Differential Private Discrete Noise-Adding Mechanism: Conditions, Properties, and Optimization

Shuying Qin, Jianping He, Chongrong Fang, and James Lam

Abstract—Differential privacy is a widely used framework for evaluating privacy loss in data anonymization. While the continuous noise-adding mechanism has been extensively studied, there is a dearth of research on discrete random mechanisms for discretely distributed data. This study addresses this gap by examining the primary differential privacy conditions and properties for general discrete random mechanisms, and investigating the trade-off between data privacy and data utility. We establish sufficient and necessary conditions for discrete \( \epsilon \)-differential privacy and sufficient conditions for discrete \(( \epsilon, \delta )\)-differential privacy, with closed-form expressions for differential privacy parameters. These conditions can be applied to evaluate the differential privacy properties of discrete noise-adding mechanisms with various types of noise. Moreover, we propose an optimal discrete \( \epsilon \)-differential private noise-adding mechanism under the utility-maximization framework. Here, the utility is characterized by the similarity of the statistical properties between the mechanism’s input and output. Our findings suggest that the optimal class of discrete noise probability distributions in the mechanism is staircase-shaped.

Index Terms—Differential privacy, discrete random mechanism, noise adding process, Wasserstein distance.

I. INTRODUCTION

A. Background

DATA anonymization is a widely used approach to safeguard data privacy in data publishing. To evaluate potential privacy loss, several privacy frameworks have been proposed, including information-theoretic privacy [2], differential privacy [3], and privacy based on secure multiparty computation [4]. Among these frameworks, differential privacy has received considerable attention due to its strong privacy guarantee. Introduced by Dwork et al. [5], the differential privacy framework protects either the record’s value or its presence in the data, depending on the definition of the neighboring datasets. Many studies have adopted the adjacency-based definition to protect the exact value of the record, as the generalized neighborhood property enables the privacy protection of specific data attributes [6], [7], [8], [9], [10].

The random noise-adding mechanism is a classic method for perturbing numerical data to achieve differential privacy. This mechanism is widely used in various fields, such as distributed optimization [11], [12], [13], control and network systems [14], [15], [16], collaborative computing [17], and others [18], [19], [20], [21], [22]. Additionally, it facilitates quantitative analysis of privacy guarantees when processing telemetry data [23], census data [24], and medical data [25]. However, the differential privacy preservation achieved by the random noise-adding mechanisms comes at the cost of data utility. Consequently, many research articles have focused on studying the trade-off between the privacy level and utility of these mechanisms [26], [27], [28], [29], [30], [31].

B. Motivations

In practice, a significant amount of data is discrete, including census records and traffic data. Random noise-adding mechanisms are commonly used to preserve data privacy. To ensure the interpretability of the protected numerical discrete data, the random noises added to the mechanisms should have discrete distributions, leading to the need for discrete noise-adding mechanisms. While researchers have extensively studied the continuous noise-adding mechanisms in terms of their privacy and utility properties, the conditions for discrete random mechanisms for discretely distributed data are unclear, and it is unknown whether the properties for continuous noise-adding mechanisms are applicable to the discrete ones. These issues call for discrete privacy-critical studies to ensure the deployment of discrete differential private mechanisms.

Moreover, to improve the utility of the data, it is crucial to consider the trade-off between the differential privacy level and the utility of the discrete random noise-adding mechanisms.
Most of the existing studies [26], [27], [32] model the utility function as a general function that depends on the noise added to the query output, which is reasonable but indirect. A key challenge is whether a more intuitive utility function exists on the level of data distortion caused by noises. One possible approach is to consider the similarity degree of the statistical properties between the original and noise-added data. Since the data are captured as random variables, this similarity degree can be quantified by the distance between the probability distributions [33]. The highly related divergence measures include Kullback-Leibler (KL) divergence, Jensen–Shannon (JS) divergence, and the Wasserstein distance. Our main interest is in building the utility metric. Furthermore, once the utility model is established, it is desirable to provide an implementable discrete random noise-adding mechanism for the utility optimization problem.

### C. Contributions

Building on the observations above, this paper investigates the differential privacy conditions, properties, and utility optimization for discrete random noise-adding mechanisms. We adopt an adjacency-based definition of differential privacy for generalized neighboring properties. Regarding the differential privacy analysis, we find that the conditions for discrete differential privacy mechanisms are more straightforward than those for continuous ones, with certain similarities in the differential privacy parameters [10]. Moreover, the differential privacy properties hold well in most discrete scenarios, such as the discrete Gaussian and Laplacian mechanisms. Our results for the discrete Gaussian mechanism partially align with the literature [34] on parameter computation results.

Regarding the trade-off between privacy level and utility, we propose using the Wasserstein distance as a utility measure. Our innovation lies in the utility model measuring the distance between the mechanism input and output, emphasizing the geometric properties of the distributions. Compared with the KL and JS divergence, the Wasserstein utility provides a more comprehensive analysis, smoothly representing the distance between two distributions defined in the space of probability measures, which is the basis of differential privacy. Additionally, it facilitates an in-depth analysis of all possible outcome events of the noise-adding mechanism. We then solve the utility-maximization problem using the Simplex Method [35]. In Table I, we compare our work with various studies on the differential privacy and utility properties of random noise-adding mechanisms.

This paper builds upon its conference version [1] and expands on its contributions in the following ways: i) we analyze the differential privacy properties of the discrete Exponential noise-adding mechanism, ii) we optimize the discrete noise-adding mechanisms for maximum data utility, iii) we provide sufficient simulations on the optimal discrete noise-adding mechanism.

The main contributions of this paper are summarized as follows:

- (Conditions.) We investigate the general differential privacy conditions for discrete noise-adding mechanisms, providing necessary and sufficient conditions for \(\epsilon\)-differential privacy and a sufficient condition for \((\epsilon, \delta)\)-differential privacy. Additionally, we derive closed-form expressions for the privacy parameters \(\epsilon\) and \(\delta\).

- (Properties.) We analyze the differential privacy properties and provide privacy guarantees for representative discrete noise-adding mechanisms based on the obtained theories. Specifically, we investigate the mechanisms under the discrete Gaussian, Laplacian, staircase-shaped, Uniform, and Exponential distributed noises.

- (Optimization.) We study the utility maximization for the \(\epsilon\)-differential private mechanisms. Defining the utility as the Wasserstein distance between the mechanism input and output probability distributions, we derive an optimal discrete staircase-shaped noise-adding mechanism. Furthermore, we conduct extensive simulations to verify its optimality.

### D. Organization

The remainder of this paper is organized as follows. In Section II, we present the related works. Section III provides the necessary preliminaries. In Section IV, we give the theoretical differential privacy conditions and a parameter computation method, analyze the differential privacy properties, and propose a discrete differential private mechanism with maximum utility. Section V provides evaluations of the mechanism’s optimality. Finally, in Section VI, we present the conclusions.

## II. RELATED WORK

Since Dwork [5] first introduced the differential privacy definition in 2006, it has become the flagship data privacy definition. Numerous attack models and scenarios are adapted to the variants and extensions of the differential privacy [36], [37]. More recently, Desfontaines et al. [38] gave a systematic taxonomy of the existing differential privacy definitions (approximately 225 kinds). They showed how new differential privacy definitions formed with the permutation and combination of seven dimensions. The original definition holds for datasets differing in one record. It is the conjunction of two situations: either two datasets have the same size and differ only on one record, or one is a copy of the other with one extra record. The former protects the value of certain records, while the latter protects the presence of records. The adjacency-based
definition is the representation of the one protecting the value of the records [6], [7], [8], [9], [10].

The mainstream of the differential privacy framework can be divided into two categories, the randomized response mechanism and the noise-adding mechanism [39]. In the randomized response mechanism, each participant reports the actual value in a differentially private manner based on the plausible deniability of responding to sensitive information. In the noise-adding mechanism, controlled probabilistic noise is added to the data we want to obfuscate before sending it to a data collector. Due to the designability of the noise, various noise-adding mechanisms have been proposed [10], [26], [27].

The majority of differential privacy researches focus on the continuous random noise-adding mechanisms to achieve anonymity for the continuously distributed data. Regarding the analysis of the general continuous random noise-adding mechanisms, He et al. [10] proposed a sufficient and necessary condition with specified privacy parameters. The fundamental theories can be applied to analyze various random noises. Then, they performed in-depth analysis of the differential privacy properties. Apart from the related analysis of the general continuous random mechanisms, differential privacy is widely discussed under a specific continuous random noise-adding mechanism. For instance, the continuous Gaussian noise-adding mechanism preserves $(\epsilon, \delta)$-differential privacy for the query functions with infinite dimensions and real values [40]. Besides, the continuous Laplacian noise-adding mechanism guarantees $\epsilon$-differential privacy [5].

