Extending the Foundations of Differential Privacy: Flexibility and Robustness

Aman Bansal
IIT Bombay
aman0456b@gmail.com

Rahul Chunduru
IIT Bombay
chrahul@cse.iitb.ac.in

Deepesh Data
UCLA
deepesh.data@gmail.com

Manoj Prabhakaran
IIT Bombay
mp@cse.iitb.ac.in

Abstract

Differential Privacy (DP) is an area that has recently seen many direct and indirect applications to machine learning. While DP provides the most rigorous notions of privacy, there are many settings where its applicability is limited.

In this work, we make foundational contributions to DP, thereby greatly expanding its applicability in multiple ways. Towards this, we define two complementary concepts, namely, Flexible Accuracy and Robust Privacy. Flexible Accuracy allows small distortions in the input (e.g., dropping outliers) before measuring accuracy of the output, allowing one to extend DP mechanisms to high-sensitivity functions. Robust Privacy goes beyond the notion of individual-level privacy considered by DP: It requires a mechanism to not include any "non-utile information" in the output, thereby offering database-level privacy. Then, we present mechanisms that can help in achieving these notions, where previously no meaningful differentially private mechanisms were available. In particular, we illustrate an application to differentially private histograms, which in turn yields mechanisms for revealing the support of a dataset or the extremal values in the data. Analyses of our constructions exploit new versatile composition theorems that facilitate modular design.

Our definitional framework, in terms of "lossy Wasserstein distance" – a 2-parameter error measure for distributions – is of independent interest. In particular, it leads us to a mechanism for differentially private sampling.

1 Introduction

In this work, we make foundational contributions to the area of Differential Privacy (DP), greatly extending its applicability in machine learning and other contexts. Our main contribution is to identify and address two limitations of the DP framework. At a high-level, these limitations follow from a seemingly natural choice: Accuracy guarantees of a mechanism are in terms of distances in the output space, and privacy demands are in terms of distances in the input space (neighboring inputs). Somewhat surprisingly, these choices turn out to be not always adequate. Our extensions can be seen as adding accuracy guarantees in terms of distances (or rather, distortions) in the input space, and privacy demands in terms of distances in the output space. Along the way, we extend the notion of DP to randomized functions over a metric space, for which distances are measured using a (generalization of) Wasserstein distance. We illustrate the applicability of our foundational extensions with applications to an important class of functions – namely, functions of a histogram of the data.

Our work could also be viewed as an extension to an earlier approach by Blum, Ligett and Roth [BLR13], which partly addressed the issues that motivate our work. Our framework subsumes that of [BLR13] and goes well beyond it, e.g., by enabling privacy mechanisms for high-sensitivity functions like maximum.

We briefly discuss the limitations of DP addressed in this work.

Limits Set by Sensitivity. Consider querying a database consisting of integer valued observations – say, ages of patients who recovered from a certain disease – for the maximum value. For the sake of privacy, one may wish to apply a DP mechanism, rather than output the maximum in the data itself. Two possible datasets which differ in only one patient are considered neighbors and a DP mechanism needs to make the outputs on these two samples indistinguishable from each other. However, the function in question is *highly sensitive* – two neighboring datasets can have their maxima differ by as much as the entire range of possible ages – and the standard DP mechanisms in the literature will add so much noise that no useful information can be retained.[1]

As we shall see, the above limitation can be attributed to a rigidly defined notion of accuracy. This same rigidity leads to another surprising limitation too. Consider the problem of reporting a *histogram* (again, say, of patients' ages). Here a standard DP mechansim, of adding a zero-mean Laplace noise to each bar of the histogram is indeed reasonable, as the histogram function has low sensitivity in each bar. Now, note that *maximum can be computed as a function of the histogram*. However, even though the histogram mechanism was sufficiently accurate in the standard sense, the maximum computed from its output is no longer accurate! This is because when a non-zero count is added to a large-valued item which originally has a count of 0, the maximum can increase arbitrarily.

In this work we develop a more relaxed notion of accuracy, called *flexible accuracy*, that lets us address both of the above issues. In particular, it not only enables new DP mechanisms for maximum, but also allows one to derive the mechanism from a new DP mechanism for histograms. A new composition theorem enables us to *transfer the accuracy guarantees on histogram to accuracy guarantees on the maximum function* (see Theorem 4.1).

Limitations due to Focus on Individuals. Differential Privacy focuses on making outputs from *neighboring* databases indistinguishable, where neighborhood usually refers to databases obtained by adding or deleting a small number of data items (or a single one). However, such a notion of neighborhood of the databases may not capture all pairs of databases that *should be* indistinguishable from each other.

Consider training a machine learning model on either dataset D_1 or dataset D_2, acquired from two large hospitals. Suppose both the datasets are representative and yield very similar models. In this case, we may reasonably require that querying a model should not reveal whether it was trained on D_1 or D_2. Indeed, since the models are "similar," one may expect them to yield results which are indistinguishable from each other. Unfortunately, this is not generally true: Similarity of outputs is measured from the point of view of honest users, in terms of a distance in the output space (or rather, the Wasserstein distance over that space, since the output is probabilistic); but the extent of their indistinguishability is measured from the point of view of an adversarial user, in terms of total variation distance or the ratio of probabilities (as in DP), which are not influenced by the metric space associated with the outputs. For instance, if the outputs from the model trained in D_1 have an even value for the least significant digit, and for D_2 it is an odd value, the total variation distance between the two output distributions is maximum, while the Wasserstein distance can be very small.

In short, one may demand – without necessarily compromising on accuracy – indistinguishability between datasets *which result in outputs that are close to each other.* Robustness is a complementary notion of privacy we introduce to addresses this demand. In contrast, DP only guarantees indistinguishability between datasets *which are close to each other.* We recommend using a DP mechanism that is also robust.

1.1 Our Contributions

Our contributions are in three parts:

- *Definitions:* In Section 3, we define *flexible accuracy* and *robust privacy* to address the above limitations. As an intermediate step, we define *lossy Wasserstein distance*, which is of independent interest.

[1] In fact, *all datasets* with low maximum values have high sensitivity *locally*, by considering a neighboring dataset with a single additional data item with a large value. As such, mechanisms which add noise based on the local sensitivity rather than global sensitivity [NRS07, ZCP+15] also do not fare any better.

- *Composition Theorems:* In Section 4, we derive versatile composition theorems that can be used to build mechanisms for functions of interest, from mechanisms for basic functions.
- *Mechanisms:* In Section 5, we give a new DP mechanism for releasing a sanitized histogram, which, via our composition theorems yield DP mechanisms with (flexible) accuracy guarantees for *all* histogram-based statistics (see Theorem 5.4), unlike any of the prior work (see [Vad17]). These functions include several high-sensitivity functions like maximum and minimum, range, maximum margin separator, etc. We also design mechanisms with robust privacy, via our composition results.

