Integer Subspace Differential Privacy

Prathamesh Dharangutte[1], Jie Gao[1], Ruobin Gong[2], Fang-Yi Yu[3]

[1]Department of Computer Science, Rutgers University
[2]Department of Statistics, Rutgers University
[3]Department of Computer Science, George Mason University
{ptd39, jg1555, ruobin.gong}@rutgers.edu, fangyiyu@gmu.edu

Abstract

We propose new differential privacy solutions for when external invariants and integer constraints are simultaneously enforced on the data product. These requirements arise in real world applications of private data curation, including the public release of the 2020 U.S. Decennial Census. They pose a great challenge to the production of provably private data products with adequate statistical usability. We propose integer subspace differential privacy to rigorously articulate the privacy guarantee when data products maintain both the invariants and integer characteristics, and demonstrate the composition and post-processing properties of our proposal. To address the challenge of sampling from a potentially highly restricted discrete space, we devise a pair of unbiased additive mechanisms, the generalized Laplace and the generalized Gaussian mechanisms, by solving the Diophantine equations as defined by the constraints. The proposed mechanisms have good accuracy, with errors exhibiting sub-exponential and sub-Gaussian tail probabilities respectively. To implement our proposal, we design an MCMC algorithm and supply empirical convergence assessment using estimated upper bounds on the total variation distance via L-lag coupling. We demonstrate the efficacy of our proposal with applications to a synthetic problem with intersecting invariants, a sensitive contingency table with known margins, and the 2010 Census county-level demonstration data with mandated fixed state population totals.

1 Introduction

Motivation. Differential privacy (DP) is a formal mathematical framework that quantifies the extent to which an adversary can learn about an individual from sanitized data products. However, data curators may be mandated, by law or by policy, to release sanitized data products that obey certain externally determined constraints, in a manner that existing differential privacy solutions are not designed to address. Two challenging types of constraints are *invariants* (Ashmead et al. 2019), which are exact statistics calculated from the confidential data (such as sum, margins of a contingency table, etc), and *integer* characteristics, pertaining to data of count nature. The challenge is exacerbated when the sanitized data product is required to respect both.

Data privatization that simultaneously satisfies mandated invariants and preserves integral characteristics is of paramount importance to a myriad of data products by official statistical agencies. One of the prominent example is the Disclosure Avoidance System (DAS) of the 2020 Decennial Census of the United States. The Census Bureau is mandated to observe a set of invariants for its decennial census data products, such as its constitutional obligation to enumerate state populations for the apportionment house seats. As the decennial census data products are count data, the bureau must ensure that the sanitized release are also integer-valued. In addition, the bureau strives to maintain the consistency of the data products with common knowledge to various degrees, such as the the non-negativity of count data and possible logical relationships between different tabulated quantities. The challenge thus remains: how to design sanitized data products that respect the pre-specified constraints, while maintaining rigorous privacy and good statistical properties? This is the central problem we address in this paper.

Our contribution. In this paper, we develop the *integer subspace differential privacy* scheme, through which we formulate and implement mathematically rigorous privacy mechanisms that meet the following *utility objectives*: 1) respecting mandated invariants, 2) maintaining integral characteristics, and 3) achieving good statistical properties including *unbiasedness*, *accuracy*, as well as *probabilistic transparency* of the mechanism specification. Invariants not only restrict the universe of databases of interest, but also partially destroys the relevance of neighbors, since databases differing by precisely one entry may or may not meet the same invariants. To this end, we adapt the classical differential privacy definition to incorporate invariant constraints with two modifications (i) we limit the comparison of databases only to those that satisfy the same invariants, in a manner similar to pufferfish privacy (Kifer and Machanavajjhala 2014); and (ii) for databases that meet the same invariants, we use their metric distance as a multiplicative factor to the privacy loss parameter under the more general smooth differential privacy framework (Chatzikokolakis et al. 2013; Desfontaines and Pejó 2019). We show that integer subspace differential privacy preserves *composition* and *post-processing* properties.

To design a differential privacy mechanism that simultaneously respects invariants and integer requirement poses a number of new challenges. For additive mechanisms, the perturbation noise vector needs to be limited to the "null space" of the invariants so as not to violate the constraints. When

the invariants are linear, this requires solving a linear system with integer coefficients, i.e. Diophantine equations, limiting the privacy noise to a (transformed) discrete lattice space. To this end, we propose generalized discrete Laplace and Gaussian mechanisms for the discrete lattice space. Both mechanisms are transparently specified additive mechanisms that enjoy demonstrated unbiasedness and accuracy via bounds on tail probabilities. The tail bounds call for extra technicality as the lattice spaces we deal with are generally not spherically symmetric. To implement the proposed mechanisms, we design a Gibbs-within-Metropolis sampler, with a transition kernel that is always limited to the desired subspace to achieve sampling efficiency. To provide empirical guidance on the Markov chain's convergence to target, we supply an assessment scheme using the L-lag coupling method proposed by (Biswas, Jacob, and Vanetti 2019). We demonstrate the efficacy of our proposal with applications to a synthetic problem with intersecting invariants, a contingency table with known margins concerning delinquent children (Federal Committee on Statistical Methodology 2005), and the 2010 Census county-level demonstration data with mandated fixed state population totals.

Related work. There has been an extensive line of work on differential privacy and its alternative definitions; see a recent survey and the references therein (Desfontaines and Pejó 2019). On the other hand, invariants pose a relatively recent challenge to formal privacy, and we focus our review below on prior work concerning invariants. It has been recognized that invariants which are non-trivial functions of the confidential data compromise the differential privacy guarantee in its classic sense, and that the invariants must be incorporated into the privacy guarantee (Gong and Meng 2020; Seeman, Slavkovic, and Reimherr 2022). Towards designing formally private mechanisms that meet prescribed invariant constraints, the only prior work in this direction is named *subspace differential privacy* (Gao, Gong, and Yu 2022) where the inputs satisfy a set of linear constraints and the sanitized outcome shall meet the same invariants. However, subspace differential privacy only considers data products that are real-valued. To impose the additional integral requirement on the data outputs is not trivial: as the TopDown algorithm (Abowd et al. 2022) already demonstrates, simple rounding or projection will either violate the invariants or introduce biases. He, Machanavajjhala, and Ding (2014) consider known constraints of the dataset (e.g. population counts for each state) which restricts the possible set of datasets, but the output of their mechanism need not satisfy the known constraints. We focus on mechanisms whose output also satisfy the counting constraints. Another line of work (Cormode, Kulkarni, and Srivastava 2017) consider additional structural properties for the mechanism's output (e.g. monotone, symmetric, fair). However, those properties are independent of the private data and can be seen as internal distributional consistency. Lastly we note a related, but different line of previous work concerning *internal consistency* of the data outputs (Barak et al. 2007; Hay et al. 2009), such as maintaining the sum of counts over some partitioning of a set to be identical as the total sum. This type of consistency requirement is independent of the value of the private data, and can be satisfied by certain types of DP mechanisms, such as local DP schemes when each data item is independently perturbed and normal algebraic operations are applied on the perturbed output. In contrast, invariant constraints considered in this work are in general non-trivial functions of the confidential data.