Recently, researchers have paid attention to the discrete random differential private mechanisms. For instance, the Exponential mechanism is a well-known discrete noise-adding mechanism that aims to protect non-numerical discrete data [41]. It can guarantee $\epsilon$-differential privacy. Moreover, Canonne et al. [34] studied the differential privacy properties for the discrete Gaussian noise-adding mechanism. They obtained that the discrete Gaussian mechanism guarantees essentially the same level of privacy and accuracy as the continuous one. Apart from the related properties, Koskela et al. [42] proposed a Fourier transform-based numerical method to compute the differential privacy parameters. Specifically, they provided the lower and upper $(\epsilon, \delta)$-differential privacy bounds for the discrete Gaussian noise-adding mechanism.

In addition to the extensive research on the differential privacy properties, some works consider the fundamental trade-off between privacy level and utility of the random noise-adding mechanisms. Gupta et al. [26] found that the optimal differential private mechanism with a fixed query sensitivity is achieved by adding Geometric distributed noise. Based on the decision theory, they took the information loss caused by the random noise uncertainty as the utility measure. In this line of research, the fixed sensitivity was generalized to an arbitrary value, and the optimal noise with the staircase-shaped distribution was derived in [27]. The utility metric they adopted was the same as the one in [26]. The utility model is rational and risk-averse, but it is hard to determine how the added noise affects the original data directly. Based on the related work mentioned above, we investigate the discrete differential private noise-adding mechanism in this paper.

### III. Preliminaries

In this section, we introduce the discrete random noise-adding mechanisms, the differential privacy definition, and the proposed utility metric for the random mechanisms.

#### A. Preliminaries of Discrete Random Mechanisms

First, we specify the discrete quantitative data discussed in this paper by introducing a set of discrete numbers with interval $\Delta$, which is given by

$$Z_\Delta = \{ k | k = k_0 \Delta, k_0 \in \mathbb{Z} \}, \Delta \in \mathbb{R}^+.$$  

Denote $Z_\Delta^+$ as a set of positive numbers in $Z_\Delta$, and $Z_\Delta^n$ as a set of $n$-dimensional column vectors $L = [x_1, x_2, \ldots, x_n]^T$, where $x_i \in Z_\Delta, i \in V = \{1, 2, \ldots, n\}$.

Then, we introduce a discrete random noise-adding mechanism, which is a randomized function. Let $\Omega, \Theta, S \subseteq Z_\Delta^+$ represent the $n$-dimensional input, noise, output space, respectively. Note that $S \triangleq \Omega \oplus \Theta$, where $\oplus$ refers to the sum of elements with the same dimension. The general discrete random noise-adding mechanism $A : \Omega \rightarrow S$ is given by

$$A(x) = g(x) + h(\vartheta), \forall x, \vartheta \in \mathbb{R}^n, g(x) \in \Omega, h(\vartheta) \in \Theta,$$

where $\vartheta = [\vartheta_1, \ldots, \vartheta_n]^T$. The term $g(\cdot)$ is a function of $x$ satisfying $|g(x) - g(y)| \leq K |x - y|$ (where $K$ is a constant) and $g(x) \neq g(y)$ when $x \neq y$, similar to the continuous Lipschitz property. The arbitrary form of the function allows the inputs to be post-processed. Besides, we define a discretization function $h : \mathbb{R}^n \rightarrow Z_\Delta^n$, which maps the continuously distributed noise to the discrete one. It can avoid cases where the interpretability and validity of the original discrete data are destroyed due to the continuously distributed noises. In this paper, we propose a general discretization method for the variable $\vartheta_i$ in every dimension, which is shown as:

$$p_{\vartheta_i}(k) = \int_{k}^{k+\Delta} f(\vartheta_i) \, d\vartheta_i, \quad i \in V, \quad k \in Z_\Delta.$$  

The term $f(\vartheta_i)$ refers to the probability density function of the original continuous random noise. The term $p_{\vartheta_i}(k)$ is the probability of the discretized random variable $\vartheta_i$ when $\vartheta_i = k$.

In summary, the discrete random noise-adding mechanism $A$ represents the process of adding discrete random noise to the discretely distributed data. We abbreviate it as the discrete random mechanism $A$. Besides, we distinguish the mechanisms based on the discrete noise distributions. For instance, we denote the approach to perturb the data by adding discrete Gaussian distributed random noise as the Gaussian mechanism. Similarly, we define the Laplacian mechanism, the staircase mechanism, and the Exponential mechanism.  

1 One limitation of the mechanism is that the original data and the random noise are independent due to the unrelated functions $g$ and $h$. This condition is easily met, as we can decouple the data from the noise by computer or design the specific noise independent of the data. So the assumption will not limit the mechanism’s practical applicability and generalization.
B. Background on Differential Privacy

In this subsection, we introduce the differential privacy (DP) properties for the discrete random mechanism \( \mathcal{A} \). A differential private mechanism means it realizes the privacy protection of the discrete data measured by differential privacy.

First, we adopt the adjacency definition to illustrate the protected data. Consider two \( n \)-dimensional data that differ only in one dimension. To protect the privacy of specific data attributes and generalize the neighborhood property, we focus on the value of the record rather than its presence. Based on this observation, we give the definition of \( m \)-adjacency for two discrete vectors which has also been investigated in [6], [7], [8], [9], [10].

**Definition 1:** (\( m \)-adjacency). Given \( m \in \mathbb{Z}_+ \), the pair of vectors \( x, y \in \mathbb{Z}_m^d \) is \( m \)-adjacent if, for a given \( i_0 \in V \), we have
\[
\forall i \in V, |x_i - y_i| \leq \begin{cases} 
    m, & i = i_0; \\
    0, & i \neq i_0.
\end{cases}
\]  

The adjacency characterizes the sensitivity in the original DP definitions [3]. The sensitivity is usually given by \( \Delta f = \max_{x \neq y} \| f(x) - f(y) \| \), where \( f \) is a query function. Note that the absolute value in our adjacency can be converted into the \( L_1 \) or \( L_2 \) norm. Besides, when focusing on the query results of the datasets, the sensitivity can be expressed in the form of adjacency. More importantly, we obtain that the pair of \( m \)-adjacent vectors has the same size and differs only in one record with the same dimension, where the difference is no more than \( m \). Hence, we focus on the specific dimension in the neighboring vectors rather than the properties of all the dimensions. Thus, the multi-dimensional problem is transformed into a one-dimensional problem.

Next, we present the adjacency-based definition of \( (\epsilon, \delta) \)-differential privacy for a discrete random mechanism \( \mathcal{A} \).

**Definition 2:** \( (\epsilon, \delta) \)-differential privacy. A discrete random mechanism \( \mathcal{A} \) is \( (\epsilon, \delta) \)-DP if for any pairs of \( m \)-adjacent vectors \( x \) and \( y \), and for all \( O \subseteq S \), we have
\[
\Pr \{ \mathcal{A}(x) \in O \} \leq e^\epsilon \Pr \{ \mathcal{A}(y) \in O \} + \delta.
\]

Intuitively speaking, a DP mechanism will not reveal more than a fixed amount of data information. The adjacency-based definition will not lead to weaker privacy guarantees in the majority of cases as the adjacency bounds the data differences in any dimension. It implies that the output of each dimension is equally bounded or even consistent in a probabilistic sense. Then, the privacy leakage of the overall mechanism output is limited, which is obtained by the cumulative multiplication of the probability values of every dimension. However, due to the randomness of the noise, we can only guarantee the desired privacy protection level with a high probability. The privacy guarantees can be quantized by two privacy parameters \( \epsilon \) and \( \delta \). The positive number \( \epsilon \) measures the privacy maintained by the mechanism, and the term \( e^\epsilon \) quantifies the privacy loss across the mechanism outputs. For cases where the upper bound \( \epsilon \) does not hold, the parameter \( \delta \) functions to compensate for outputs by allowing a small probability of error. Smaller \( \epsilon \) and \( \delta \) represent more robust privacy-preserving capability.

**Remark 1:** The adjacency-based DP definition can be transformed with other DP definitions. For example, the classical DP definition protects the presence of data [3]. The presence and absence are characterized by binary variables, with 0 and 1 representing the presence of data in both datasets and only one dataset, respectively. It is equivalent to the case of 1-adjacency under \( Z_1^d \). Besides, the adjacency-based DP definition is also interchangeable with the DP definitions considering queries. Consider two vectors of medical records \( x, y \), where each record is a Boolean denoting whether a person has diabetes or not. The two vectors are neighboring with only one distinct record. A malicious user executes the counting query \( Q \), finding that \( \| Q(x) - Q(y) \| = 1 \). Then, the vectors are 1-adjacent as the emphasis is placed on the query results rather than the original data. More importantly, the adjacency-based DP definition generalizes the neighboring property. It implies that we can adapt the adjacency to any data attributes we are interested in. The data difference can be any positive value \( m \), not just a fixed value only reflecting the presence or absence of data.