1.2 Related Work

DP, defined by Dwork et al. [DMNS06] has developed into a highly influential framework for providing formal privacy guarantees (see [DR14] for more details). The notion of flexible accuracy we define is motivated by the difficulty in handling outliers in the data. Some of the work leading to DP explicitly attempts to address the privacy of outliers [CDM+05, CDMT05], as did some of the later works within the DP framework [DL09, BNS19, TS13]. These results rely on having a distribution over the data, or respond only when the answer is a "stable value".

Incidentally, Wasserstein distance has been used in privacy mechanisms in the Pufferfish framework [KM14, SWC17], which also relies on a distribution over data.

Blum et al. [BLR13] introduced notions of *usefulness* and *distributional privacy*, that were motivated by similar limitations of DP as those which motivated flexible accuracy and robust privacy, respectively. As described later, our framework goes well beyond what these notions allowed.

1.3 Paper Organization

We give preliminary definitions in Section 2. We define our new notions, namely, the lossy Wasserstein distance, flexible accuracy, and robust privacy in Section 3, where we also state the important properties that the lossy Wasserstein distance satisfies. We give our new composition theorems for flexible accuracy and robust privacy, along-with a new composition theorem for differential privacy via pre-processing in Section 4. We present and analyze several new mechanisms – for histogram and any histogram-based-statistic (e.g., max and support), and also for robust privacy – in Section 5. In the same section, we also present a mechanism for querying randomized functions that achieves differential privacy (extending the notion beyond deterministic functions) as well as our notions of flexible accuracy and robust privacy. We empirically compare our mechanisms against the state-of-the-art in Section 6.

2 Preliminaries

We denote by $\mathbb{N}$ all non-negative integers (including zero). We refer to a metric space $(\Omega, \mathfrak{d})$, where $\mathfrak{d}$ is a metric over the set Ω. Here, Ω can be discrete or continuous, and for generality, we shall not assume that it is discrete, and use probability density functions to specify distributions over Ω. For example, given a distribution ϕ over $\Omega \times \Omega$, its first marginal is given by $\phi_1(x) = \int_\Omega \phi(x,y)\,\mathrm{d}y$ and the second marginal is given by $\phi_2(y) = \int_\Omega \phi(x,y)\,\mathrm{d}x$.

We will use upper case letters (P, Q, X, Y, etc.) to denote random variables, as well as the probability distributions associated with them. For a random variable P, we denote its probability density function by $P(\cdot)$.

Definition 1 (Total Variation Distance)**.** Let P and Q be two probability distributions on a sample space Ω. The total variation distance between P and Q, denoted by $\Delta(P,Q)$, is defined as

$$\Delta(P,Q) = \frac{1}{2}\int_\Omega |P(\omega) - Q(\omega)|\,\mathrm{d}\omega.$$

2.1 Differential Privacy

Let $\mathcal{X}$ denote a universe of possible "databases" with a symmetric neighborhood relation $\sim$. In typical applications, two databases $\boldsymbol{x}$ and $\boldsymbol{x}'$ are considered neighbors if one is obtained from the other by removing the data corresponding to a single "individual." A *mechanism* $\mathcal{M}$ over $\mathcal{X}$ is an algorithm which takes $\boldsymbol{x} \in \mathcal{X}$ as input and samples an output from an output space $\mathcal{Y}$, according to some distribution. We shall denote this distribution by $\mathcal{M}(\boldsymbol{x})$.

Definition 2 (Differential Privacy [DMNS06, DKM+06])**.** A randomized algorithm $\mathcal{M} : \mathcal{X} \to \mathcal{Y}$ is (ϵ, δ)-differentially private (DP), if for all neighboring databases $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ and all measurable subsets $S \subseteq \mathcal{Y}$, we have

$$\Pr[\mathcal{M}(\boldsymbol{x}) \in S] \leq e^{\epsilon} \Pr[\mathcal{M}(\boldsymbol{x}') \in S] + \delta.$$

Definition 3 (Laplace Distribution)**.** Let b be a positive real number. The Laplace distribution with *scaling parameter* b and mean μ, denoted by $\mathrm{Lap}(x|\mu, b)$, is defined by the following density function:

$$\mathrm{Lap}(x|\mu, b) := \frac{1}{2b} e^{\frac{-|x-\mu|}{b}}, \quad x \in \mathbb{R}.$$

We denote a random variable that is distributed according to the Laplace distribution with the scaling parameter b and mean μ by $\mathrm{Lap}(b, \mu)$. If mean μ is zero, then we will simply denote it by $\mathrm{Lap}(b)$.

2.1.1 Laplace Mechanism

Let $f : \mathcal{X} \to \mathbb{R}$ be a function defined from $\mathcal{X}$ to $\mathbb{R}$. When f is a deterministic map, then the Laplace mechanisms $\mathcal{M}_{\mathrm{Lap}}^{f,b} : \mathcal{X} \to \mathbb{R}$ (given below) is well known for achieving $(\epsilon, 0)$-DP (denoted simply by ϵ-DP and also called pure DP) for f [DMNS06, DR14].

Laplace mechanism $\mathcal{M}_{\mathrm{Lap}}^{f,b}$: On input $\boldsymbol{x} \in \mathcal{X}$, output $f(\boldsymbol{x}) + \mathrm{Lap}(b)$.

In this paper, we use the Laplace mechanism with appropriate parameters to achieve robustness. We also extend the above mechanism $\mathcal{M}_{\mathrm{Lap}}^{f,b}$ to obtain differential privacy for sampling queries, i.e., when f is a randomized function.

3 Lossy Wasserstein Distance, Flexible Accuracy, and Robust Privacy

In this section, we formally define lossy Wasserstein distance, flexible accuracy, and robust privacy.

3.1 Lossy Wasserstein Distance

Central to the formalization of all the results in this work is a new notion of distance between distributions over a metric space, that we call *lossy Wasserstein distance.* Lossy Wasserstein distance generalizes the notion of Wasserstein distance, or Earth Mover Distance, which is the minimum cost of transporting probability mass ("earth") of one distribution to make it match the other. Loss refers to the fact that some of the mass is allowed to be lost during this transportation. We shall use the "infinity norm" version, where the cost paid is the maximum distance any mass is transported.

Formally, consider a metric space with ground set Ω, and metric $\mathfrak{d}$, where Wasserstein distance can be defined. For example, one may consider $\Omega = \mathbb{R}^n$ and the metric $\mathfrak{d}$ being an ℓ_p-metric.

For $\theta \in [0, 1]$, and distributions P, Q over the metric space $(\Omega, \mathfrak{d})$, we define the set of *θ-lossy couplings of P and Q*, $\Phi^{\theta}(P, Q)$ as consisting of joint distributions ϕ over Ω^2 with marginals ϕ_1 and ϕ_2 such that $\Delta(\phi_1, P) + \Delta(\phi_2, Q) \leq \theta$. Note that $\Phi^0(P, Q)$ consists of joint distributions with marginals exactly equal to P and Q.