To impose invariants and preserve integrality for Census data, the TopDown algorithm (Abowd et al. 2022) employs a discrete Gaussian mechanism (Canonne, Kamath, and Steinke 2020) to inject integer-valued noise during the so-called *measurement* phase, to form noise-infused counts to be tabulated into multi-way contingency tables at various levels of geographic resolution. This is followed by the so-called *estimation* phase, a constrained optimization to impose invariants (such as state population totals, etc) and controlled rounding to integer solutions (see Table 1 of Abowd et al. 2022). In general, this approach first imposes unconstrained noise and performs post-processing to meet the additional constraints. One limitation of this solution is that it is not easy to characterize the probabilistic distribution of the resulting privacy noise after post-processing, as this distribution crucially depends on the particular values of the input data. Thus, the post-processing approach may destroy the probabilistic transparency of the privatized data product (Gong 2022). Our solution maintains the transparency of the privacy mechanism as well as the statistical independence between the confidential data and the privacy errors, which are of paramount importance to downstream applications derived from these noisy data products. Furthermore, since invariants and integer value constraints are our first priority besides privacy, we supply a rigorously updated definition of differential privacy under this setting by limiting the discussion only among databases that satisfy these constraints.

2 Integer Subspace Differential Privacy

In this work, we model private data as a histogram $\vec{x} = (x_1, \dots, x_d)^\top \in \mathbb{N}^d$, where $x_i \in \mathbb{N}$ is the number of individuals with feature $i \in [d] := \{1, \dots, d\}$. A trusted curator holds the database $\vec{x}$, and provides an interface to the database through a randomized mechanism $M : \mathbb{N}^d \to \mathbb{Z}^d$ where we require the output to always take integer values. As motivated in introduction, our goal is to design good mechanisms whose output is close to the original histogram that satisfies not only certain privacy notions, but also invariant constraints and integral requirements.

The notion of differential privacy ensures that no individual's data has much effect on the probabilistic characteristic of the output from the mechanism M (Dwork et al. 2006; Barber and Duchi 2014). Formally, we say that a randomized mechanism $M : \mathbb{N}^d \to \mathbb{Z}^d$ is (ϵ, δ)-*differentially private* if for some $\epsilon, \delta \geq 0$, all neighboring databases $\vec{x}$ and $\vec{x}'$ with $\|\vec{x} - \vec{x}'\| = 1$, and all event $S \subseteq \mathbb{Z}^d$,

$$\Pr\left[M\left(\vec{x}\right) \in S\right] \leq e^{\epsilon} \cdot \Pr\left[M\left(\vec{x}'\right) \in S\right] + \delta, \qquad (1)$$

where $\|\vec{x} - \vec{x}'\|$ is a distance norm of two databases $(\vec{x}, \vec{x}')$. Depending on the application, the norm may be chosen either as ℓ_1 norm or ℓ_2 norm.

The notion of differential privacy in (1) satisfies a more general notion of *smooth differential privacy* (Chatzikokolakis et al. 2013). M is (ϵ, δ)-*smooth differentially private* if

for all datasets $\vec{x}, \vec{x}'$ and all event S ,

$$\Pr[M(\vec{x}) \in S] \leq e^{\epsilon\|\vec{x}-\vec{x}'\|} \Pr[M(\vec{x}') \in S] + \frac{e^{\epsilon\|\vec{x}-\vec{x}'\|}-1}{e^{\epsilon}-1}\delta. \quad (2)$$

Previous works mostly consider the pure smooth differential privacy with $\delta = 0$, and we generalize it to the approximated case.[1] Compared to (1), the definition in (2) replaces the neighboring relationship of $(\vec{x}, \vec{x}')$ by a norm function.

From the data curator's perspective, in addition to privacy concerns, there often exists external constraints that the privatized output M must meet. These constraints can often be represented as counting functions defined below.

Definition 1 (counting invariant constraints)**.** Given a collection of subsets $\mathcal{A} = \{A_1, \ldots, A_k\}$ on $[d]$ with $k \leq d$, we say a function $f : \mathbb{N}^d \to \mathbb{Z}^d$ is $\mathcal{A}$-invariant if

$$\sum_{i\in A} x_i = \sum_{i\in A} f(\vec{x})_i, \text{ for all } \vec{x} \in \mathbb{N}^n \text{ and } A \in \mathcal{A}.$$

Alternatively, given a collection of counting invariant constraints $\mathcal{A}$, we define an equivalence relationship $\equiv_{\mathcal{A}}$ so that $\vec{x} \equiv_{\mathcal{A}} \vec{x}'$ if $\sum_{i\in A} x_i = \sum_{i\in A} x'_i$ for all $A \in \mathcal{A}$. Then an $\mathcal{A}$-invariant f satisfies $\vec{x} \equiv_{\mathcal{A}} f(\vec{x})$ for all $\vec{x}$.

Note that when a privatized output is subject to counting invariant constraints as above, we may not use the standard neighboring relation because two neighboring datasets differing by one element may not satisfy the same counting constraint. In particular, the feasible outputs of an $\mathcal{A}$-invariant mechanism lie within an integer subspace. Therefore, we employ smooth differential privacy, and only control the privacy loss on datasets that satisfy the same counting constraints, i.e., only "secret pairs" of databases within the same equivalent classes. This formulation was also used in previous work such as pufferfish privacy (Kifer and Machanavajjhala 2014).

Definition 2. We say a mechanism M is (ϵ, δ)-differentially private on an equivalence relation $\equiv$ if for all equivalent pair $\vec{x} \equiv \vec{x}'$ and event E on the output space

$$\Pr[M(\vec{x}) \in E] \leq e^{\epsilon\|\vec{x}-\vec{x}'\|} \Pr[M(\vec{x}') \in E] + \frac{e^{\epsilon\|\vec{x}-\vec{x}'\|}-1}{e^{\epsilon}-1}\delta.$$

Moreover, given a collection of counting constraint $\mathcal{A}$, we say M is *$\mathcal{A}$-induced integer subspace differentially private* with *privacy loss budget* (ϵ, δ), if M is $\mathcal{A}$-invariant and (ϵ, δ)-differentially private on an equivalence relation $\equiv_{\mathcal{A}}$.

Many invariant constraints can be formulated using counting constraints. Below are some examples.

Example 3. Suppose $\mathcal{A} = \{[d]\}$ is a singleton. All datasets with the same number of agents are equivalent under $\equiv_{\{[d]\}}$, so the differential privacy on $\equiv_{\{[d]\}}$ reduces to the original (bounded) differential privacy.