C. Wasserstein Distance

In this subsection, we present a general definition of the utility metric, Wasserstein distance. It reflects the similarity of the statistical properties of the mechanism’s input and output.

**Definition 3:** \( (p \)-Wasserstein distance). The \( p \)-Wasserstein distance between two probability measures \( u \) and \( v \) on \( \mathbb{R}^d \) is
\[
W^p(u, v) = \inf_{X \sim u, Y \sim v} (\mathbb{E} \| X - Y \|^p)^{\frac{1}{p}}, \quad p \geq 1,
\]

where \( X \) and \( Y \) are two \( d \)-dimensional random vectors distributed as \( u \) and \( v \). The infimum is taken over all joint distributions of the random variables \( X \) and \( Y \).

Intuitively, the distance \( W^p(u, v) \) is the minimal effort required to reconstruct \( u \)'s mass distribution into the \( v \)'s. The effort is quantified by moving every unit of mass from \( x \) to \( y \) with the cost \( \| x - y \|^p \). In this paper, we focus on the special case of 1-Wasserstein distance on \( \mathbb{R}^d = 1 \). By referring to [43], an explicit formula is given by
\[
W^1(X, Y) = \int_\mathbb{R} |F_X(t) - F_Y(t)| \, dt,
\]

where \( F_X(\cdot) \), \( F_Y(\cdot) \) are the cumulative distribution functions (CDF) of continuous random variables \( X \) and \( Y \), respectively.

IV. MAIN RESULTS

In this section, we first propose the DP conditions for the discrete random mechanism \( \mathcal{A} \), followed by the computation methods for the DP parameters \( \epsilon \) and \( \delta \). Next, we analyze the DP properties for five representative mechanisms. Then, we consider the trade-off between privacy level and utility, deriving a \( \epsilon \)-DP mechanism with the maximum utility.
In this paper, we consider the discrete-distributed input and the added noise have already been post-processed\(^2\). The simplified discrete random mechanism \(A\) is rewritten as:

\[
A(x) = x + \theta,
\]

where \(x \in \Omega \subseteq \mathbb{Z}_\Delta^n\), \(\theta \in \Theta \subseteq \mathbb{Z}_\Delta^n\).

A. DP Conditions and Parameter Computation

In this subsection, a necessary and sufficient condition for the \(\epsilon\)-DP mechanism and a sufficient condition for the \((\epsilon, \delta)\)-DP mechanism are given by Theorem 1 and Theorem 2, respectively, with the closed-form expressions for DP parameters. First of all, we consider the \(\epsilon\)-DP conditions.

**Theorem 1:** The discrete random mechanism \(A\) satisfies \(\epsilon\)-DP if and only if there exists a positive constant \(c_0\) such that

\[
\sup_{\forall m_i \in [-m, m], m_0 \in \mathbb{Z}_\Delta, \forall i \in V} \frac{p_{\theta_i}(k - m_0)}{p_{\theta_i}(k)} = c_0,
\]

where \(k \in \mathbb{Z}_\Delta\). Moreover, we have that \(c_0\) is an increasing function of \(m\). The privacy parameter \(\epsilon\) is specified by

\[
\epsilon = \log(c_0).
\]

The proof of Theorem 1 is given in Appendix A.

Remark 2: We explore the similarities and differences between the discrete DP conditions in Theorem 1 and the continuous results in [10]. First, the essence of the criteria for the \(\epsilon\)-DP mechanisms is the same. That is, any adjacent probability ratio for the noise probability distribution should have an upper bound \(c_0\). In other words, the privacy loss \(e^\epsilon\) is finite. Furthermore, the discrete conditions are the simplification of the continuous ones. For the continuous random noise distributions, due to the uncountability of the real number set, the potential infinite local maximum and minimum should be considered in any given interval. Nevertheless, for discrete probability distributions, we only focus on whether the probability value at any single point is zero (as described in (9)). The difference shows that the DP parameter \(\epsilon\) is highly related to how we discretize a continuous probability distribution.

In summary, Theorem 1 helps verify whether a discrete random mechanism \(A\) is \(\epsilon\)-DP or not, only relying on the properties of the discrete noise probability distribution. This idea is distinct from the existing work [44], which goes through all the mechanisms outputs to validate the DP properties for the mechanisms by DP definition.

Next, we consider a more relaxed notion, \((\epsilon, \delta)\)-DP. We propose a sufficient condition to verify the \((\epsilon, \delta)\)-DP properties, along with the expressions for the DP parameters \(\epsilon\) and \(\delta\).

**Theorem 2:** Let \(\Theta \subseteq \mathbb{Z}_\Delta^n\) be the set of discrete random variables \(\theta\). Suppose that \(\Theta_0\) and \(\Theta_1\) are two subsets of \(\Theta\), which satisfies \(\Theta = \Theta_0 \cup \Theta_1\) and \(\Theta_0 \cap \Theta_1 = \emptyset\). Assume

\[
\sum_{\theta \in \Theta_0} p_{\theta}(k) \leq \delta,
\]

and the condition (7) holds when \(\theta \in \Theta_1\), that is,

\[
\sup_{\forall m_i \in [-m, m], m_0 \in \mathbb{Z}_\Delta, \forall i \in V} \frac{p_{\theta_i}(k - m_0)}{p_{\theta_i}(k)} = c_0,
\]

where \(\forall i \in V, k \in \mathbb{Z}_\Delta\). Then the discrete random mechanism \(A\) is \((\epsilon, \delta)\)-DP, and the privacy parameter \(\epsilon\) is given by

\[
\epsilon = \log(c_0).
\]

The proof of Theorem 2 is given in Appendix B.

Theorem 2 is applied for the mechanisms where the \(\epsilon\)-DP conditions cannot be strictly met. The main idea is to limit the probability values that exceed the upper bound \(c_0\) in the range \(\delta\), then the rest of the probability distributions still provide a privacy guarantee.

To further explain the rationale of Theorem 2, we consider the extreme limitations of the two DP parameters \(\epsilon\) and \(\delta\).

- \(\Theta_1 \rightarrow \Theta\) and \(\Theta_0 \rightarrow \emptyset\). It evolves in the \(\epsilon\)-DP since

\[
\delta = \lim_{\Theta_0 \rightarrow \emptyset} \sum_{\theta \in \Theta_0} p_{\theta}(k) = 0,
\]

that is, Theorem 1 is satisfied.

- \(\Theta_0 \rightarrow \Theta\) and \(\Theta_1 \rightarrow \emptyset\). Then, we have

\[
\delta = \lim_{\Theta_1 \rightarrow \emptyset} \sum_{\theta \in \Theta_0} p_{\theta}(k) = 1
\]

and

\[
c_0 = \lim_{\Theta_1 \rightarrow \emptyset} \sup_{\forall m_i \in [-m, m], m_0 \in \mathbb{Z}_\Delta, \forall i \in V} \frac{p_{\theta_i}(k - m_0)}{p_{\theta_i}(k)} = 1,
\]

thus \(\epsilon = \log(c_0) = 0\). Substituting \(\epsilon\) and \(\delta\) into (4), we have \(\text{Pr}\{A(x) \in O\} \leq \text{Pr}\{A(y) \in O\} + 1\). Then, one implies that any mechanism \(A\) satisfies \((0, 1)\)-DP.

Note that only discussing the limitations of DP parameters in the second case is not sufficient since it can be applied to arbitrary discrete random mechanisms. Thus, it is worth specifying tight bounds of \(\epsilon\) and \(\delta\) for every discrete random mechanism, which will be further discussed in Section IV-B.

Remark 3: The DP conditions derived from the adjacency-based DP definition are suitable for those from the classical privacy definitions. For instance, denote adding or removing one record as unbounded-DP (a classical choice) and changing exactly one record as bounded-DP (our choice). Under the DP guarantees, \(-\)bounded-DP implies \(2\epsilon\)-bounded DP, as changing a record can be seen as removing it and adding a new one in the right place [45]. In other words, the conclusions

\(\)
derived from different definitions will be similar, along with the exact values of the privacy parameters that can be converted to each other. The transferability implies that the analysis in Theorem 1 and Theorem 2 can promote relevant research on more general cases. □

B. DP Properties and Privacy Guarantees

In this subsection, we apply the obtained conditions to discuss the DP properties for two kinds of discrete random mechanisms. The first kind is achieved by adding discrete noises discretized from the continuous ones. Here, four representative mechanisms are selected, that is, the Gaussian, the Laplacian, the staircase, and the Uniform mechanisms. The second kind is obtained by adding noises that are originally discretely distributed. The Exponential mechanism is representative. For each mechanism, we derive the DP properties (ε-DP or (ε, δ)-DP) and the DP parameters.

