2012.
[21] 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.
[22] I. Mironov, "On significance of the least significant bits for differential privacy," in *Proc. ACM Conf. Comput. Commun. Security (CCS '12)*, New York, NY, USA, 2012, pp. 650–661.
[23] 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.
[24] 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 '11)*, New York, NY, USA, 2011, pp. 229–240.
[25] F. K. Dankar and K. El Emam, "The application of differential privacy to health data," in *Proc. Joint EDBT/ICDT Workshops (EDBT-ICDT '12)*, New York, NY, USA, 2012, pp. 158–166.
[26] A. Friedman and A. Schuster, "Data mining with differential privacy," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD '10)*, New York, NY, USA, 2010, pp. 493–502.
[27] 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.
[28] J. Lei, "Differentially private m-estimators," in *Proc. 23rd Annu. Conf. Neural Inf. Process. Syst.*, Granada, Spain, 2011, pp. 361–369.
[29] L. Wasserman and S. Zhou, "A statistical framework for differential privacy," *J. Amer. Statist. Assoc.*, vol. 105, no. 489, pp. 375–389, 2010.
[30] C. Dwork, M. Naor, T. Pitassi, G. N. Rothblum, and S. Yekhanin, "Pan-private streaming algorithms," in *Proc. 1st Symp. Innovat. Comput. Sci. (ICS '10)*, Beijing, China, 2010.
[31] 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 '10)*, Philadelphia, PA, USA, 2010, pp. 1106–1125.
[32] A. Blum and A. Roth, "Fast private data release algorithms for sparse queries," *ArXiv*, 2011 [Online]. Available: arXiv:1111.6842, to be published
[33] J. Hsu, S. Khanna, and A. Roth, "Distributed private heavy hitters," in *Proc. 39th Int. Colloqu. Conf. Automata, Lang., Program. (ICALP '12)*, Berlin/Heidelberg, Germany, 2012, vol. 1, pp. 461–472.
[34] J. Hsu, A. Roth, and J. Ullman, "Differential privacy for the analyst via private equilibrium computation," *ArXiv*, 2012 [Online]. Available: arXiv:1211.0877, to be published
[35] J. Blocki, A. Blum, A. Datta, and O. Sheffet, "The Johnson-Lindenstrauss transform itself preserves differential privacy," in *Proc. IEEE 53rd Annu. Symp. Foundat. Comput. Sci. (FOCS '12)*, Washington, DC, USA, 2012, pp. 410–419.
[36] M. Hardt and A. Roth, "Beyond worst-case analysis in private singular vector computation," in *Proc. 45th Annu. ACM Symp. Theory Comput. (STOC '13)*, New York, NY, USA, 2013, pp. 331–340.
[37] 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.
[38] A. Gupta, A. Roth, and J. Ullman, "Iterative constructions and private data release," *Theory of Cryptograp.*, pp. 339–356, 2012.
[39] S. P. Kasiviswanathan, K. Nissim, S. Raskhodnikova, and A. Smith, "Analyzing graphs with node differential privacy," in *Theory of Cryptography*. New York, NY, USA: Springer, 2013, pp. 457–476.
[40] V. Karwa, S. Raskhodnikova, A. Smith, and G. Yaroslavtsev, "Private analysis of graph structure," in *Proc. VLDB Endowment*, 2011, vol. 4, pp. 1146–1157.
[41] G. Cormode, C. Procopiuc, D. Srivastava, E. Shen, and T. Yu, "Differentially private spatial decompositions," in *Proc. IEEE 28th Int. Conf. Data Eng.*, 2012, pp. 20–31.
[42] M. Hardt and K. Talwar, "On the geometry of differential privacy," in *Proc. 42nd ACM Symp. Theory of Comput. (STOC '10)*, New York, NY, USA, 2010, pp. 705–714.
[43] 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 '13)*, New York, NY, USA, 2013, pp. 351–360.
[44] A. Ghosh, T. Roughgarden, and M. Sundararajan, "Universally utility-maximizing privacy mechanisms," in *Proc. 41st Annu. ACM Symp. Theory Comput. (STOC '09)*, New York, NY, USA, 2009, pp. 351–360.
[45] H. Brenner and K. Nissim, "Impossibility of differentially private universally optimal mechanisms," in *Proc. 51st Annu. IEEE Symp. Foundat. Comput. Sci. (FOCS '10)*, Oct. 2010, pp. 71–80.
[46] M. Gupte and M. Sundararajan, "Universally optimal privacy mechanisms for minimax agents," in *Proc. Symp. Principles Database Syst.*, 2010, pp. 135–146.
[47] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proc. 48th Annu. IEEE Symp. Foundat. Comput. Sci. (FOCS '07)*, Washington, DC, USA, 2007, pp. 94–103.
[48] K. Nissim, S. Raskhodnikova, and A. Smith, "Smooth sensitivity and sampling in private data analysis," in *Proc. 39th Annu. ACM Symp. Theory Comput. (STOC '07)*, New York, NY, USA, 2007, pp. 75–84.
[49] P. Kairouz, S. Oh, and P. Viswanath, "Extremal mechanisms for local differential privacy," *CoRR*, vol. abs/1407.1338, 2014.
[50] Q. Geng and P. Viswanath, "The optimal mechanism in differential privacy: Multidimensional setting," *CoRR*, vol. abs/1312.0655, 2013.

**Quan Geng** received his B.S. in electronic engineering from Tsinghua University 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 is a Quantitative Analyst in Tower Research Capital LLC. He has interned at Microsoft Research Asia, Qualcomm Flarion Technologies and Tower Research Capital LLC. His research interests include information theory, wireless communication, machine learning, and differential privacy.

**Peter Kairouz** received his M.S. in ECE from the University of Illinois at Urbana-Champaign (UIUC) in 2012 and his B.E. in ECE from the American University of Beirut (AUB) in 2010. He is a Ph.D. student at UIUC. He was a research intern at Qualcomm Inc. from May 2012 to August 2012 and May 2013 to August 2013. He has received numerous scholarships and awards including the Roberto Padovani Scholarship from Qualcomm's Research Center in 2012, the Distinguished Graduating Student Award from AUB's ECE department in 2010, and the Benjamin Franklin Scholarship from the United States Agency for International Development in 2007. His research interests include statistical data privacy and security, machine learning, and big data.

**Sewoong Oh** received his Ph.D. from the department of Electrical Engineering at Stanford University in 2011. He is an Assistant Professor of Industrial and Enterprise Systems Engineering at UIUC. He was a Postdoctoral Researcher at the Laboratory for Information and Decision Systems (LIDS) at MIT. His research interests are in statistical inference and 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.