[1]When $\delta = 0$, (2) reduces to $\Pr[M(\vec{x}) \in E] \leq e^{\epsilon\|\vec{x}-\vec{x}'\|} \cdot \Pr[M(\vec{x}') \in E]$. However, for the approximated smooth differential privacy, the error term grows as the distance $\|\vec{x} - \vec{x}'\|$ increases. When $\|\cdot\|$ is a norm, a mechanism on the histogram is (ϵ, δ)-DP if and only if the mechanism is (ϵ, δ)-smooth DP.

Example 4. Privatized census data products must ensure that the total population of each state is reported exactly as enumerated. Given k states, and $A_l \subset [d]$ be the collection of features that belong to state l. we can define counting invariants $\mathcal{A} = \{A_1, \ldots, A_k\}$ where $A_1, \ldots, A_k$ forms a partition on the universe $[d]$.

Example 5. We can encode the invariant condition on a $\sqrt{d} \times \sqrt{d}$ two-dimensional contingent table where the sum of each row and each column sums are fixed as a collection of counting constraints. Formally, given $\sqrt{d} \in \mathbb{N}$, the set $\mathcal{A}$ contains the following subsets on $[d]$, for all i and j $R_i = \{i\sqrt{d} + \ell : 0 \leq \ell < \sqrt{d}\}$ and $C_j = \{j + \ell\sqrt{d} : 0 \leq \ell < \sqrt{d}\}$. That is $\mathcal{A} = \{R_i, C_j : 0 \leq i, j < \sqrt{d}\}$.

Composition and post-processing properties for Definition 2 also hold due to similar arguments as the differential privacy, but with the additional requirement of equivalence over databases. Let $\equiv_{(\mathcal{A}_1,\mathcal{A}_2)}$ be the equivalence relation such that $x \equiv_{(\mathcal{A}_1,\mathcal{A}_2)} x'$ implies $x \equiv_{\mathcal{A}_1} x'$ and $x \equiv_{\mathcal{A}_2} x'$.

Proposition 6 (Composition)**.** *Let M_1 be (ϵ_1, δ_1)- differentially private on equivalence relation $\equiv_{\mathcal{A}_1}$ and M_2 be (ϵ_2, δ_2)- differentially private on equivalence relation $\equiv_{\mathcal{A}_2}$. For any pair of databases x, x' having $x \equiv_{(\mathcal{A}_1,\mathcal{A}_2)} x'$, the composed mechanism $M_{1,2}(x) = (M_1(x), M_2(x))$ is $(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$-differentially private on $\equiv_{\mathcal{A}_1,\mathcal{A}_2}$.*

Proposition 7 (Post-processing)**.** *If M is (ϵ, δ)-differentially private over $\equiv$, then for a arbitrary randomized mapping $F : \mathcal{Y} \to \mathcal{Z}$, $F \circ M$ is (ϵ, δ)-differentially private over $\equiv$.*

Details of Proposition 6 and 7 can be found in Appendix A of this paper's full version (Dharangutte et al. 2022). Note that when privacy mechanisms that obey different invariant constraints are imposed, it might be possible to infer aspects of the confidential data following logical consequences from both. It is not clear how to define post-processing for invariant constraints in those situations.

3 Generalized Laplace and Gaussian Mechanisms

Now we study $\mathcal{A}$-induced integer differentially private mechanisms that satisfy counting invariant constraints. The main challenge is that counting invariant constraints significantly restrict the feasible output space, especially when the output is required to take only integer values. To resolve this issue, we first show that counting invariant constraints on integer-valued output can be written as a lattice space. Then we propose revised Laplace and Gaussian mechanisms on a lattice space. We discuss how to implement these mechanisms via sampling and MCMC in the Implementation section.

3.1 Counting Invariants and Lattice Spaces

In this section, we show the output space of integer datasets satisfying counting invariant constraints are lattice spaces. A *lattice* Λ is a discrete additive subgroup of $\mathbb{R}^d$. Given a matrix $\boldsymbol{B} \in \mathbb{R}^{d\times m}$ consisting of m basis $\vec{b}_1, \ldots, \vec{b}_m$ with $1 \leq m \leq d$, a lattice generated by the basis is $\Lambda(\boldsymbol{B}) = \left\{\boldsymbol{B}\vec{v} = \sum_{i=1}^m v_i\vec{b}_i : \vec{v} \in \mathbb{Z}^m\right\}$.

Proposition 8. *Given a collection of counting invariants $\mathcal{A} = \{A_1, \ldots, A_k\}$, there exists a lattice $\Lambda_{\mathcal{A}} = \Lambda(\boldsymbol{C}_{\mathcal{A}})$with basis $\boldsymbol{C}_{\mathcal{A}} \in \mathbb{Z}^{d\times(d-k)}$ so that a function $f : \mathbb{N}^d \to \mathbb{Z}^d$ is $\mathcal{A}$-invariant if and only if for all $\vec{x} \in \mathbb{N}^d$*

$$f(\vec{x}) - \vec{x} \in \Lambda_{\mathcal{A}}.$$

Note that as $\boldsymbol{C}_{\mathcal{A}}$ is an integer-valued matrix, $\Lambda_{\mathcal{A}}$ is a subset of the integer grid $\mathbb{Z}^d$. As we impose more counting invariants, the lattice $\Lambda_{\mathcal{A}}$ becomes sparser, making it harder to find a feasible solution for an $\mathcal{A}$-invariant mechanism.

Proof. Given a collection of counting invariant constraints $\mathcal{A} = \{A_1, \ldots, A_k\}$ and a dataset $\vec{x}$, finding a feasible output $\vec{y}$ with $\vec{y} \equiv_{\mathcal{A}} \vec{x}$ is equivalent to solving the following linear equations in (3), where $\boldsymbol{A} \in \{0,1\}^{k\times d}$ is an *incidence matrix*, for all $i \in [d]$ and $l \le k$ $A_{l,i} = 1$ if $i \in A_l$ and zero otherwise. Then $\vec{y} \equiv_{\mathcal{A}} \vec{x}$ if and only if

$$\boldsymbol{A}\vec{y} = \boldsymbol{A}\vec{x}. \tag{3}$$

Since we require $\vec{y} \in \mathbb{Z}^d$, the problem of solving $\vec{y}$ is known as solving the linear Diophantine equation (Gilbert and Pathria 1990; Schrijver 1998; Greenberg 1971), and can be done by computing the *Smith normal form* of $\boldsymbol{A}$. Specifically, given an integer-valued matrix $\boldsymbol{A} \in \mathbb{Z}^{k\times d}$ of rank k, there exists $\boldsymbol{U} \in \mathbb{Z}^{k\times k}, \boldsymbol{V} \in \mathbb{Z}^{d\times d}$ and $\boldsymbol{D} \in \mathbb{Z}^{k\times d}$ with

$$\boldsymbol{UAV} = \boldsymbol{D} \tag{4}$$

so that $\boldsymbol{U}$ and $\boldsymbol{V}$ are unimodular matrices that are invertible over the integers, and $\boldsymbol{D}$ is a diagonal matrix (i.e. the Smith normal form of $\boldsymbol{A}$) with $D_{l,l} \neq 0$ for all $l \in [k]$.