For the first kind of mechanism, we discretize continuous probability density functions (PDF) of the noises to obtain the discrete probability mass functions (PMF) with the proposed discretization method\(^3\) shown in (2).

1) Gaussian mechanism:

The PMF of the discrete Gaussian distribution shown in Fig. 1 is given by

\[
p_{\theta_i}(k) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(k-m)^2}{2\sigma^2}}, \quad k \in \mathbb{Z}_\Delta, \tag{13}
\]

where the parameters \(\mu\) and \(\sigma\) are the mean and the standard deviation of the original continuous distribution, respectively.

Based on Theorem 1, one implies that the Gaussian mechanism is not \(\epsilon\)-DP. There exist intervals, such as \(k \geq \mu + m_0\), that as \(k\) increases, the lower bound of the probability ratio

\[
\frac{p_{\theta_i}(k-m_0)}{p_{\theta_i}(k)} = \frac{\int_{k-m_0}^{k+\Delta} e^{-\frac{(z-m_0)^2}{2\sigma^2}} dz}{\int_{k}^{k+\Delta} e^{-\frac{(z-m_0)^2}{2\sigma^2}} dz} \\
\geq \frac{\Delta e^{-\frac{1}{2\sigma^2}((k+\Delta-m_0-\mu)^2)}}{\Delta e^{-\frac{1}{2\sigma^2}(k-\mu)^2}} = e^{\frac{1}{2\sigma^2}(m_0-\Delta)(2k+\Delta-m_0-2\mu)}
\]

goes infinity. For these cases, the privacy loss is not bounded. So, we consider the relaxed DP properties. Based on Theorem 2, we have that the Gaussian mechanism \(\mathcal{A}\) is \((\epsilon, \delta)\)-DP. The two DP parameters \(\epsilon\) and \(\delta\) are specified by

\[
\epsilon = \frac{1}{2\sigma^2} (m-\Delta)(2M+\Delta-m-2\mu) \tag{14}
\]

and

\[
\delta = 2 \sum_{k=M}^{+\infty} p_{\theta_i}(k) = \frac{2}{\sqrt{2\pi}\sigma} \sum_{k=M}^{+\infty} \int_{k}^{k+\Delta} e^{-\frac{(z-m_0)^2}{2\sigma^2}} dz, \tag{15}
\]

where \(M\) is an arbitrarily large constant.

Remark 4: The DP properties derived from the conditions for the general Gaussian mechanism under the adjacency-based DP definition in (4) are similar to the existing conclusions. First, it is proved in [34] that the discrete Gaussian mechanism only preserves \((\epsilon, \delta)\)-DP despite different discretization methods, as the DP parameter

\[
\delta = \text{Pr} \left[ \theta_i > \frac{e\sigma^2}{\Delta} \right. - \frac{\Delta}{2} \left. - \epsilon \cdot \text{Pr} \left[ \theta_i > \frac{e\sigma^2}{\Delta} \right. + \frac{\Delta}{2} \right]
\]

is always positive. Then, we discuss the similarity of the upper bounds on the permissible privacy probability error \(\delta\). Set the adjacency \(m = 1\) (the same as the sensitivity in [34]), the parameter \(\delta\) proposed in [34] is

\[
\delta \leq \frac{1}{\sqrt{2\pi}\sigma^2} e^{-2M^2/\sigma^2}. \tag{16}
\]

Meanwhile, the parameter based on our results (15) is

\[
\delta \leq \frac{2}{\sqrt{2\pi}\sigma^2} \sum_{k=M}^{+\infty} \int_{k}^{k+1} e^{-\frac{k^2}{2\sigma^2}}. \tag{17}
\]

The parameter bounds shown in (16) and (17) are slightly different. However, they are almost the same under a large standard deviation \(\sigma\). It is caused by the distinct discretization methods. The discrete Gaussian distribution in [34] is a natural analog of the continuous Gaussian, that is,

\[
p_{\theta_i}(k) = e^{-(k-\mu)^2/2\sigma^2} \sum_{k' \in \mathbb{Z}} e^{-(k'-\mu)^2/2\sigma^2}, \tag{18}
\]

which is a more accurate but complicated discretization approach than (2). In brief, the analysis of DP properties is more general as it relies less on the noise probability distribution. Meanwhile, the similarity implies that the conclusion derived from the adjacency-based and classical DP definitions is convertible. □

2) Laplacian mechanism: The discrete Laplacian distribution is discretized under (2) from the PDF \(f(z) = \frac{1}{2\lambda} e^{-\frac{|z-\mu|}{\lambda}}\).

With simplification, we have

\[
p_{\theta_i}(k) = \begin{cases} 
1 - e^{-\Delta/\lambda} e^{\frac{k-\lambda}{\lambda}}, & k \geq \mu; \\
\frac{2}{e^{\lambda/2} - 1} e^{-\frac{k-\mu}{\lambda}}, & k \leq \mu - \Delta,
\end{cases}
\]

where \(k \in \mathbb{Z}_\Delta, \mu, \lambda\) are the same position and scaling parameters as the continuous distribution, respectively.
Theorem 1, for any mechanism guarantees the staircase-shaped distribution in Fig. 3 is obtained by DP properties under discrete scenarios. The PMF of the discrete maintaining the data utility, we are interested in exploring its equivalent to the ones derived from the classical definition. In [5], which is highly related to the scaling parameter \( \lambda \) strongly correlate with the parameters and properties of the Laplacian. So we conclude that the DP properties for the discrete mechanisms implies that the DP properties for the discrete mechanisms \( \epsilon \)-DP mechanism, where the DP parameter \( \epsilon \) is given by \( \epsilon = \log e^{\frac{m+\Delta}{\lambda}} \), (19) which is highly related to the scaling parameter \( \lambda \). Then, one implies that the DP properties for the discrete mechanisms strongly correlate with the parameters and properties of the specific discrete probability distributions.

Remark 5: The DP properties for the Laplacian mechanism derived from the adjacency-based DP definition in (4) are equivalent to the ones derived from the classical definition. In [5], the DP parameter is given by \( \epsilon = \frac{\Delta f}{\epsilon} \), where \( \Delta f \) represents the sensitivity. It is equivalent to the expression (19) since the adjacency characterizes the sensitivity and \( \Delta \) is an arbitrarily small discretization interval.

3) Staircase mechanism: Since the study in [27] pointed out that the continuous \( \epsilon \)-DP staircase mechanism performs best in maintaining the data utility, we are interested in exploring its DP properties under discrete scenarios. The PMF of the discrete staircase-shaped distribution in Fig. 3 is obtained by

\[
p_{\theta_i}(k) = \begin{cases} 
\frac{1 - \rho}{2a} \rho^k, & \text{if } ja\Delta \leq k < (j+1)a\Delta; \\
\frac{1 - \rho}{2a} \rho^k, & \text{if } -(j+1)a\Delta \leq k < -ja\Delta, 
\end{cases}
\]

So we conclude that the Laplacian mechanism \( A \) is a discrete \( \epsilon \)-DP mechanism, where the DP parameter \( \epsilon \) is specified by

\[
\epsilon = \log e^{\frac{m+\Delta}{\lambda}},
\]

which is highly related to the scaling parameter \( \lambda \). Then, one implies that the DP properties for the discrete mechanisms strongly correlate with the parameters and properties of the specific discrete probability distributions.

4) Uniform mechanism: The PMF of the discrete Uniform noise added

\[
p_{\theta_i}(k) = \frac{\Delta}{b-a+\Delta},
\]

where \( k \in \mathbb{Z}_\Delta, a \in \mathbb{Z}^+, \rho \in \{ x \mid 0 < x < 1, x \in \mathbb{R} \}, j \in \mathbb{N} \). Here \( a \) and \( \rho \) represent the width and height of the staircase-shaped distribution, respectively.

It is easily obtained that the staircase mechanism is \( \epsilon \)-DP, consistent with the result in [27]. For any given adjacency \( m \), where \( m \in \{ m_0 \mid c_0\Delta < m_0 \leq (c+1) a\Delta, a, c \in \mathbb{Z}^+ \} \), there exists an upper bound \( c_0 \) satisfying

\[
\sup_{\forall m_0 \in [-m, m], m_0 \in \mathbb{Z}_\Delta} \frac{p_{\theta_i}(k-m)}{p_{\theta_i}(k)} = \frac{1}{\rho^\Delta} = c_0.
\]

Then, the corresponding DP parameter \( \epsilon \) is shown as

\[
\epsilon = \log(c_0) = \log(\rho^\lceil \frac{m}{\Delta} \rceil),
\]

where the term \( \lceil t \rceil \) represents the smallest integer no less than \( t \).