Definition 4 (θ-Lossy ∞-Wasserstein Distance)**.** Let P and Q be two distributions over a metric space $(\Omega, \mathfrak{d})$. For $\theta \in [0,1]$, the θ-lossy ∞-Wasserstein distance between P and Q is defined as:

$$W_\theta^\infty(P,Q) = \inf_{\phi \in \Phi^\theta(P,Q)} \max_{(x,y) \leftarrow \phi} \mathfrak{d}(x,y). \tag{1}$$

For simplicity, we write $W^\infty(p,q)$ to denote $W_0^\infty(p,q)$.

Lossy ∞-Wasserstein distance is a generalization of the guarantee of being "Probably Approximately Correct" (PAC). A PAC guarantee states that a randomized quantity G is, except with some small probability γ, within an approximation radius β of a desired *deterministic* quantity f: i.e., $\Pr_{g \leftarrow G}[|g-f| > \beta] \leq \gamma$. Representing f by a point distribution F_f, this can be equivalently written as $W_\gamma^\infty(F_f, G) \leq \beta$. It also generalizes total variation distance between two distributions $\frac{1}{2}\|F - G\|_1$ (treating distributions as probability vectors), since $W_\gamma^\infty(F,G) = 0$ iff $\|F-G\|_1 \leq \gamma$.

Now we discuss further important properties of W_θ^∞, namely, the triangle inequality (Lemma 3.1) and the effect of adding independent noise (Lemma 3.2), both of which we prove in Appendix A. These properties will be useful in proving our results.

Lemma 3.1. *For distributions P, Q, and R over a metric space $(\Omega, \mathfrak{d})$, and for $\gamma_1, \gamma_2 \in [0,1]$, we have.*

$$W_{\gamma_1+\gamma_2}^\infty(P,R) \leq W_{\gamma_1}^\infty(P,Q) + W_{\gamma_2}^\infty(Q,R). \tag{2}$$

To study the effect of adding independent noise on W_θ^∞, we need the metric to satisfy additional conditions. Specifically, we shall require that the metric corresponds to the metric in a normed vector space (e.g., $\mathbb{R}^d$).

Lemma 3.2. *Let X, Y and Z be three random variables over a normed vector space, Ω with Z being independent of X and Y, and let $\mathfrak{p}_0$ denote the distribution with all its mass at 0. Then, the lossy ∞-Wasserstein distance defined using the metric induced by the norm of Ω satisfies the following: $\forall \gamma, \gamma_1$ such that $\gamma, \gamma_1 \geq 0$ and $\gamma_1 \leq \gamma/2$, we have*

$$W_\gamma^\infty(X+Z, Y+Z) \overset{\text{(a)}}{\leq} W_\gamma^\infty(X,Y) \overset{\text{(b)}}{\leq} W_{\gamma-2\gamma_1}^\infty(X+Z, Y+Z) + 2W_{\gamma_1}^\infty(\mathfrak{p}_0, Z). \tag{3}$$

3.1.1 Average Version of the Lossy Wasserstein Distance

Our definition of W_θ^∞ uses a worst case notion of distance. Many of the results using this notion have analogues using an average case version. We present this definition below, as it may be of interest elsewhere.

Definition 5 (θ-Lossy Average Wasserstein Distance)**.** Let P and Q be two probability distributions over a metric space $(\Omega, \mathfrak{d})$, and let $\theta \in [0,1]$. The *θ-lossy average Wasserstein distance* between P and Q is defined as:

$$W_\theta(P,Q) = \inf_{\phi \in \Phi^\theta(P,Q)} \mathop{\mathbb{E}}_{(x,y) \leftarrow \phi} [\mathfrak{d}(x,y)]. \tag{4}$$

The following lemma relates lossy average Wasserstein and lossy ∞-Wasserstein distances, and is proved in Appendix A.1.

Lemma 3.3. *For any two distributions P, Q, and $0 \leq \beta' < \beta \leq 1$,*

$$W_\beta(P,Q) \leq W_\beta^\infty(P,Q) \leq \frac{W_{\beta'}(P,Q)}{(\beta - \beta')}.$$

3.2 Flexible Accuracy

The high-level idea of flexible accuracy is to allow for some *distortion of the input* before measuring accuracy. We would like to define "natural" distortions of a database, that are meaningful for the function in question. For many functions, removing a few entries (say, outliers) would be a natural distortion. On the other hand, *adding* new entries – even just one – is often not a reasonable distortion. As such, distortion is defined not using a metric over databases, but a *quasi-metric* (which is not required to be symmetric). We shall use quasi-metrics with range $\mathbb{R}_{\geq 0} \cup \{\infty\}$, where ∞ indicates that one database cannot be distorted into another one.

Definition 6 (Measure of Distortion)**.** A *measure of distortion* on a set $\mathcal{X}$ is a function $\partial : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ which forms a quasi-metric over $\mathcal{X}$.

An important example of a measure of distortion is ∂_{drop}. It is defined when each element in $\mathcal{X}$ is a finite multiset over a ground set $\mathcal{G}$. Formally, $\boldsymbol{x} \in \mathcal{X}$ is a function $\boldsymbol{x} : \mathcal{G} \to \mathbb{N}$ that outputs the multiplicity of each element of $\mathcal{G}$ in $\boldsymbol{x}$. Then, for finite $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$, we define

$$\partial_{\mathrm{drop}}(\boldsymbol{x}, \boldsymbol{x}') := \begin{cases} \frac{\sum_{g \in G} \boldsymbol{x}(g) - \boldsymbol{x}'(g)}{\sum_{g \in G} \boldsymbol{x}(g)} & \text{if } \forall g \in \mathcal{G}, \boldsymbol{x}(g) \geq \boldsymbol{x}'(g), \\ \infty & \text{otherwise.} \end{cases} \tag{5}$$

That is, $\partial_{\mathrm{drop}}(\boldsymbol{x}, \boldsymbol{x}')$ measures the fraction of elements in $\boldsymbol{x}$ that are to be dropped for it to become $\boldsymbol{x}'$ (unless $\boldsymbol{x}'$ cannot be derived thus). We present other useful examples of measures of distortion in Section 5.1.5, which also allow moving the data points (when $\mathcal{G}$ is a metric-space).

Informally, flexible accuracy with a distortion bound α guarantees that on input $\boldsymbol{x}$, a mechanism shall produce an output that corresponds to $f(\boldsymbol{x}')$ for some $\boldsymbol{x}'$ such that $\partial(\boldsymbol{x}, \boldsymbol{x}') \leq \alpha$. In addition to such input distortion, we may allow the output to be also probably approximately correct, with an approximation error parameter β and an error probability parameter γ. Formally, the probabilistic approximation guarantee of the output is given as a bound of β on a γ-lossy ∞-Wasserstein distance.