With the above decomposition, we can characterize the integer solutions of (3). $\vec{y}$ is a solution of (3), if and only if $\boldsymbol{A}(\vec{y} - \vec{x}) = \boldsymbol{0}$ and it is equivalent to $\boldsymbol{DV}^{-1}(\vec{y} - \vec{x}) = \boldsymbol{0}$, since $\boldsymbol{U}$ is unimodular. Because $\boldsymbol{D}$ is diagonal and has rank k, $\vec{y}$ is a solution if and only if there exists $\vec{w} \in \mathbb{Z}^d$ so that $\vec{y} = \vec{x} + \boldsymbol{V}\vec{w}$ and $w_l = 0$ for all $l \le k$. Additionally, if we define $\boldsymbol{C}_{\mathcal{A}} \in \mathbb{Z}^{d\times(d-k)}$ that consists the bottom $d-k$ rows of $\boldsymbol{V}$, then $\vec{y}$ is a solution if and only if $\vec{y} - \vec{x} \in \{\boldsymbol{C}_{\mathcal{A}}\vec{v} : \vec{v} \in \mathbb{Z}^{d-k}\} = \Lambda(\boldsymbol{C}_{\mathcal{A}})$. We call $\mathcal{A}$ *full rank* if the rank of $\boldsymbol{A} \in \{0,1\}^{k\times d}$ is k, the associated $\boldsymbol{C}_{\mathcal{A}}$ as the *basis of* $\mathcal{A}$, and $\Lambda_{\mathcal{A}} := \Lambda(\boldsymbol{C}_{\mathcal{A}})$. The integer solutions of (3) is the lattice $\Lambda_{\mathcal{A}}$ shifted by $\vec{x}$. Therefore, if $M : \mathbb{N}^d \to \mathbb{Z}^d$ is $\mathcal{A}$-invariants the output given input $\vec{x}$ is contained in $\vec{x} + \Lambda_{\mathcal{A}} := \{\vec{x} + \vec{z} : \vec{z} \in \Lambda_{\mathcal{A}}\}$, $M(\vec{x}) \in \vec{x} + \Lambda_{\mathcal{A}}$. □

3.2 Generalized Laplace Mechanism

Now we can define a generalized Laplace mechanism whose output space satisfies Proposition 8. We adopt the classical exponential mechanism to this restricted output space and show the resulting mechanisms are as required. Because the output space Proposition 8 is unbounded, we show the utility guarantee by controlling the dimension of the lattice spaces which may be of interest by itself. First, we formally define the generalized Laplace mechanism.

Definition 9. Given a collection of counting invariant condition $\mathcal{A}$ of full rank and $\epsilon > 0$, the *generalized Laplace mechanism* is $M_{Lap,\mathcal{A},\epsilon}(\vec{x}) = \vec{x} + \vec{z}$ with $\vec{z}$ sampled from

$$q_\epsilon(\vec{z}) \propto \exp\left(-\epsilon\|\vec{z}\|\right) \mathbf{1}[\vec{z} \in \Lambda_{\mathcal{A}}] \tag{5}$$

where $\Lambda_{\mathcal{A}}$ is defined in Proposition 8.

Note that if the counting constraint is *vacuous*, i.e. $\mathcal{A} = \emptyset$ with $d = 1$, and the output can be real number, the generalized Laplace mechanism reduces to the original Laplace mechanism on histograms. Additionally, if we require integer-valued output with $\mathcal{A} = \emptyset$, the generalized Laplace mechanism becomes the double geometric mechanism.

Theorem 10. *Given full rank $\mathcal{A}$ and $\epsilon > 0$, mechanism $M_{Lap,\mathcal{A},\epsilon} : \mathbb{N}^d \to \mathbb{Z}^d$ in Definition 9 is $\mathcal{A}$-induced integer subspace with $(\epsilon, 0)$.*

The proof is similar to the privacy guarantee of Laplace mechanisms, and is presented in Appendix B of this paper's full version (Dharangutte et al. 2022). The implementation of the generalized Laplace mechanisms in Definition 9 is non-trivial and is elaborated in Implementation section.

After the privacy guarantee, we turn to discuss the utility of the generalized Laplace mechanism. First, (6) establishes the unbiasedness of the generalized Laplace mechanism. The tail bound needs additional work, because the output spaces of the conventional exponential mechanism are often finite, which is not the case here. We resolve this issue using the bounded-dimension property of lattices spaces (Lemma 11), and prove the error bound is sub-exponential in (7). There is a long line of work (Landau 1915; Tsuji 1953; Bentkus and Gotze 1999) that approximate the number of lattice points in sphere by the volume. Using these ideas, we prove the following lemma (with proof in Appendix B) of this paper's full version (Dharangutte et al. 2022) for Theorem 12.

Lemma 11. *Let $\boldsymbol{B} \in \mathbb{R}^{d\times m}$ be a rank $m \le d$ matrix. For all $r > 0$ large enough, $|\{\vec{z} \in \Lambda(\boldsymbol{B}) : \|\vec{z}\|_2 \le r\}| \le \frac{2V_m}{\sqrt{\det(\boldsymbol{B}^\top\boldsymbol{B})}} r^m$ where V_m is the volume of the m dimensional unit sphere.*

Theorem 12 (unbiasedness and accuracy)**.** *Given $\mathcal{A}$ of k counting constraints and $\epsilon > 0$, $M_{Lap,\mathcal{A},\epsilon}$ in Definition 9 is unbiased*

$$\mathbb{E}[M_{Lap,\mathcal{A},\epsilon}(\vec{x})] = \vec{x}, \tag{6}$$

and, if $\|\cdot\|$ is ℓ_2-norm, there exists a constant $K > 0$ so that for all $\vec{x} \in \mathbb{N}^d$, and large enough $t > 0$,

$$\Pr[\|M_{Lap,\mathcal{A},\epsilon}(\vec{x}) - \vec{x}\|_2 \ge t] \le Kt^{d-k}\exp(-\epsilon t). \tag{7}$$

Moreover, K can be $\frac{4\cdot 2^{d-k}V_{d-k}}{\sqrt{\det(\boldsymbol{C}_{\mathcal{A}}^\top\boldsymbol{C}_{\mathcal{A}})}}$ where V_{d-k} is the volume of the $(d-k)$ dimensional unit sphere.