5) Exponential mechanism: Unlike the above four discrete random mechanisms, the Exponentially distributed noise added
in the mechanism is inherently discretely distributed. The PMF of the discrete Exponential noise in Fig. 5 is given by
\[ p_{\theta}(k) = \eta e^{-\eta k}, \quad k \in \mathbb{Z}_+, \] (23)
where \( \eta \) is the rate parameter of the Exponential distribution.

According to Theorem 1, one implies that the Exponential mechanism is \( \epsilon \)-DP. Since for any \( m_0 \in [-m,m] \), \( m_0 \in \mathbb{Z}_\Delta \) and \( k > m_0 \), we have the bounded probability ratio
\[ \left| \frac{p_{\theta+m_0}(k)}{p_{\theta}(k)} \right| = \frac{\eta e^{-\eta(k-m_0)}}{\eta e^{-\eta k}} = e^{\eta m_0} < e^{\eta m}. \]

Meanwhile, the DP parameter \( \epsilon \) is
\[ \epsilon = \log e^{\eta m} = \eta m. \] (24)

Thus, we conclude that the privacy cost \( \epsilon \) is proportional to the adjacency \( m_0 \), consistent with the result in [46].

**Remark 6:** The Exponential mechanism in this paper is slightly different from the existing one [46]. The main difference is that the discrete data we protect are numerically represented while the other protects non-numerical ones. For one thing, the commonly adopted Exponential mechanism \( A_{u,\epsilon} \) with the quality score function \( u(\cdot) \) and the privacy parameter \( \epsilon \) is given by
\[ A_{u,\epsilon}(x,m) \sim e^{\frac{u(x,m)}{\epsilon}}. \]

For another, the Exponential mechanism \( A \) in this paper is
\[ A(x) = x + \theta, \quad \theta \sim \text{Exp}(\eta), \]
where \( \text{Exp}(\eta) \) denotes the Exponential distribution in (23). No matter whether the data are numerical or not, both Exponential mechanisms protect data under the \( \epsilon \)-DP guarantees.

In summary, the DP properties for the typical mechanisms are listed in Table II. We conclude that the discrete random mechanisms can preserve any given adjacency-based DP levels with the design of noise probability distributions.

### C. The Optimal DP Mechanism

Based on the DP conditions, we further study the trade-off between privacy level and utility. In this subsection, a utility-maximization (Wasserstein distance-minimization) problem is formulated, subject to the DP constraints. Then, we give an equivalent linear programming form of the problem to make the non-convex optimization solvable. Ultimately, we derive an optimal \( \epsilon \)-DP staircase mechanism.

First, we model the input data as a random variable generated by a given probability distribution. We denote \( x, \theta, x + \theta \) as the random variables of the mechanism input, added noise, and output, respectively. In a probabilistic sense, the same input data have the same probability of output after the mechanism (shown in (39)). So, we focus on the \( i_0 \)-th dimension data to be protected. For a concise expression, we omit the subscript \( i_0 \). Then, all three random variables are one-dimensional variables, with the corresponding PMFs as \( p_x(\cdot), p_\theta(\cdot), p_{x+\theta}(\cdot) : \mathcal{Z}_\Delta \rightarrow \mathbb{R} \). Based on the convolution sum in the probability theory, the PMF is computed as
\[ p_{x+\theta}(k) = \sum_i p_x(i) p_\theta(k-i), \quad k \in \mathbb{Z}_\Delta. \] (25)

Then, we construct a primal utility optimization problem.

**Utility model:** We build the utility model as a minimization framework. We aim to maximize the mechanism utility by maintaining the similarity between input and output probability distributions as much as possible. Extend the Wasserstein distance based on the continuous distributions in (5) to the discrete Wasserstein distance, we have
\[ W_0(x,x+\theta) = \sum_{k \in \mathbb{Z}_\Delta} |P_x(k) - P_{x+\theta}(k)|, \]
where \( P(\cdot) \) is the CDF for discrete random variables. The objective is to minimize the Wasserstein distance, that is,
\[ \min_{p_\theta(k)} W_0(x,x+\theta). \] (26)

Note that the optimization variable is the noise PMF \( p_\theta(\cdot) \).

**Constraints:** The major constraint is the \( \epsilon \)-DP guarantees. Once the privacy cost \( \epsilon \) is given, based on Theorem 1, we have
\[ c_\theta = \sup_{m_0 \in [-m,m], m_0 \in \mathbb{Z}_\Delta} \frac{p_{\theta}(k-m_0)}{p_{\theta}(k)} = e^\epsilon. \] (27)
Then, to avoid the cases where two negative elements lead to a positive result, the second constraint is given by
\[ p_\theta(k) > 0, \quad \forall k \in \mathbb{Z}_\Delta. \] (28)

The last constraint is that the optimization variable should satisfy the basic properties of probability, that is,
\[ \sum_{k \in \mathbb{Z}_\Delta} p_\theta(k) = 1. \] (29)

**Optimization:** Combining (26)–(29), we formulate the following primal optimization problem:
\[
\begin{align*}
\text{P}_0 : \quad & \min_{p_\theta(k)} W_0(x,x+\theta) \\
\text{s.t.} \quad & \epsilon \leq \log(c_\theta); \\
& \sum_{k \in \mathbb{Z}_\Delta} p_\theta(k) = 1; \\
& p_\theta(k) > 0, \forall k \in \mathbb{Z}_\Delta.
\end{align*}
\] (30a, 30b, 30c, 30d)

To solve the non-convex optimization problem \( \text{P}_0 \), we propose an approximate conventional convex optimization problem \( \text{P}_2 \). First, we introduce the input probability distribution
\[ p_x = (\ldots, p_x(k-\Delta), p_x(k), p_x(k+\Delta), \ldots)^T \] (31)
and the noise probability distribution

\[ p_\theta = \left(\ldots, p_\theta(k-\Delta), p_\theta(k), p_\theta(k+\Delta), \ldots\right)^T, \]  

where \( k \in \mathbb{Z}_\Delta \) and \( p_x(k), p_\theta(k) \) denote the probability value at the point \( k \) of the input and the noise, respectively. With these two notions, we give an equivalent problem \( P_1 \), where the equivalence is proved in Theorem 3. 

\[ P_1 : \min_{p_\theta} W_1(x, x + \theta) \]  

\[ \text{s.t. } A p_\theta \leq b, \quad |p_\theta| = 1, \quad p_\theta > 0, \]  

where \( W_1(x, x + \theta) = \sum_k |p_x^T M_k p_\theta| \). The matrices \( M_k \) and \( A, b \) are shown in (47) and (50), respectively.

**Theorem 3:** The problem \( P_0 \) is equivalent to \( P_1 \), which means that they have the same optimal solutions. The proof of Theorem 3 is given in Appendix C.

Then, to make the optimization variable \( p_\theta(k) \) independent of the calculation with the absolute values, we adopt the absolute value inequality. It is shown as

\[ W_1(x, x + \theta) = \sum_k |p_x^T M_k p_\theta| \leq \sum_k |p_x^T M_k| p_\theta \equiv W_2(x, x + \theta). \]  

Finally, we obtain an approximate optimization problem with the standard linear programming form:

\[ P_2 : \min_{p_\theta} W_2(x, x + \theta) \]  

\[ \text{s.t. } A p_\theta \leq b, \quad |p_\theta| = 1, \quad p_\theta > 0, \]  

where the objective function \( W_2 \) is given in (33). We solve the problem with the Simplex Method [35]. We derive that the optimal discrete random mechanism is realized by the class of staircase-shaped probability distributions, as shown in Fig. 6. The parameters of the optimal distribution (that is, the height and the width of the stairs) are determined by the input distribution \( p_x \) involved in the objective function, as well as the privacy cost \( \epsilon \) and the adjacency parameter \( m \) involved in the constraints. The in-depth analysis of how three parameters affect the mechanism is provided in Section V.

**V. SIMULATION**

In this section, we validate the utility guaranteed by the optimal \( \epsilon \)-DP staircase mechanism.