Definition 7 ((α, β, γ)-accuracy)**.** Let ∂ be a measure of distortion on a set $\mathcal{X}$ and $f : \mathcal{X} \to \mathcal{Y}$ be a randomized function such that $\mathcal{Y}$ admits a metric. A mechanism $\mathcal{M}$ is said to be (α, β, γ)*-accurate for* f *with respect to* ∂, if for each $x \in \mathcal{X}$, there is a random variable X' with support contained in $\{\boldsymbol{x}' | \partial(\boldsymbol{x}, \boldsymbol{x}') \leq \alpha\}$ such that $W_\gamma^\infty(f(X'), \mathcal{M}(\boldsymbol{x})) \leq \beta$.

Flexible accuracy generalizes existing accuracy definitions. In particular:

- As mentioned in Section 3.1, $(0, \beta, \gamma)$-accuracy already extends the PAC guarantees.
- $(\alpha, 0, 0)$-accuracy for specific functions like median was implicitly used in other contexts in the DP literature [BSU16].
- Blum et al. [BLR13] introduced *usefulness* to measure accuracy with respect to a "perturbed" function. While adequate for the function classes they considered (half-space queries, range queries etc.), it is not applicable to queries like maximum. Flexible accuracy generalizes usefulness (see Section 5.1.6).

As we show later, flexible accuracy lets us develop DP mechanisms for highly sensitive functions (*e.g.*, max), for which existing DP mechanisms offered only limited, if not vacuous, guarantees.

3.3 Robust Privacy

We define a mechanism whose output is in a metric space to be robustly private if, roughly, it holds that whenever two input distributions result in output distributions that are close in Wasserstein distance, then the output distributions are also "indistinguishable." Unlike in the definition of differential privacy, where an input neighborhood is specified, here the neighborhood is implicitly defined by the mechanism itself.

To understand robust privacy, consider utility of a mechanism as being unaffected by small perturbations of the output. Now, the output of the mechanism may contain hints as to the input, which do not contribute to its utility. Robust privacy deals with removing such "non-utile signals" in the output. In general, modifying the mechanism to remove these hints could also result in degradation of the utility, and render more information non-utile. A robust mechanism could be seen as a *fixed point* of this iteration of removing non-utile signals.

Before formally defining robust privacy, we define closeness and indistinguishability of two distributions. Below, $\theta, \rho, \epsilon, \delta \in \mathbb{R}_{\geq 0}$, with $\delta, \theta \leq 1$.

Definition 8 ((ρ, θ)-closeness, (ϵ, δ)-indistinguishability)**.** A pair of distributions D_0 and D_1 over a set $\mathcal{Y}$ with a metric $\mathfrak{d}$ are said to be (ρ, θ)-*close* w.r.t. $\mathfrak{d}$ if $W_\theta^\infty(D_0, D_1) \leq \rho$ (where W_θ^∞ is defined w.r.t. $\mathfrak{d}$). D_0 and D_1 are said to be (ϵ, δ)-*indistinguishable* if for all measurable subsets $S \subseteq \mathcal{Y}$, and $b \in \{0, 1\}$, it holds that $\Pr_{y \leftarrow D_b}[y \in S] \leq e^\epsilon \cdot \Pr_{y \leftarrow D_{1-b}}[y \in S] + \delta$.

Closeness models that the two distributions are almost equally useful in all applications that are robust to small perturbations (e.g., the application does not rely on steganographic information encoded in the less significant digits). The definition of indistinguishability above is as used in differential privacy (Definition 2): a mechanism $\mathcal{M}$ is (ϵ, δ)-DP iff for all $\boldsymbol{x} \sim \boldsymbol{x}'$, $\mathcal{M}(\boldsymbol{x})$ and $\mathcal{M}(\boldsymbol{x}')$ are (ϵ, δ)-indistinguishable.

For a distribution P over $\mathcal{X}$, we write $\mathcal{M}(P)$ to denote the output distribution of $\mathcal{M}$ when its input is drawn from P.

Definition 9 ($(\rho, \theta, \epsilon, \delta)$-robust privacy)**.** A mechanism $\mathcal{M} : \mathcal{X} \to \mathcal{Y}$ is said to be $(\rho, \theta, \epsilon, \delta)$-*robustly private* w.r.t. a metric $\mathfrak{d}$ over $\mathcal{Y}$ if, for all distributions P, Q over $\mathcal{X}$ such that $\mathcal{M}(P)$ and $\mathcal{M}(Q)$ are (ρ, θ)-close w.r.t. $\mathfrak{d}$, $\mathcal{M}(P)$ and $\mathcal{M}(Q)$ are also (ϵ, δ)-indistinguishable.

We remark that by only referring to the output of $\mathcal{M}$ rather than a desired function f, the definition of robust privacy avoids baking in a specific notion of accuracy into it. Thus, as in the case of DP, it can be used in combination with separately defined notions of accuracy. Further, again like DP, this definition does not rely on data distributions (in contrast with the distributional privacy notion of [BLR13]).

4 Composition Theorems

It is often convenient to design a mechanism as the function composition of two mechanisms, $\mathcal{M} = \mathcal{M}_2 \circ \mathcal{M}_1$. We present "composition theorems" which yield flexible accuracy, differential privacy, and robust privacy guarantees for $\mathcal{M}$ in terms of those for $\mathcal{M}_1$ and $\mathcal{M}_2$.

4.1 Flexible Accuracy Under Composition

As we shall need distortion measure between two distributions in our accuracy guarantees, it is rather useful to extend the distortion measure to distributions. This can be done in same way as W^∞, but with respect to a quasi-metric rather than a metric.

Definition 10 (Extension of a Measure of Distortion to Distributions)**.** For a measure of distortion ∂ over a set $\mathcal{X}$, we define $\widehat{\partial}$ as its extension to distribution, which maps a pair of distributions P, Q over $\mathcal{X}$ to a real number as

$$\widehat{\partial}(P, Q) = \inf_{\phi \in \Phi^0(P,Q)} \sup_{\substack{(x,y): \\ \phi(x,y) \neq 0}} \partial(x, y).$$

If P is a point distribution with all its mass on a point x, we denote $\widehat{\partial}(P, Q)$ as $\widehat{\partial}(x, Q)$.

The following lemmas, proven in Appendix B.1, show that flexible accuracy w.r.t. ∂ implies a similar condition w.r.t. $\widehat{\partial}$.

Lemma 4.1. *If ∂ is a measure of distortion over A, then $\widehat{\partial}$ is a quasi-metric.*

Lemma 4.2. *If $\mathcal{M} : A \to B$ is an (α, β, γ)-accurate mechanism for f w.r.t. ∂, then for any random variable X over A, there is a random variable X^* such that*

$$\widehat{\partial}(X, X^*) \leq \alpha, \qquad W_\gamma^\infty(f(X^*), \mathcal{M}(X)) \leq \beta.$$

Now we define two new sensitivity notions: *distortion sensitivity* for a function and *error sensitivity* for a mechanism. These notions will be used in our composition theorem for flexible accuracy.