Note that as the dimension $d-k$ increases $2^{d-k}V_{d-k} \to 0$, so K is determined by $1/\sqrt{\det(\boldsymbol{C}_{\mathcal{A}}^\top\boldsymbol{C}_{\mathcal{A}})}$.

3.3 Generalized Gaussian Mechanism

In this section, we propose generalized Gaussian mechanisms that relate to Lattice-based cryptography discussed in Section 4. We first define Gaussian random variables on lattices. Given a lattice $\Lambda \subset \mathbb{R}^d$, and $\vec{c} \in \mathbb{R}^d$, the spherical *Gaussian random variable* $\vec{Z}$ on Λ with variance σ and center $\vec{c}$ is

$$\Pr[\vec{Z} = \vec{z}] \propto \exp\left(-\frac{1}{2\sigma^2}\|\vec{z} - \vec{c}\|_2^2\right), \forall \vec{z} \in \Lambda. \tag{8}$$

Note that the generalized Gaussian mechanism utilizes an ℓ_2 norm to measure the noise scale.

Definition 13. Given a collection of counting invariant condition $\mathcal{A}$ of full rank, $\epsilon > 0$, and $\delta > 0$, the *generalized Gaussian mechanism* is defined as $M_{G,\mathcal{A},\epsilon,\delta}(\vec{x}) = \vec{x} + \vec{z}$ where $\vec{z}$ is the Gaussian random variable on $\Lambda_{\mathcal{A}}$ with center at $\mathbf{0}$ and variance $\sigma^2_{\epsilon,\delta} = \frac{2c_{\mathcal{A}} \ln 1/\delta}{\epsilon^2}$ for some constant $c_{\mathcal{A}} = \mathcal{O}(\max\{(d-k)\ln(d-k), \ln K\})$ that only depends on the dimension and the collection of counting constraints where K is defined in Theorem 12.

While the variance scale in the original Gaussian mechanisms is independent of the dimension, generalized Gaussian mechanism's $\sigma_{\epsilon,\delta}$ in Definition 13 depends on the dimension. This may be due to a lack of symmetry of the lattice space. Recall that the proof of the original Gaussian mechanism reduces the high dimensional case to the one dimensional setting, thanks to the Gaussian being spherically symmetric. However, our lattice space $\Lambda_{\mathcal{A}}$ is generally not spherically symmetric, and simple reduction does not work.

Theorem 14. *Given full rank $\mathcal{A}$, ϵ and δ with $0 < \delta < \epsilon < 1/e$, mechanism $M_{G,\mathcal{A},\epsilon,\delta} : \mathbb{N}^d \to \mathbb{Z}^d$ in Definition 13 is $\mathcal{A}$-induced integer subspace differentially private with (ϵ, δ).*

After the privacy guarantee, we turn to show the utility of the generalized Gaussian mechanism. First, since the distribution function of the discrete Gaussian is an even function, the additive errors of the discrete Gaussian mechanism is unbiased (9). Furthermore, similar to Gaussians on the Euclidean space, Gaussians on lattices are sub-Gaussian.[2]

Theorem 15 (unbiasedness and accuracy)**.** *Given the condition in Theorem 14, $M_{G,\mathcal{A},\epsilon,\delta}$ in Definition 13 is unbiased,*

$$\mathbb{E}[M_{G,\mathcal{A},\epsilon,\delta}(\vec{x})] = \vec{x}, \text{ for all } \vec{x} \in \mathbb{N}^d. \tag{9}$$

Additionally, there exists a constant $K > 0$ defined in Theorem 12 so that for all $\vec{x} \in \mathbb{N}^d$, and large enough $t > 0$,

$$\Pr[\|M_{G,\mathcal{A},\epsilon,\delta}(\vec{x}) - \vec{x}\|_2 \geq t] \leq K t^{d-k} e^{-\frac{t^2}{2\sigma^2_{\epsilon,\delta}}}. \tag{10}$$

Last, we remark that previous work has used discrete Gaussian mechanisms (Canonne, Kamath, and Steinke 2020)for integer valued noises. But these mechanisms do not satisfy (non-trivial) invariants constraints. Note that when there is no invariant constraint, the resulting lattice space becomes an integer grid and our generalized Gaussian mechanism reduces to the discrete Gaussian mechanism.

4 Implementation

We discuss implementation issues of discrete Gaussian and generalized Laplace in lattice spaces. In general the design of exact samplers for complex privacy mechanisms (such as the Exponential mechanism, to which our generalized Laplace mechanism is a special case) is a widely acknowledged challenge in the literature. Section 3 of Abowd et al. (Abowd et al. 2022) specifically discussed this challenge in the context of invariants and integer constraints and concluded that direct sampling from the Exponential mechanism is "infeasible" for their TopDown algorithm. We address this problem in two ways. For certain discrete Gaussian, we can use exact samplers; for generalized Laplace we develop an efficient and practical solution using Markov chain Monte Carlo (MCMC). Our experiments all use the MCMC sampler for practicality.

Sampling discrete Gaussian on a lattice. There are several efficient algorithms to sample discrete variables within negligible statistical distance of any discrete Gaussian distribution whose scales are not exceedingly narrow (Gentry, Peikert, and Vaikuntanathan 2008; Peikert 2010). Moreover, (Brakerski et al. 2013) provide an exact Gaussian sampler.

Theorem 16 (Lemma 2.3 in Brakerski et al. (2013))**.** *There is a probabilistic polynomial-time algorithm that takes as input a basis $\boldsymbol{B} = (\vec{b}_1, \ldots, \vec{b}_m)$ for a lattice $\Lambda(\boldsymbol{B}) \subset \mathbb{R}^d$, a center $\vec{c} \in \mathbb{R}^d$ and parameter $\sigma = \Omega(\max_i \|\vec{b}_i\|_2 \ln d)$ and outputs a vector that is distributed exactly as (8).*

We note that, however, it is believed to be computationally difficult to sample efficiently on general lattice spaces if the target distribution is not sufficiently dispersed. Most lattice-based cryptographic schemes (Folláth 2014) are based on the hardness assumption of solving the short integer solution (SIS) problem and the learning with errors (LWE) problem (Brakerski et al. 2013), and the ability of an efficient sampler of a narrow distribution on general lattice efficiently would enable us to solve short integer solution and break lattice-based cryptographic schemes.