**A. Simulation Scenario**

First, we briefly describe the mechanism inputs and the noises. Note that the mechanism randomness only comes from the random noise added to the mechanism. We generate the mechanism input with the designated discrete distribution and keep it constant in each simulation. The random noise is determined by the specific discrete random mechanism. In this subsection, we consider three \( \epsilon \)-DP mechanisms, that is, the Laplacian mechanism, the staircase mechanism with a lower stair width (\( a = 5 \)), and the one with a higher stair width (\( a = 20 \)). Once the DP parameter \( \epsilon \) and the adjacency parameter \( m \) are given, the noise distribution can be uniquely determined according to Table II. We set the minimum discrete interval \( \Delta = 1 \) in the following simulations.

**B. The Effect of Three Factors on the Mechanism Utility**

Next, we discuss how three parameters mentioned in Section IV-C affect the utility of the optimal \( \epsilon \)-DP mechanism.

- **Effect of input distribution \( p_x \)**

  We select the Gaussian and the Poisson distributions as two input examples. The discrete Gaussian PMF is a discretized result, with the mean parameter \( \mu \) and the variance \( \sigma^2 \):

  \[ p_x(k) = \int_k^{k+1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx, \quad k \in \mathbb{Z}_\Delta = 1. \]

  The discrete Poisson PMF with parameter \( \gamma \) is given by

  \[ p_x(k) = \frac{e^{-\gamma} \gamma^k}{k!}, \quad k \in \mathbb{Z}_\Delta = 1. \]

  Fig. 7 shows the optimal distributions with discrete Gaussian input and Poisson input with the DP constraints set by \( \epsilon = 2 \) and \( m = 15 \). Overall, the two optimal distributions are all staircase-shaped, with the height and width of the stairs influenced by the input distributions.

- **Effect of privacy cost \( \epsilon \)**

  We assume a moderate adjacency \( m = 15 \) and take Gaussian input (\( \mu=0, \sigma=10 \)) as an example. In the trade-off problem, higher privacy cost (a larger \( \epsilon \)) implies less utility loss, as reflected in smaller Wasserstein distance. This trend can be verified in Fig. 8. Notice that the result with Laplacian noise (green cross line) and staircase-shaped noise with smaller stair width (purple triangle line) are comparable, especially at the high privacy level (smaller \( \epsilon \)). This is because the two distributions are similar under the parameter settings subject to the same DP constraints. Besides, the optimality of our mechanism is better represented as \( \epsilon \) increases.

- **Effect of the adjacency \( m \)**

  With the DP parameter \( \epsilon \) set as 2, the effect of the adjacency on the Wasserstein distance is shown in Fig. 9. It is observed that the Wasserstein distance and the adjacency are positively correlated. If the elements in the two datasets differ significantly, then even with the optimal \( \epsilon \)-DP mechanism, the utility guarantee is limited. Since larger differences require noises of greater amplitudes, the mechanism utility is significantly sacrificed.

**Remark 7:** In Fig. 9, we compare the mechanism utility with the optimal staircase-shaped distribution (red star line) and two
standard staircase-shaped distributions with explicit parameters (purple triangle line and pink square line). Under the same $\epsilon$-DP guarantees, the standard fixed staircase mechanism performs slightly worse in maintaining the utility. The reason is that we couple the optimal noise probability distribution with the input distribution. Due to the arbitrariness of the inputs, we cannot give a closed-form expression of the optimal noise distribution independent of the input. Instead, we obtain the optimal mechanism without fixed parameters by solving an equivalent problem through linear programming.

C. Verification of the Mechanism Optimality

To further validate the mechanism optimality, we compare the statistical properties of the mechanism utility with three other mechanisms under the same $\epsilon$-DP guarantees. To eliminate the uncertainty of the discrete random noises, we conduct 100 simulation runs for each simulation and conduct statistics on the mechanism utility. Notice that a smaller Wasserstein distance implies a higher mechanism utility.

Fig. 10 shows the utility of four $\epsilon$-DP mechanisms with different privacy levels and adjacency settings. Besides, we summarize the average, maximum, and minimum Wasserstein distance. Overall, our proposed mechanism has a higher probability of better utility. However, there are some exceptions. For example, with higher privacy protection ($\epsilon = 0.5$), its utility is similar to the staircase mechanism; in the low privacy situation, the performance is close to the Laplacian mechanism. It is partially caused by the uncertainty of the random noise. In conclusion, the optimal $\epsilon$-DP staircase mechanism ensures the maximum mechanism utility in the vast majority of cases.

VI. CONCLUSION

We consider the DP conditions, properties, and the trade-off between the mechanism utility and privacy level for...
discrete random noise-adding mechanisms. For the general DP mechanisms, a necessary and sufficient condition for ε-DP and a sufficient condition for (ε, δ)-DP are derived, followed by the numerical parameter expressions. Then, we analyze the DP properties for several typical mechanisms. Furthermore, we take the Wasserstein distance as the utility metric and build the trade-off issue as a utility-maximization optimization problem. The proposed optimal mechanism is staircase-shaped, with the parameters depending on the mechanism inputs and the DP requirements. Finally, extensive simulations are performed to verify the mechanism optimality. Future directions include the DP conclusions extension for more individuals and the correlation with homomorphic encryption.

**APPENDIX A**

**PROOF OF THEOREM 1**

Proof: \(\Leftarrow\): We prove the necessity by contradiction. Assume that

\[
\sup_{\forall m_0 \in [-m, m]} \frac{p_{\theta_i}(k - m_0)}{p_{\theta_i}(k)} = \infty, \quad k \in \mathbb{Z}_\Delta.
\]

Then, for any given large constant \(M\), there exists \(k_0 \in \mathbb{Z}_\Delta\) and \(m_0 \in \mathbb{Z}_\Delta \setminus \{0\}\) such that

\[
\frac{p_{\theta_i}(k_0 - m_0)}{p_{\theta_i}(k_0)} \geq M.
\]

Construct a pair of \(m_0\)-adjacent vectors \(x_i, y_i \in \mathbb{Z}_\Delta\) satisfying

\[
\forall i \in V, \quad |x_i - y_i| \leq \begin{cases} m_0, & i = i_0; \\ 0, & i \neq i_0. \end{cases}
\]

Based on the discrete property of \(\mathbb{Z}_\Delta\) and the sign of \(m_0\), we divide \(m_0 \in [-m, m]\) into three parts: \([-\Delta, \Delta]\), \(\{j \mid 2\Delta \leq j \leq m, j \in \mathbb{Z}_\Delta\}\) and \(\{j \mid -m \leq j \leq -2\Delta, j \in \mathbb{Z}_\Delta\}\). Denote \(O \subseteq S\), where \(O_i\) is the set of the \(i\)-th column element in \(O\). Note that DP is guaranteed if (4) holds for any given \(O\). In the following three parts, we construct the output range \(O_{i_0}\) to derive the contradiction for the necessity proof.

- \(m_0 \in \{-\Delta, \Delta\}\)
  
  Define \(O_{i_0} = \{k \mid k = y_{i_0} + k_0\}\). From (6), we have
  
  \[
  \frac{\Pr\{A(x_{i_0}) \in O_{i_0}\}}{\Pr\{A(y_{i_0}) \in O_{i_0}\}} = \frac{p_{x_{i_0} + \theta_{i_0}}(k)}{p_{y_{i_0} + \theta_{i_0}}(k)} = \frac{p_{x_{i_0} + \theta_{i_0}}(y_{i_0} + k_0)}{p_{y_{i_0} + \theta_{i_0}}(y_{i_0} + k_0)} \leq M.
  \]

- \(m_0 \in \{j \mid 2\Delta \leq j \leq m, j \in \mathbb{Z}_\Delta\}\)
  
  Define \(O_{i_0} = \{k \mid y_{i_0} + k_0 \leq k \leq y_{i_0} + k_0 + m_0 - \Delta\}\). Since \(p_{\theta_{i_0}}\) is bounded, there exists a constant \(C \geq 1\), s.t.
  
  \[
  \sum_{k=k_0}^{k_0+m_0-\Delta} p_{\theta_{i_0}}(k) = \sum_{k=k_0}^{k_0+m_0-\Delta} p_{\theta_{i_0}}(k_0) + \sum_{k=k_0+m_0-\Delta}^{k_0} p_{\theta_{i_0}}(k) \leq C \cdot p_{\theta_{i_0}}(k_0).
  \]

  Then, one follows that
  
  \[
  \frac{\Pr\{A(x_{i_0}) \in O_{i_0}\}}{\Pr\{A(y_{i_0}) \in O_{i_0}\}} \leq C \cdot p_{\theta_{i_0}}(k_0).
  \]

  

  \[
  = \sum_{k=k_0}^{k_0+m_0-\Delta} p_{x_{i_0} + \theta_{i_0}}(k) \leq \sum_{k=k_0}^{k_0+m_0-\Delta} p_{x_{i_0} + \theta_{i_0}}(k_0) + \sum_{k=k_0+m_0-\Delta}^{k_0} p_{x_{i_0} + \theta_{i_0}}(k) \leq C \cdot p_{\theta_{i_0}}(k_0).
  \]

  Thus, we prove that (7) is a necessary condition for the \(\epsilon\)-DP mechanism \(A\).