Definition 11 (Distortion sensitivity)**.** Let $f : A \to B$ be a randomized function where B admits Wasserstein distances. Let ∂_1, ∂_2 be measures of distortion on A, B respectively. Then, the *distortion-sensitivity* of f w.r.t. (∂_1, ∂_2) and $\theta \in [0,1]$ and $\omega \geq 0$, is defined as the function $\sigma_f^{\theta,\omega} : \mathbb{R}_{\geq 0} \cup \{\infty\} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ given by

$$\sigma_f^{\theta,\omega}(\alpha) = \sup_{\substack{x,Y: \\ \widehat{\partial}_2(f(x),Y) \leq \alpha}} \inf_{\substack{X: \\ W_\theta^\infty(f(X),Y) \leq \omega}} \widehat{\partial}_1(x, X) \tag{6}$$

where $x \in A$, and the random variables X and Y are distributed over A and B, respectively. Above, infimum over an empty set is defined to be ∞.

In all the applications of our main results in this paper, we apply distortion sensitivity with $\theta = \omega = 0$. So, for simplicity, we write σ_f^θ to denote $\sigma_f^{\theta,0}$ and σ_f to denote $\sigma_f^{0,0}$.

To intuitively understand the notion of distortion sensitivity, assume $\theta = \omega = 0$. With this, if a function f has distortion sensitivity σ_f, then arbitrarily distorting $f(x)$ within an α radius can be modeled as distorting x within a $\sigma_f(\alpha)$ radius.

Remark 4.1. *The distortion sensitivity of the identity function $f : A \to A$ with respect to $(\widehat{\partial}, \widehat{\partial})$ for any distortion measure $\widehat{\partial}$ over A satisfies $\sigma_f(\alpha) \leq \alpha$.*

Definition 12 (Error sensitivity)**.** Let $\mathcal{M} : A \to B$ be any mechanism for a function $f : A \to B$, where both A and B have associated Wasserstein distances. Let $\widehat{\partial}$ be a measure of distortion on A. Then, for $\alpha_2, \gamma_2 \geq 0$, the error-sensitivity $\tau_{\mathcal{M}}^{\alpha_2,\gamma_2} : \mathbb{R}_{\geq 0} \times [0,1] \to \mathbb{R}_{\geq 0}$ of $\mathcal{M}$ w.r.t. f is defined as:

$$\tau_{\mathcal{M},f}^{\alpha_2,\gamma_2}(\beta_1, \gamma_1) = \sup_{X,X': W_{\gamma_1}^\infty(X,X') \leq \beta_1} \ \inf_{Y: \widehat{\partial}(X',Y) \leq \alpha_2} W_{\gamma_2}^\infty(\mathcal{M}(X), f(Y)). \tag{7}$$

In other words, if $\tau_{\mathcal{M},f}^{\alpha_2,\gamma_2}(\beta_1, \gamma_1) = \beta_2$, then for distributions X, X' over A which are (β_1, γ_1)-close to each other, one can α_2-distort X' to Y in such a way that $\mathcal{M}(X)$ and $f(Y)$ are (β_2, γ_2)-close to each other. Note that $\tau_{\mathcal{M},f}$ generalizes the parameters of flexible accuracy: Any mechanism $\mathcal{M}$ for computing a function f is $(\alpha, \tau_{\mathcal{M},f}^{\alpha,\gamma}(0,0), \gamma)$-accurate.

We need the following lemma, shown in Appendix B.1, to prove our composition theorem for flexible accuracy.

Lemma 4.3. *Suppose $f : A \to B$ has distortion sensitivity $\sigma_f^{\theta,\omega}$, w.r.t. (∂_1, ∂_2). Then, for random variables X_0 over A, and Y over B such that $\widehat{\partial}_2(f(X_0), Y) \leq \alpha$, there exists a distribution X over A such that $W_\theta^\infty(f(X), Y) \leq \omega$ and $\widehat{\partial}_1(X_0, X) \leq \sigma_f^{\theta,\omega}(\alpha)$.*

Having defined the distortion and error sensitivities of a mechanism, we shall now see how they play in a composition $M_2 \circ M_1$ for $f_2 \circ f_1$ where M_1, M_2 are mechanisms with (α, β, γ) accuracy guarantees.

Theorem 4.1 (Flexible Accuracy Composition)**.** *Let $M_1 : A \to B$ and $M_2 : B \to C$ be mechanisms, respectively with $(\alpha_1, \beta_1, \gamma_1)$-accuracy for $f_1 : A \to B$ and $\tau_{\mathcal{M}_2,f_2}$ error sensitivity for $f_2 : B \to C$, with respect to measures of distortion ∂_1, ∂_2 defined on A, B and metrics $\mathfrak{d}_1, \mathfrak{d}_2$ defined on B, C respectively. Then, for any $\theta \in [0,1]$, $\omega \geq 0$, $\alpha_2 \geq 0$ and $\gamma_2 \in [0,1]$, the mechanism $M_2 \circ M_1 : A \to C$ is (α, β, γ)-accurate for the function $f_2 \circ f_1$ w.r.t. ∂_1 and $\mathfrak{d}_2$, where $\alpha = \alpha_1 + \sigma_{f_1}^{\theta,\omega}(\alpha_2)$, $\beta = \tau_{M_2,f_2}^{\alpha_2,\gamma_2}(\beta_1, \gamma_1) + \tau_{f_2,f_2}^{0,\theta}(\omega, \theta)$, and $\gamma = \gamma_2 + \theta$.*

Theorem 4.1 is a general composition theorem for flexible accuracy. In all the applications in this paper, we use a simplified version of the above composition theorem with $\theta = \omega = 0$. An illustration of how the composition theorem works is given as a pebbling game in Figure 1.

Proof of Theorem 4.1. To compare $f_2 \circ f_1$ and $\mathcal{M}_2 \circ \mathcal{M}_1$, we consider the hybrid mechanism $\mathcal{M}_2 \circ f_1$, For a given element $x \in A$, since $\mathcal{M}_1$ is $(\alpha_1, \beta_1, \gamma_1)$-accurate mechanism for f_1, there exists a random variable X' with distribution X' such that,