MCMC sampling for generalized Laplace on a lattice. Since a discrete Gaussian mechanism only provides (ϵ, δ)-DP, it is important to develop pure-DP mechanisms using the generalized Laplace mechanism on a lattice. We devise an MCMC sampling scheme to instantiate the generalized Laplace mechanism (Definition 9). As a practical solution, MCMC has been theoretically studied in the literature; e.g. Wang, Fienberg, and Smola (2015) for stochastic gradient Monte Carlo and Ganesh and Talwar (2020) for discretized Langevin MCMC. The challenge to designing such a sampling scheme is twofold. First, the invariants render the target state space highly constrained. To achieve sampling efficiency requires efficient proposals, which ideally satisfy the invariants themselves in order to maintain a reasonable acceptance rate (c.f. Gong and Meng 2020). Second, MCMC algorithms are generally not exact samplers, and may fail to converge to the target distribution in finite time, thus incur additional cost to privacy that can be difficult to quantify for a given application. To this end, we 1) target only the additive noise component of the mechanism, ensuring that any privacy leakage due to sampling is independent of the underlying confidential data; and 2) devise an empirical convergence assessment for the upper bound on the total variation (TV) distance between the chain's marginal distribution the target, based on the L-lag coupling method proposed by (Biswas, Jacob, and Vanetti 2019).

Proposal design. We wish to sample the additive noise $\vec{z}$ employed by the generalized Laplace mechanism. For an incidence matrix $\boldsymbol{A} \in \mathbb{Z}^{k \times d}$ specifying the invariants, recall its Smith normal form $\boldsymbol{D} = \boldsymbol{UAV}$ in (4). Consider a jumping distribution with probability mass function

[2] This is nontrivial, because the support of Gaussian in (8) is not uniform. For instance, suppose Λ is not a lattice space, but has 2^{r^2} points in each integer radius r shell, the a Gaussian on such a space is not sub-Gaussian.

$g(\vec{u}) = g_{\boldsymbol{A}}\left(\boldsymbol{V}^{-1}\vec{u}\right)$, where

$$g_{\boldsymbol{A}}(\vec{e}) = \prod_{j=1}^{k} \mathbf{1}\{e_j = 0\} \prod_{j=k+1}^{d} \eta_j(e_j), \qquad (11)$$

for $\vec{e} = (e_1, \ldots, e_d)$ a *pre-jump* object whose first k entries are exactly zero, and $\vec{u} = \boldsymbol{V}\vec{e}$ the proposed jump. The η_j's in (11) can be set to any interger-valued univariate probabilities that are unbiased and symmetric around zero. A straightforward choice is the double geometric distribution, with mass function $\eta_j(e) = \frac{1-a}{1+a}a^{\|e\|_1}$ and parameter a. Note that the jumping distribution g connects to $g_{\boldsymbol{A}}$ in (11) without a Jacobian because $\vec{u}$ and $\vec{e}$ are integer-valued and $\boldsymbol{V}$ is invertible. The proposed jump $\vec{u} \sim g$ is unbiased and symmetric, i.e. $\mathbb{E}_g(\vec{u}) = \mathbf{0}$ and $g(\vec{u}) = g(-\vec{u})$, and always respects the desired invariants specified by $\boldsymbol{A}$:

$$\boldsymbol{A}\vec{u} = \boldsymbol{U}^{-1}\boldsymbol{U}\boldsymbol{A}\boldsymbol{V}\vec{e} = \boldsymbol{U}^{-1}\boldsymbol{D}\vec{e} = \boldsymbol{U}^{-1}\mathbf{0} = \mathbf{0}.$$

A Metropolis sampling scheme. The target distribution $q_\epsilon(\vec{z})$ is given in (5) and is known only up to a normalizing constant. Algorithm 1 in Appendix D of this paper's full version (Dharangutte et al. 2022) presents a Gibbs-within-Metropolis sampler that produces a sequences of dependent draws $\left(\vec{z}^{(l)}\right)_{0 \le l \le \text{nsim}}$ from the target distribution q_ϵ in (5) known only up to a normalizing constant. We use an additive jumping distribution whose element-wise construction is described in (11). The algorithm incurs a transition kernel that dictates how the chain moves from an existing state to the next one: $\vec{z}^{(l)} \sim K\left(\vec{z}^{(l-1)}, \cdot\right)$. The initial distribution π_0 may be chosen simply as g for convenience. The choice of η_j is a tuning decision for the algorithm, and should be made to encourage fast mixing of the chain. We discuss our choices for examples in the Experiments section. Note that steps 5 through 7 of Algorithm 1 updates the proposed jump in a Gibbs sweep (i.e. one dimension at a time), utilizing the fact that the pre-jump object $\vec{e}$ is element-wise independent under $g_{\boldsymbol{A}}$. Doing so facilitates the L-lag coupling, to be discussed next, for assessing empirical convergence of the chain. In practice, these updates may be performed simultaneously.

Empirical convergence assessment using L-lag coupling. We perform empirical convergence assessment of the proposed algorithm using L-lag coupling. Construct a joint transition kernel $\tilde{K}$ of two Markov chains, each having the same target distribution q_ϵ induced by the marginal transition kernel K as defined in Algorithm 1. The joint kernel $\tilde{K}$, given by Algorithm 2 in Appendix D, is a maximal coupling kernel such that the L-th lag of the two chains will couple in finite time with probability one. The random L-lag meeting time, which Algorithm 3 in Appendix D samples, provides an estimate of the upper bound on the TV distance,

$$d_{TV}(q_\epsilon^{(l)}, q_\epsilon) \le \mathbb{E}\left[\max\left(0, \left\lceil \left(\tau^{(L)} - L - l\right)/L \right\rceil\right)\right]$$

between the target distribution q_ϵ and the marginal distribution of the chain at time l, denoted $q_\epsilon^{(l)}$ (Biswas, Jacob, and Vanetti 2019, Theorem 2.5). The upper bounds are obtained as empirical averages over independent runs of coupled Markov chains. The lag $L > 0$ would need to be set, and we discuss its choice in the next section and Appendix D. Compared to earlier work which uses MCMC to instantiate privacy mechanisms, our use of L-lag coupling provides real-time (rather than asymptotic) assessment on the MCMC convergence behavior. The number of iterations needed to ensure convergence is empirically assessed via the upper bound on the total variation (note that this is still an estimate). In general, there is a tradeoff between the number of iterations and an extra privacy loss budget δ', in the sense that Eq. (5) can be absorbed as another additive error in the DP guarantee.

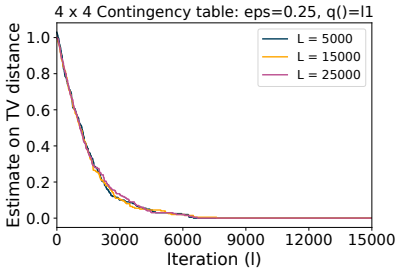

Figure 1: Estimated TV bound from target

County	Low	Medium	High	Very High	Total
Alpha	-1	-1	0	2	0
Beta	5	4	-2	-7	0
Gamma	-5	-3	5	3	0
Delta	1	0	-3	2	0
Total	0	0	0	0	0

Table 1: Generalized Laplace additive noise ($\epsilon = 0.25$, $q() = \ell_1$-norm), which preserves row and column margins.