\(\Rightarrow\): Next, we prove the sufficiency. Based on (6), we have

\[
\Pr\{A(x) \in O\} = \Pr\{x + \theta \in O\}
\]

\[
= \sum_{i=1}^{n} \Pr\{x_i + \theta_i \in O\} \prod_{i=1, i \neq i_0} \Pr\{x_i + \theta_i \in O\}.
\]

and

\[
\Pr\{A(y) \in O\} = \Pr\{y + \theta \in O\}
\]

\[
= \sum_{i=1}^{n} \Pr\{y_i + \theta_i \in O\} \prod_{i=1, i \neq i_0} \Pr\{y_i + \theta_i \in O\}.
\]

Due to \(x_i = y_i, i \neq i_0\), we have

\[
\prod_{i=1, i \neq i_0} \Pr\{x_i + \theta_i \in O\} = \prod_{i=1, i \neq i_0} \Pr\{y_i + \theta_i \in O\}.
\]

Besides, with the condition in (7), it follows that

\[
\Pr\{x_i + \theta_i \in O\} = \sum_{i=1, i \neq i_0} \Pr\{y_i + \theta_i \in O\}.
\]

Combining (37)–(40), it yields that

\[
\Pr\{A(x) \in O\} \leq c_6 \Pr\{A(y) \in O\} \leq e^{\log(\epsilon)} \Pr\{A(y) \in O\},
\]

which satisfies the definition of \(\epsilon\)-DP.
Furthermore, comparing (4) and (41), we can easily derive the DP parameter $\epsilon$, that is, $\epsilon = \log(c_b)$. From (7), we note that the upper bound $c_b$ relies on the adjacency $m_0$. When $m_1 \leq m_2$ holds, we have

$$\sup_{m_0 \in [-m_1,m_1]} p_{y_0+m_0}(k) \leq \sup_{m_0 \in [-m_2,m_2]} p_{y_0}(k),$$

where $m_0, m_1, m_2, k \in \mathbb{Z}_\Delta$. Hence, we refer that $c_b$ is an increasing function of $m$.

**APPENDIX B**

**Proof of Theorem 2**

**Proof:** Construct a pair of $m$-adjacent vectors $x$ and $y$ with $x_i = y_i + m$ and $x_i = y_i$ (when $i \neq i_0$), where $x_i, y_i \in \mathbb{Z}_\Delta$.

Then, we obtain the following result:

$$\Pr\{A(x) \in \mathcal{O}\} = \prod_{i=1}^{n} \Pr\{A(x_i) \in \mathcal{O}_i\}$$

$$= \Pr\{A(x_{i_0}) \in \mathcal{O}_{i_0}\} \prod_{i=1,i \neq i_0}^{n} \Pr\{A(x_i) \in \mathcal{O}_i\}$$

$$= \left[\sum_{\Theta_k} \sum_{\theta \in \Theta_k} \Pr\{A(y_{i_0}) \in \mathcal{O}_{i_0}\} \Pr\{A(y_0) \in \mathcal{O}_0\}\right]$$

$$+ \sum_{\Theta_k} \Pr\{A(y_{i_0}) \in \mathcal{O}_{i_0}\} \prod_{i=1,i \neq i_0}^{n} \Pr\{A(y_i) \in \mathcal{O}_i\}$$

$$\leq c_b \sum_{\Theta_k} \Pr\{A(y_{i_0}) \in \mathcal{O}_{i_0}\} \prod_{i=1,i \neq i_0}^{n} \Pr\{A(y_i) \in \mathcal{O}_i\}$$

where $k_0 \in \mathbb{Z}_\Delta$ and $n \in \mathbb{R}$. The term $n \to \infty$ implies that the finite input (43) can replace the infinite one (31).

First, we prove the equivalence of the two objective functions, $W_0$ and $W_1$. We aim to convert the CDF into PMF, which contains the optimization variables more explicitly.

- The input cumulative probability at point $k$: $P_x(k)$,

$$P_x(k) = p_x(-\infty) + \cdots + p_x(k - \Delta) + p_x(k).$$

Combining (43) and (44), we easily have

$$P_x(k) = \begin{cases} 0, & k \leq k_0; \\ 1, & k \geq k_0 + n \Delta. \\ \end{cases}$$

Thus, $P_x(k)$ is the cumulative probability at point $k$.

Then, for $k \in \left\{ k' \mid k_0 + \Delta \leq k' \leq k_0 + n \Delta, k' \in \mathbb{Z}_\Delta \right\}$, we have

$$P_x(k) = p_x(k_0 + \Delta) + \cdots + p_x(k - \Delta) + p_x(k) = \sum_{i=k_0+\Delta}^{k} p_x(i) \cdot \left[ \sum_{i=k_0+\Delta}^{k} p_x(i) \right]$$

where the column number of the matrix $M^k_i$ is given by

$$r = 1 + \left[ k - (k_0 + \Delta) \right]/\Delta.$$

The column number of the matrix $M^k_i$ depends on the noise distribution $P_{\theta}$. Specifically, we have $M^k_i = 0$ for $k < k_0 + \Delta$ and $M^k_i = 1$ for $k > k_0 + n \Delta$, where 0 and 1 are two matrices full of elements 0 and 1, respectively.

- The output cumulative probability at point $k$: $P_{x+\theta}(k)$,

$$P_{x+\theta}(k) = p_{x+\theta}(-\infty) + \cdots + p_{x+\theta}(k - \Delta) + p_{x+\theta}(k).$$

With the finite representation of the input distribution $p_x$ in (43), the output probability at point $k$ in (25) is reformulated as following:

$$p_{x+\theta}(k) = \sum_{i=k_0+\Delta}^{k_0+n\Delta} p_{x}(i) p_{\theta}(k-i), \; k, i \in \mathbb{Z}_\Delta.$$  (46)

Then, we substitute every element in (45) with (46) and make simplification towards $p_{x}$ and $p_{\theta}$. Consequently, we have

$$P_{x+\theta}(k) = \sum_{i=-\infty}^{k-k_0-(n+1)\Delta} \left( \sum_{j=k_0+n\Delta}^{k_0+n\Delta} p_{x}(j) p_{\theta}(i) \right)$$

$$+ \sum_{i=k-k_0-n\Delta}^{k-k_0-\Delta} \left( \sum_{j=k_0+n\Delta}^{k-k_0} p_{x}(j) p_{\theta}(i) \right)$$

**APPENDIX C**

**Proof of Theorem 3**

**Proof:** The key idea of the equivalence proof is to make the optimization variables $P_{\theta}$ explicitly involved in the problem. To make the derivation clearer, we represent the mechanism input probability distribution as a finite one:

$$p_x = \left( p_x(k_0 + \Delta), p_x(k_0 + 2\Delta), \ldots, p_x(k_0 + n\Delta) \right)^T,$$  (43)

where $k_0 \in \mathbb{Z}_\Delta$ and $n \in \mathbb{R}$. The term $n \to \infty$ implies that the finite input (43) can replace the infinite one (31).
where the column vector \([k_1]\) corresponds to the element at the corresponding position in the column vector \(p_\theta\), that is, \(p_\theta(k_1)\). We have \(k_1 = k - k_0 - n\Delta\), \(k_2 = k - k_0 - \Delta\). Before the column \([k_2]\), all elements in the matrix \(M_k^1\) are 1, and after the column \([k_2]\), all elements are 0. So far, we make the two CDFs at the point \(k\), \(P_x(k)\) and \(P_{x+\theta}(k)\), related to the input distribution \(p_x\) and the noise distribution \(p_\theta\). All the information at point \(k\) is contained in the matrix \(M_k\), where

\[
M_k = M_k^1 - M_k^2.
\]

Then, we have

\[
W_0 = \sum_k |P_x(k) - P_{x+\theta}(k)| = \sum_k |p_x^T(M_k^1 - M_k^2)p_\theta| = \sum_k |p_x^TM_kp_\theta| = W_1.
\] (48)

Next, we prove the equivalence of the constraints.