$$\widehat{\partial}_1(x, X') \leq \alpha_1 \tag{8}$$

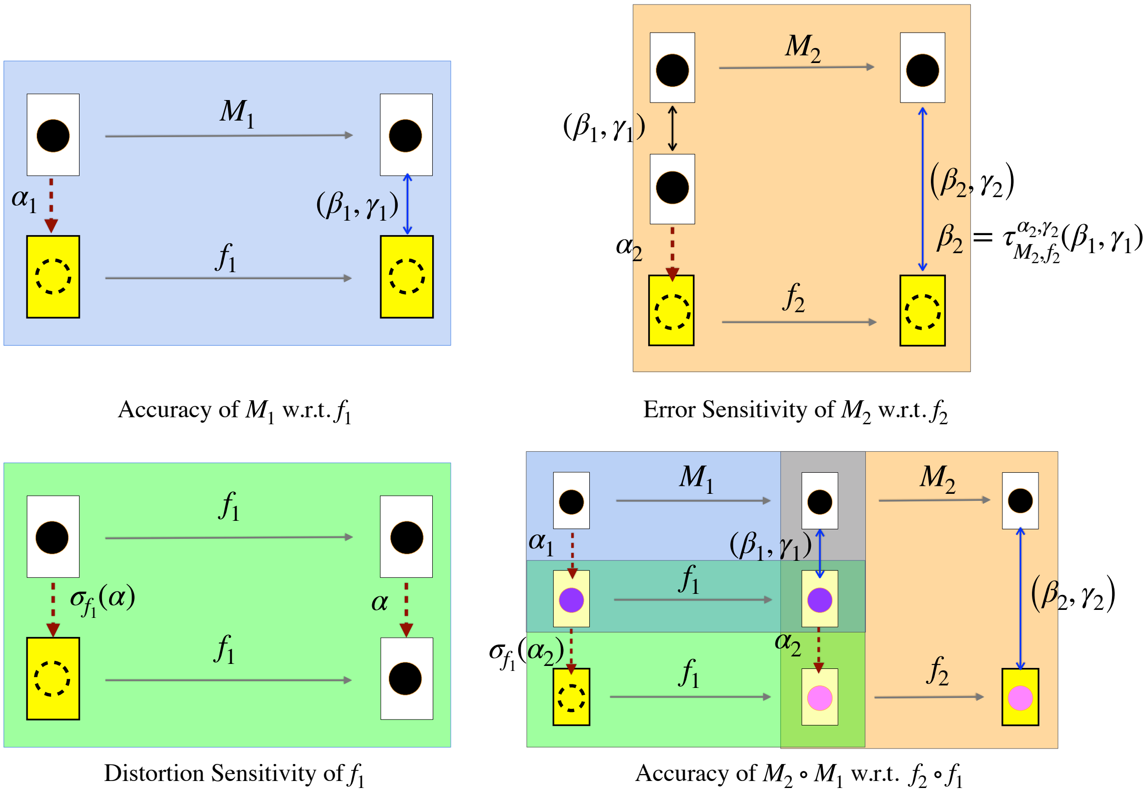

Figure 1: An illustration of the composition theorem, Theorem 4.1. Dotted arrows indicate closeness in terms of distortion between histograms (or distributions thereof), and the solid arrows indicate closeness in terms of lossy Wasserstein distance. Each figure shows the corresponding guarantee (accuracy, error sensitivity or distortion sensitivity) as a pebbling game: The white boxes with black pebbles corresponds to given histograms, and the yellow boxes indicate histograms that are guaranteed to exist, such that the given closeness relations hold. This allows those boxes to be pebbled. Accuracy guarantee of $M_2 \circ M_1$ is derived by first applying the pebbling rule of accuracy of M_1 (to obtain the purple pebbles), then that of the error sensitivity of M_2 (to get the pink pebbles) and finally using the pebbling rule of the distortion sensitivity of f_1 to pebble the remaining yellow box.

$$W^{\infty}_{\gamma_1}(f_1(X'), \mathcal{M}_1(x)) \leq \beta_1. \tag{9}$$

Now, applying the mechanism $\mathcal{M}_2$ on $\mathcal{M}_1(x)$, we incur an overall error by at most $\tau^{\alpha_2,\gamma_2}_{\mathcal{M}_2,f_2}(\beta_1, \gamma_1)$ to the output of function f_2 over a distorted input (see Definition 12). Therefore, there exists a random variable Y^* such that,

$$\widehat{\partial}_2(f_1(X'), Y^*) \leq \alpha_2, \tag{10}$$

$$W^{\infty}_{\gamma_2}(f_2(Y^*), \mathcal{M}_2(\mathcal{M}_1(x))) \leq \tau^{\alpha_2,\gamma_2}_{\mathcal{M}_2,f_2}(\beta_1, \gamma_1) \tag{11}$$

From (10) and Lemma 4.3, there exists X over A such that

$$\widehat{\partial}_1(X', X) \leq \sigma^{\theta,\omega}_{f_1}(\alpha_2) \tag{12}$$

$$W^{\infty}_{\theta}(f_1(X), Y^*) \leq \omega. \tag{13}$$

Now by Lemma 4.1, since ∂_1 being a quasi-metric, so is $\widehat{\partial}_1$. Hence, combined with (8), we have

$$\widehat{\partial}_1(x, X) \leq \alpha_1 + \sigma_{f_1}^{\theta,\omega}(\alpha_2). \tag{14}$$

Now, using triangle inequality from Lemma 3.1, we can write

$$\begin{aligned} W_{\gamma_2+\theta}^{\infty}(f_2(f_1(X)), \mathcal{M}_2(\mathcal{M}_1(x))) \leq\ & W_{\gamma_2}^{\infty}(f_2(Y^*), \mathcal{M}_2(\mathcal{M}_1(x))) \\ & + W_{\theta}^{\infty}(f_2(f_1(X)), f_2(Y^*)) \end{aligned} \tag{15}$$

We can bound the first term on the RHS of (15) using (11). This yields

$$W_{\gamma_2+\theta}^{\infty}(f_2(f_1(X)), \mathcal{M}_2(\mathcal{M}_1(x))) \leq \tau_{\mathcal{M}_2,f_2}^{\alpha_2,\gamma_2}(\beta_1, \gamma_1) + W_{\theta}^{\infty}(f_2(f_1(X)), f_2(Y^*)). \tag{16}$$

Note that we have $W_{\theta}^{\infty}(f_1(X), Y^*) \leq \omega$ from (13). Applying f_2 on $f_1(X)$ and Y^* and using Definition 12 gives $W_{\theta}^{\infty}(f_2(f_1(X)), f_2(Y^*)) \leq \tau_{f_2,f_2}^{0,\theta}(\omega, \theta)$. Substituting this back in (16) gives $W_{\gamma_2+\theta}^{\infty}(f_2(f_1(X)), \mathcal{M}_2(\mathcal{M}_1(x))) \leq \tau_{\mathcal{M}_2,f_2}^{\alpha_2,\gamma_2}(\beta_1, \gamma_1) + \tau_{f_2,f_2}^{0,\theta}(\omega, \theta)$. This, together with (14), concludes the proof of Theorem 4.1.

□

4.2 Differential Privacy and Robust Privacy Under Composition

We prove "pre-processing" theorems for both differential privacy as well as robust privacy. The pre-processing theorem for DP can be viewed as complementing the "post-processing" theorem for DP (see [DR14, Proposition 2.1]), which states that if $\mathcal{M}_1$ is (ϵ, δ)-DP, then for any mechanism $\mathcal{M}_2$, the composed mechanism $\mathcal{M}_2 \circ \mathcal{M}_1$ would remain (ϵ, δ)-DP. Our pre-processing theorem for DP states that if $\mathcal{M}_2$ is private, then so would $\mathcal{M}_2 \circ \mathcal{M}_1$ be (i.e., pre-processing does not hurt privacy), provided that $\mathcal{M}_1$ is well behaved.Following is a notion of being well-behaved that suffices for our purposes.