5 Experiments

Intersecting counting constraints. Consider a synthetic example in which a set of three counting constraints $\mathcal{A} = (A_1, A_2, A_3)$ are defined over 14 records, where the intersection of all subsets of constrains are nonempty. The constraints are schematically depicted by Fig. 3 in Appendix E of this paper's full version (Dharangutte et al. 2022), and may be encoded by an incident matrix $\boldsymbol{A} \in \mathbb{Z}^{3\times 14}$, with each row corresponding to one of A_1, A_2 and A_3, and each element being 1 at indices corresponding to the record within that constraint and 0 otherwise. We apply Algorithm 1 to instantiate the generalized Laplace mechanism, with both ℓ_1-norm and ℓ_2-norm targets, to privatize a data product that conforms to the constraint $\boldsymbol{A}$. To ensure adequate dispersion of the target distribution, we set $\epsilon = 0.25$, a value on the smaller end within the range of meaningful privacy protection (e.g. Dwork 2011). The pre-jump proposal distributions η_j are double geometric distributions, with parameter $a = \exp(-1)$ for the ℓ_1-norm target and $a = \exp(-1.5)$ for the ℓ_2-norm target. Details on the tuning, the noise distribution and convergence assessment can be found in Appendix E of this paper's full version (Dharangutte et al. 2022). In particular, Fig. 4 in Appendix E shows that the chains empirically converge around

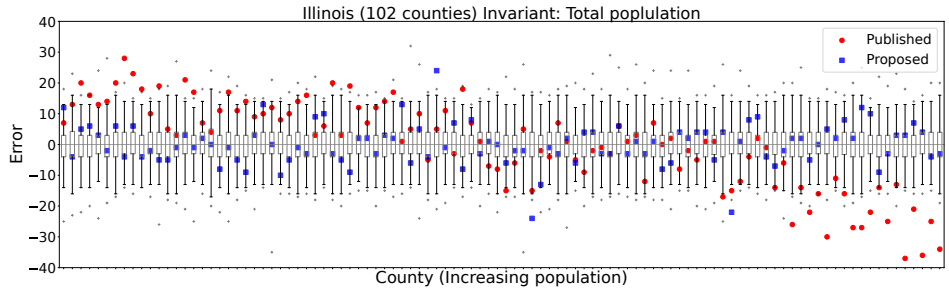

Figure 2: Privacy noise from the generalized Laplace mechanism (Definition 9) via Algorithm 1 (blue squares: one instance; boxplot: 1000 instances) for county populations of Illinois in increasing county sizes. State population total is invariant. The proposed noises are integer-valued and unbiased. For comparison are DAS errors from the Nov 2020 vintage 2010 Census demonstration data (Van Riper, Kugler, and Schroeder 2020) (red dots).

10^5 iterations for the ℓ_1 and 10^6 for the ℓ_2-norm target.

Delinquent children by county and household head education level. The Federal Committee on Statistical Methodology published a fictitious dataset concerning delinquent children in the form of a 4×4 contingency table, tabulated across four counties by education level of household head (Table 4 in Federal Committee on Statistical Methodology 2005, reproduced in Table 2 of Appendix E), to illustrate various traditional SDL techniques (Slavković and Lee 2010).

The charge is to extend privacy protection to the sensitive individual records while preserving the margins of the contingency table for data publication. To do so, we apply Algorithm 1 to instantiate the generalized Laplace mechanism with both ℓ_1- and ℓ_2-norm targets and with $\epsilon = 0.25$. Fig. 1 shows the evolution of the TV upper bound on the chain's marginal distributions to the ℓ_1-norm target, estimated with 200 independent coupled chains, which appear to converge after about 10^4 iterations and are stable at various choices of L. Table 1 shows one instance of the proposed additive noise, obtained after the chain achieves empirical convergence. They are integer-valued, with zero row and column totals which would preserve the margins of the confidential table. Convergence assessment for the ℓ_2-norm target and discussions on the choice of proposal are in Appendix E.

2010 U.S. Census county-level population data. We consider the publication of county-level population counts subject to the invariant of state population size, and compare with the privacy-protected demonstration files produced by preliminary versions of the 2020 Census DAS. The confidential values are the 2010 Census Summary Files (CSF), curated by IPUMS NHGIS and are publicly available (Van Riper, Kugler, and Schroeder 2020). Employed for this example are the November 2020 vintage demonstration data, protected by pure differential privacy with $\epsilon = 0.192 = 4\text{ (total)} \times 0.16\text{ (county level)} \times 0.3\text{ (population query)}$.

We demonstrate Algorithm 2 using the generalized Laplace mechanism under ℓ_1 norm, with ϵ set to accord to the Census Bureau's specification and $a = \exp(-2.5)$. Fig. 2 showcases the proposed county-level errors applied to the population of Illinois. The x-axis is arranged in increasing true county population sizes. Blue squares shows an instance of the proposed noise vector. The boxplots summarize 1000 proposed noise vectors (thinned at 0.01%), with whiskers indicating 1.5 times the interquartile range and cross-marks indicating extreme realizations. The proposed errors are integer-valued and centered around zero, confirming their marginal unbiasedness. Note that the published DAS privacy errors (red dots) show a clear negative bias as a function of increasing underlying county population size. The bias is likely due to optimization-based post-processing during the estimation phase which imposes non-negativity on the constituent geographic areas with small population counts. Similar observations were made for other states (Appendix E).

6 Discussion and Future Work

This paper provides solutions for sanitizing private data that are required to simultaneously observe pre-specified invariants and integral characteristics. The proposed *integer subspace differential privacy* scheme allows for rigorous statements of privacy guarantees while maintaining good statistical properties, including unbiasedness, accuracy, and probabilistic transparency of the privatized output. An efficient MCMC scheme is devised to instantiate the proposed mechanism, alongside tools to assess empirical convergence.

One direction for future work is to design exact sampling algorithms for the proposed generalized discrete Laplace and Gaussian mechanisms. In the case that the normalizing constant to the target distribution q_ϵ is known exactly, it may be feasible to design, for example, efficient rejection sampling based on intelligent choices of proposal distributions. It is also conceivable that a perfect sampling scheme, such as *coupling from the past* (Propp and Wilson 1996), may be designed. The recent work of Seeman, Reimherr, and Slavković (2021) proposes an exact sampling scheme for Exponential mechanisms using *artificial atoms*. However, the technique requires the state space be compact, which is not the case for the constrained yet unbounded target distribution q_ϵ considered in this work. Another direction for future work is to extend invariant-respecting privacy protection data products that are required to be *binary* (i.e., exhibiting or not exhibiting a private attribute), as well as to obey *inequality* constraints (e.g. non-negativity of count data). The challenge there is again on the design of efficient methods to generate noise vectors in an extremely constrained discrete space.

Acknowledgements

Prathamesh Dharangutte and Jie Gao would like to acknowledge support from NSF through grants CCF-2118953, IIS-2207440, CCF-2208663 and CNS-2137245. Ruobin Gong would like to acknowledge support from NSF through grant DMS-1916002.