- The DP constraint in (30b)

We rewrite the constraint as an inequality constraint (detailed expression in (27)), that is,

\[
\forall k, \forall m_0 \in [-m, m], p_\theta(k - m_0) \leq c_b = e^\epsilon
\Rightarrow p_\theta(k - m_0) - e^\epsilon \cdot p_\theta(k) \leq 0.
\] (49)

Then, we transform (49) into a matrix form:

\[
\begin{bmatrix}
-k_1 & [k_2] & [k_3]
\end{bmatrix}
\begin{bmatrix}
-\epsilon \\
0 \\
\vdots \\
0 \\
0 \\
1 \\
-\epsilon \\
0 \\
0 \\
0 \\
\vdots \\
0 \\
0 \\
0 \\
\end{bmatrix}
\begin{bmatrix}
p_\theta \\
0 \\
0 \\
0 \\
0 \\
0 \\
-\epsilon \\
0 \\
0 \\
0 \\
0 \\
0 \\
-\epsilon \\
0 \\
\end{bmatrix}
\leq
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
\end{bmatrix}.
\] (50)

where \(k_1 = k - m, k_2 = k, k_3 = k + m\). The three columns correspond to the elements \(p_\theta(k_1), p_\theta(k_2), p_\theta(k_3)\), respectively. The row number of the matrix \(A_k\) is \(2m + 1\), the same as the number of constraints. The column number relies on the length of the noise distribution \(p_\theta\). Go through all the points \(k\), and put \(A_k\) together with the corresponding element \(k\), we have the equivalent form of the DP constraint:

\[
\epsilon = \log(c_b) \Leftrightarrow p_\theta(k - m_0) \leq e^\epsilon \Leftrightarrow A_p \leq b,
\] (51)

where \(A = (\cdots A_{k=0} A_{k=\Delta} A_{k=2\Delta} \cdots)^T\) and \(b = 0^T\).

- The total sum constraint in (30c)

\[
\sum_k p_\theta(k) = 1 \Leftrightarrow \sum_k [p_\theta(k)] = 1 \Leftrightarrow |p_\theta| = 1.
\] (52)

- The positive value constraint in (30d)

\[
\forall k, p_\theta(k) > 0 \Leftrightarrow (\cdots, p_\theta(\Delta) > 0, p_\theta(2\Delta) > 0, \cdots)^T
\Leftrightarrow (\cdots, p_\theta(\Delta), p_\theta(2\Delta), \cdots)^T > 0
\Leftrightarrow p_\theta > 0.
\] (53)

By now, we have given the equivalent form of three constraints in problem \(P_0\) with (51), (52) and (53), respectively.

Thus, with (48), (51)–(53), we obtain the whole equivalent form of the original optimization problem \(P_0\), that is,

\[
P_1: \min_{p_\theta} W_1(x, x + \theta) = \sum_k [p_x^T M_k p_\theta]
\quad \text{s.t. } A_p \leq b, \quad |p_\theta| = 1, \quad p_\theta > 0.
\]
QIN et al.: DIFFERENTIAL PRIVATE DISCRETE NOISE-ADDING MECHANISM: CONDITIONS, PROPERTIES, AND OPTIMIZATION 3547

[19] M. Sun, W. P. Tay, and X. He, “Toward information privacy for the internet of things: A nonparametric learning approach,” IEEE Trans. Signal Process., vol. 66, no. 7, pp. 1734–1747, Apr. 2018.

[20] K. Wei et al., “Federated learning with differential privacy: Algorithms and performance analysis,” IEEE Trans. Inf. Forensics Secur., vol. 15, pp. 3454–3469, 2020.

[21] Y. Wang, J. Lam, and H. Lin, “Consensus of linear multivariable discrete-time multiagent systems: Differential privacy perspective,” IEEE Trans. Cybern., vol. 52, no. 12, pp. 13915–13926, Dec. 2022.

[22] C. Zhang, J. Chen, J. He, and P. Cheng, "Privacy-preserving consensus-based energy management in smart grids," IEEE Trans. Signal Process., vol. 66, no. 23, pp. 6162–6176, Dec. 2018.

[23] B. Ding, J. Kulkarni, and S. Yekhanin, “Collecting telemetry data privately,” in Proc. Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 7371–7380.

[24] S. L. Garfinkel, J. M. Abowd, and S. Powazek, “Issues encountered deploying differential privacy,” in Proc. Workshop Privacy Electron. Soc., 2018, pp. 133–137.

[25] Z. Lv and F. Piccialli, “The security of medical data on internet based on differential privacy technology,” ACM Trans. Internet Techn., vol. 21, no. 3, pp. 1–18, 2021.

[26] M. Gupte and M. Sundararajan, “Universally optimal privacy mechanisms for minimax agents,” in Proc. ACM SIGMOD-SIGACT-SIGART Symp. Princ. Database Syst., 2010, pp. 135–146.

[27] Q. Geng and P. Viswanath, “The optimal noise-adding mechanism in differential privacy,” IEEE Trans. Inf. Theory, vol. 62, no. 2, pp. 925–951, Feb. 2016.

[28] K. Balan, L. Sankar, and A. D. Sarwate, “Robust privacy–utility tradeoffs under differential privacy and hamming distortion,” IEEE Trans. Inf. Forensics Secur., vol. 13, no. 11, pp. 2816–2830, Nov. 2018.

[29] Q. Geng, W. Ding, R. Guo, and S. Kumar, “Tight analysis of privacy and utility tradeoff in approximate differential privacy,” in Proc. Int. Conf. Artif. Intell. Statist., 2020, pp. 89–99.

[30] T. Asikis and E. Pournaras, “Optimization of privacy–utility tradeoffs under informational self-determination,” Future Gener. Comput. Syst., vol. 109, pp. 488–499, 2020.

[31] M. Kumar, M. Rossby, B. A. Moser, and B. Freudenthaler, “Deriving an optimal noise adding mechanism for privacy-preserving machine learning,” in Proc. Int. Conf. Database Expert Syst. Appl., 2019, pp. 108–118.

[32] A. Ghosh, T. Roughgarden, and M. Sundararajan, “Universally utility-maximizing privacy mechanisms,” SIAM J. Comput., vol. 41, no. 6, pp. 1673–1693, 2012.

[33] J. Chung, P. Kannappan, C. Ng, and P. Sahoo, “Measures of distance between probability distributions,” J. Math. Anal. Appl., vol. 138, no. 1, pp. 280–292, 1989.

[34] C. L. Canonne, G. Kamath, and T. Steinke, “The discrete Gaussian for differential privacy,” in Proc. Adv. Neural Inf. Process. Syst., vol. 33, 2020, pp. 15676–15688.

[35] J. A. Nelder and R. Mead, “A simplex method for function minimization,” Comput. J., vol. 7, no. 4, pp. 308–313, 1965.

[36] W. Wang, L. Ying, and J. Zhang, “On the relation between identifiability, differential privacy, and mutual-information privacy,” IEEE Trans. Inf. Theory, vol. 62, no. 9, pp. 5018–5029, Sep. 2016.

[37] P. Cuff and L. Yu, “Differential privacy as a mutual information constraint,” in Proc. ACM SIGSAC Conf. Comput. Commun. Secur., 2016, pp. 43–54.

[38] D. Desfontaines and B. Pejó, “Sok: Differential privacies,” Proc. Priv. Enhancing Technol., vol. 2, pp. 288–313, 2020.

[39] X. Xiong, S. Liu, D. Li, Z. Cai, and X. Niu, “A comprehensive survey on local differential privacy,” Secur. Commun. Netw., vol. 2020, no. 1–2, pp. 1–29, 2020.

[40] F. Liu, “Generalized Gaussian mechanism for differential privacy,” IEEE Trans. Knowl. Data Eng., vol. 31, no. 4, pp. 747–756, Apr. 2019.

[41] J. Dong, D. Durfee, and R. Rogers, “Optimal differential privacy-composition for exponential mechanisms,” in Proc. Int. Conf. Mach. Learn., 2020, pp. 2597–2606.

[42] A. Koskela, J. Jäkö, L. Prediger, and A. Honkela, “Tight differential privacy for discrete-valued mechanisms and for the subsampled Gaussian mechanism using FFT,” in Proc. Int. Conf. Artif. Intell. Statist., 2021, pp. 3358–3366.

[43] V. M. Panaretos and Y. Zemel, “Statistical aspects of Wasserstein distances,” Ann. Rev. Statist. Appl., vol. 6, pp. 405–431, 2019.

[44] C. Dwork et al., “The algorithmic foundations of differential privacy,” Found. Trends Theor. Comput. Sci., vol. 9, no. 3–4, pp. 211–407, 2014.

[45] D. Kifer and A. Machanavajjhala, “No free lunch in data privacy,” in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2011, pp. 193–204.

[46] F. McSherry and K. Talwar, “Mechanism design via differential privacy,” in Proc. 48th Annu. IEEE Symp. Found. Comput. Sci., 2007, pp. 94–103.