Definition 13 (Neighborhood preserving Mechanism)**.** A mechanism $\mathcal{M} : A \to B$ is *neighborhood preserving* w.r.t. neighborhood relations $\sim_A$ over A and $\sim_B$ over B, if for all $x, y \in A$ s.t. $x \sim_A y$, there exists a pair of jointly distributed random variables (X, Y) s.t. $X = \mathcal{M}(x)$, $Y = \mathcal{M}(y)$, and $\Pr[X \sim_B Y] = 1$.

Theorem 4.2 (Differential Privacy Composition)**.** *Let $\mathcal{M}_1 : A \to B$ and $\mathcal{M}_2 : B \to C$ be any two mechanisms. If $\mathcal{M}_1$ is neighborhood-preserving w.r.t. neighborhood relations $\sim_A$ and $\sim_B$ over A and B, respectively, and $\mathcal{M}_2$ is (ϵ, δ)-DP w.r.t. $\sim_B$, then $\mathcal{M}_2 \circ \mathcal{M}_1 : A \to C$ is (ϵ, δ)-DP w.r.t. $\sim_A$.*

Proof sketch. Since M_1 is a neighborhood-preserving mechanism, two elements which are neighbors in the input space of M_1 are also neighbors in the input space of M_2. Thus, for any two neighbors $x, y \in A$, $\mathcal{M}_2 \circ \mathcal{M}_1(x)$ and $\mathcal{M}_2 \circ \mathcal{M}_1(y)$ are (ϵ, δ)-indistinguishable. We formalize this intuition and provide a complete proof in Appendix B.2. □

Theorem 4.3 (Robust Privacy Composition)**.** *Let $\mathcal{M}_1 : A \to B$ and $\mathcal{M}_2 : B \to C$ be any two mechanisms. If $\mathcal{M}_2$ is $(\rho, \theta, \epsilon, \delta)$-robustly private w.r.t. a metric $\mathfrak{d}$, then $\mathcal{M}_2 \circ \mathcal{M}_1$ is also $(\rho, \theta, \epsilon, \delta)$-robustly private w.r.t. the metric $\mathfrak{d}$.*

Proof. Since robust privacy is defined entirely in terms of the output distributions, it is always preserved under pre-processing. In the following, we make this argument formal.

In order to prove that $\mathcal{M}_2 \circ \mathcal{M}_1$ is $(\rho, \theta, \epsilon, \delta)$-robustly private, we need to show that for any two input distributions P and Q, if $\mathcal{M}_2 \circ \mathcal{M}_1(P)$ and $\mathcal{M}_2 \circ \mathcal{M}_1(Q)$ are (ρ, θ)-close, then they are also (ϵ, δ)-indistinguishable. Let us denote $\mathcal{M}_1(P)$ and $\mathcal{M}_1(Q)$ by P' and Q', respectively. We now need to prove that if $\mathcal{M}_2(P')$ and $\mathcal{M}_2(Q')$ are (ρ, θ)-close, then they are also (ϵ, δ)-indistinguishable. But since $\mathcal{M}_2$ is $(\rho, \theta, \epsilon, \delta)$-robustly private, this statement trivially holds by definition of robust privacy (see Definition 9). □

5 Mechanisms

In this section, we propose and analyze concrete mechanisms for several important functions. First, we present a new DP mechanism for the histogram function with flexible accuracy and then extend it to any "histogram based statistic" (*e.g.*, max and support).

We also show that by using appropriate measures of distortion, all the error in the above mechanism can be attributed to input distortion, allowing it to be used to obtain DP mechanisms for arbitrary function families via composition. This yields DP mechanisms for complex (and highly sensitive) hypotheses classes in machine learning like *maximum-margin separators*, provided that the accompanying flexible accuracy guarantee is acceptable. Finally, we also note that, mechanisms (e.g., for half-space queries) which required [BLR13] to introduce the accuracy notion of *usefulness* can be cast in the framework of flexible accuracy.

For clear exposition of ideas, we defer the discussion on robust privacy of these mechanisms to Section 5.2, where we present a general compiler that takes any mechanism (over the reals) and, via composition, makes it robustly private, without (significantly) degrading its flexible accuracy or differential privacy guarantees.

We also present a mechanism for differentially-private sampling that also achieves our new notions of flexible accuracy and robust privacy in Section 5.3.

5.1 Exploiting Flexible Accuracy

Multisets form an abstraction of datasets that is widely applicable. Each element in the multiset can be a vector encoding attributes (including labels), or more abstractly, belongs to a ground set $\mathcal{G}$. Formally, a multiset $\boldsymbol{x}$ over the ground set $\mathcal{G}$ is a function $\boldsymbol{x} : \mathcal{G} \to \mathbb{N}$ that outputs the multiplicity of elements in $\mathcal{G}$. The *size* and *support* of $\boldsymbol{x}$ are defined as $|\boldsymbol{x}| := \sum_{i\in\mathcal{G}} \boldsymbol{x}(i)$ and $\mathrm{supp}(\boldsymbol{x}) := \{i \in \mathcal{G} : \boldsymbol{x}(i) \neq 0\}$, respectively. We shall be interested in finite-sized multisets, which we refer to as histograms. We denote the domain of all histograms over $\mathcal{G}$ by $\mathbb{H}_{\mathcal{G}}$. For DP, the standard notion of neighborhood among histograms is defined as $\boldsymbol{x} \sim_{\text{hist}} \boldsymbol{x}'$ iff $\sum_{i\in\mathcal{G}} |\boldsymbol{x}(i) - \boldsymbol{x}'(i)| \leq 1$. Later, we shall also require $\mathcal{G}$ to be a metric space, endowed with a metric $\mathfrak{d}$.

We start with a new mechanism, the truncated Laplace mechanism, $\mathcal{M}^{\tau,\mathcal{G}}_{\text{trLap}} : \mathbb{H}_{\mathcal{G}} \to \mathbb{H}_{\mathcal{G}}$ (Construction 1) for the identity function. It simply *decreases* the multiplicity of each element by a bounded quantity sampled from a carefully chosen distribution. Our choice of this noise distribution – a truncated Laplace distribution – lets us prove the following DP and accuracy guarantees.

5.1.1 Derivation of the Truncated Laplace Mechanism

As a warm up, consider a Boolean task of reporting whether a given set is empty or not. The only input distortion we are allowed is to drop some elements – i.e., we cannot report an empty set as non-empty. Since we seek to limit the extent of distortion, let us add a constraint that if a set has q or more elements, then with probability 1 (or very close to 1) we should report the set as being non-empty. Let p_k denote the probability that a set of size $k \in [0, q]$ is reported as being non-empty, so that $p_0 = 0$ and $p_q = 1$.