References

Abowd, J. M.; Ashmead, R.; Cumings-Menon, R.; Garfinkel, S.; Heineck, M.; Heiss, C.; Johns, R.; Kifer, D.; Leclerc, P.; Machanavajjhala, A.; Moran, B.; Sexton, W.; Spence, M.; and Zhuravlev, P. 2022. The Census TopDown Algorithm. *Harvard Data Science Review*.

Ashmead, R.; Kifer, D.; Leclerc, P.; Machanavajjhala, A.; and Sexton, W. 2019. Effective Privacy After Adjusting for Invariants with Applications to the 2020 Census. Technical report, US Census Bureau.

Barak, B.; Chaudhuri, K.; Dwork, C.; Kale, S.; McSherry, F.; and Talwar, K. 2007. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, 273–282.

Barber, R. F.; and Duchi, J. C. 2014. Privacy and Statistical Risk: Formalisms and Minimax Bounds. arXiv 1412.4451.

Bentkus, V.; and Gotze, F. 1999. Lattice Point Problems and Distribution of Values of Quadratic Forms. *Annals of Mathematics*, 150(3): 977–1027.

Biswas, N.; Jacob, P. E.; and Vanetti, P. 2019. Estimating convergence of Markov chains with L-lag couplings. *Advances in Neural Information Processing Systems*, 32.

Brakerski, Z.; Langlois, A.; Peikert, C.; Regev, O.; and Stehlé, D. 2013. Classical Hardness of Learning with Errors. 575–584.

Canonne, C. L.; Kamath, G.; and Steinke, T. 2020. The discrete gaussian for differential privacy. *Advances in Neural Information Processing Systems*, 33: 15676–15688.

Chatzikokolakis, K.; Andrés, M. E.; Bordenabe, N. E.; and Palamidessi, C. 2013. Broadening the Scope of Differential Privacy Using Metrics. In Cristofaro, E. D.; and Wright, M. K., eds., *Privacy Enhancing Technologies - 13th International Symposium, PETS 2013, Bloomington, IN, USA, July 10-12, 2013. Proceedings*, volume 7981 of *Lecture Notes in Computer Science*, 82–102. Springer.

Cormode, G.; Kulkarni, T.; and Srivastava, D. 2017. Constrained differential privacy for count data. *arXiv preprint arXiv:1710.00608*.

Desfontaines, D.; and Pejó, B. 2019. SoK: Differential Privacies. *CoRR*, abs/1906.01337.

Dharangutte, P.; Gao, J.; Gong, R.; and Yu, F.-Y. 2022. Integer Subspace Differential Privacy. *arXiv preprint arXiv:2212.00936*.

Dwork, C. 2011. A firm foundation for private data analysis. *Communications of the ACM*, 54(1): 86–95.

Dwork, C.; McSherry, F.; Nissim, K.; and Smith, A. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, 265–284. Springer.

Federal Committee on Statistical Methodology. 2005. Report on Statistical Disclosure Limitation Methodology. Technical report, Statistical Policy Working Paper 22 (Second Version). https://www.hhs.gov/sites/default/files/spwp22.pdf [Accessed: 05-15-2022].

Folláth, J. 2014. Gaussian Sampling in Lattice Based Cryptography. *Tatra Mountains Mathematical Publications*, 60(1): 1–23.

Ganesh; and Talwar. 2020. Faster differentially private samplers via Rényi divergence analysis of discretized Langevin MCMC. In *Adv. Neural Inf. Process. Syst.*, 7222–7233.

Gao, J.; Gong, R.; and Yu, F.-Y. 2022. Subspace Differential Privacy. In *Proceedings of the The Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22)*, 3986–3995. ArXiv:2108.11527.

Gentry, C.; Peikert, C.; and Vaikuntanathan, V. 2008. Trapdoors for hard lattices and new cryptographic constructions. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, 197–206.

Gilbert, W. J.; and Pathria, A. 1990. Linear diophantine equations. *Preprint*. ncatlab.org/nlab/files/GilbertPathria90.pdf.

Gong, R. 2022. Transparent privacy is principled privacy. *Harvard Data Science Review (Special Issue 2)*. Doi:10.1162/99608f92.b5d3faaa.

Gong, R.; and Meng, X.-L. 2020. Congenial Differential Privacy under Mandated Disclosure. In *Proceedings of the 2020 ACM-IMS on Foundations of Data Science Conference*, 59–70.

Greenberg, H. 1971. *Integer programming*. Academic Press.

Hay, M.; Rastogi, V.; Miklau, G.; and Suciu, D. 2009. Boosting the accuracy of differentially-private histograms through consistency. *arXiv preprint arXiv:0904.0942*.

He, X.; Machanavajjhala, A.; and Ding, B. 2014. Blowfish privacy: Tuning privacy-utility trade-offs using policies. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, 1447–1458.

Kifer, D.; and Machanavajjhala, A. 2014. Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems (TODS)*, 39(1): 1–36.

Landau, E. 1915. Zur analytischen Zahlentheorie der definiten quadratischen Formen (Uber die Gitterpunkte in einem mehrdimensionalen Ellipsoid), Sitzungsber. Koniglich Breissischen. *Akad. Wiss.*, 31: 458–476.

Peikert, C. 2010. An efficient and parallel Gaussian sampler for lattices. In *Annual Cryptology Conference*, 80–97. Springer.

Propp, J. G.; and Wilson, D. B. 1996. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures & Algorithms*, 9(1-2): 223–252.

Schrijver, A. 1998. *Theory of linear and integer programming*. John Wiley & Sons.

Seeman, J.; Reimherr, M.; and Slavković, A. 2021. Exact Privacy Guarantees for Markov Chain Implementations of the Exponential Mechanism with Artificial Atoms. *Advances in Neural Information Processing Systems*, 34.

Seeman, J.; Slavkovic, A.; and Reimherr, M. 2022. A Formal Privacy Framework for Partially Private Data. *arXiv preprint arXiv:2204.01102*.

Slavković, A. B.; and Lee, J. 2010. Synthetic two-way contingency tables that preserve conditional frequencies. *Statistical Methodology*, 7(3): 225–239.

Tsuji, M. 1953. On lattice points in an n-dimensional ellipsoid. *Journal of the Mathematical Society of Japan*, 5(3-4): 295–306.

Van Riper, D.; Kugler, T.; and Schroeder, J. 2020. IPUMS NHGIS Privacy-Protected 2010 Census Demonstration Data, version 20201116. Minneapolis, MN: IPUMS.

Wang, Y.-X.; Fienberg, S. E.; and Smola, A. 2015. Privacy for Free: Posterior Sampling and Stochastic Gradient Monte Carlo. In *Proceedings of the International Conference on Machine Learning*, 2493–2502.