Now, for privacy, we consider two sets to be neighbors if their sizes differ by at most one. For our scheme to be (ϵ, δ)-differential private, we require

$$\begin{aligned} p_k &\leq p_{k+1}e^\epsilon + \delta, \qquad & p_{k+1} &\leq p_k e^\epsilon + \delta \\ (1-p_k) &\leq (1-p_{k+1})e^\epsilon + \delta, \qquad & (1-p_{k+1}) &\leq (1-p_k)e^\epsilon + \delta, \end{aligned}$$

for $0 \leq k < q$, with boundary conditions $p_0 = 0$ and $p_q = 1$. We are interested in simultaneously reducing ϵ and δ subject to the above constraints. The pareto-optimal (ϵ, δ) turn out to be given by $\delta\left(\frac{e^{(q/2)\epsilon}-1}{e^\epsilon-1}\right) = \frac{1}{2}$, with corresponding values of p_k being given by

$$p_k = \delta\left(\frac{e^{k\epsilon}-1}{e^\epsilon-1}\right), \quad k \leq q/2 \qquad p_k = 1 - p_{q-k}, \quad k \geq q/2. \tag{17}$$

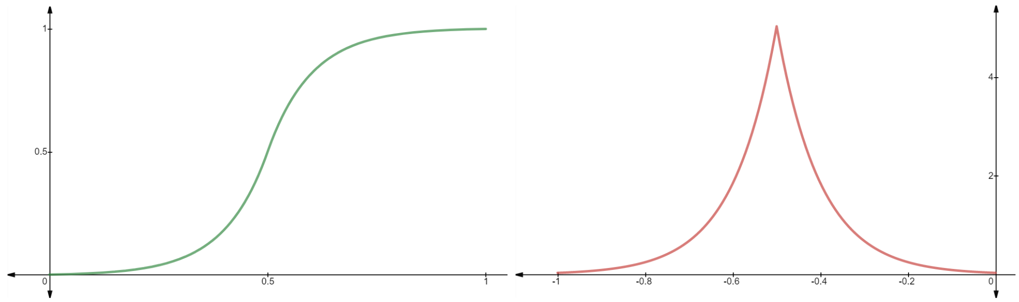

Figure 2: The probability function in the optimal mechanism for reporting whether a set is empty or not (left), which can be interpreted as adding a noise according to a truncated Laplace distribution with a negative mean (right).

In particular, we may choose $\epsilon = O\left(\frac{1}{\sqrt{q}}\right)$, and $\delta = O\left(\frac{e^{-\sqrt{q}/2}}{\sqrt{q}}\right)$, providing a useful privacy guarantee when q is sufficiently large.

In Figure 2, on the left, we plot the probabilities p_k against k/q for this choice of (ϵ, δ).

To generalize this Boolean mechanism to a full-fledged histogram mechanism, we reinterpret it. In a histogram mechanism, where again, the distortion allowed in the input is to only drop elements, we can add a *negative noise* to the count in each "bar" of the histogram. (If the reduced count is negative, we report it as 0.) We seek a noise function such that the probability of the reported count being 0 (when the actual count is $k \in [0, q]$) is the same as that of the above mechanism reporting that a set of size k is empty. That is, the probability of adding a noise $\nu \le -k$ should be $1 - p_k$. That is, if the noise distribution is given by the density function σ, we require that

$$\int_{-q}^{-k} \sigma(t) \cdot dt = 1 - p_k \tag{18}$$

$$\sigma(t) = 0 \qquad \text{for } t \notin [-q, 0]$$

Substituting the expression for p_k from (17), and then differentiating this identity with respect to k, we obtain the following expression for $\sigma(t)$, for $t \in [-q, 0]$:

$$\sigma(t) = \frac{1}{1 - e^{-\epsilon q/2}} \mathrm{Lap}(t \mid -\frac{q}{2}, \frac{1}{\epsilon}),$$

where Lap is the Laplace noise distribution with mean $-\frac{q}{2}$ and scale parameter $1/\epsilon$. The plot on the right side in Figure 2 shows this noise distribution.

The final histogram mechanism (Construction 1) is derived by adding this noise to each bar of the histogram, followed by rounding to the nearest integer (or to 0, if it is negative).

5.1.2 The Truncated Laplace Mechanism

We describe the truncated Laplace mechanism (Construction 1) in Figure 3. The goal of this mechanism is to output a histogram by distorting the given histogram as little as possible, while obtaining a DP guarantee. The following theorem summarizes the privacy and flexible accuracy guarantees achieved by $\mathcal{M}_{\text{trLap}}^{\tau,\mathcal{G}}$.

Theorem 5.1. *On histograms of size at least n, i.e., $|\boldsymbol{x}| > n$, $\mathcal{M}_{\text{trLap}}^{\tau,\mathcal{G}}$ from Construction 1 satisfies the following guarantees:*

- $\mathcal{M}_{\text{trLap}}^{\tau,\mathcal{G}}$ *is* $\left(O(\frac{1}{\sqrt{\tau n}}), e^{-\Omega(\sqrt{\tau n})}\right)$*-differentially private w.r.t.* $\sim_{\text{hist}}$*, and*

- *If* $|\mathrm{supp}(\boldsymbol{x})| \le t$, *then* $\mathcal{M}^{\tau,\mathcal{G}}_{\mathrm{trLap}}$ *is* $(\tau t, 0, 0)$*-accurate for the identity function, w.r.t. the distortion measure* ∂_{drop}.

Note that, since we obtain $(\alpha, 0, 0)$-accuracy in the result above, we did not use any metric over $\mathbb{H}_{\mathcal{G}}$ for stating it. This result shows that as the database size grows, $\mathcal{M}^{\tau,\mathcal{G}}_{\mathrm{trLap}}$ can be used to obtain (ϵ, δ)-DP guarantee where ϵ tends to 0. The following version of this result, stated for $\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\mathrm{trLap}}$, allows setting the parameter ϵ directly (and is used in our empirical comparisons):

Theorem 5.2. *On inputs* $\boldsymbol{x}$ *s.t.* $|\boldsymbol{x}| \ge \frac{2}{\epsilon\tau} e^{-\frac{\epsilon\tau}{2}}$, $\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\mathrm{trLap}}$ *is* $\left(\epsilon, \frac{e^{\epsilon}-1}{2(e^{\epsilon n\tau/2}-1)}\right)$*-DP w.r.t.* $\sim_{\mathrm{hist}}$.

Both these results are proved in Appendix C. The proofs rely on the following lemma, whose proof we sketch below.

Lemma 5.1. *For any* $\nu \ge 0$ *and on inputs* $\boldsymbol{x}$ *s.t.* $|\boldsymbol{x}| \ge \frac{2}{\epsilon\tau} \ln\left(1 + \frac{1-e^{-\frac{\epsilon\tau}{2}}}{e^{\epsilon(\nu+\frac{\tau}{2})}-1}\right)$, $\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\mathrm{trLap}}$ *is* $\left((1+\nu)\epsilon, \frac{e^{\epsilon}-1}{2(e^{\epsilon q/2}-1)}\right)$*-DP w.r.t.* $\sim_{\mathrm{hist}}$.

Proof. We shall in fact prove that a mechanism which outputs $\hat{\boldsymbol{y}}$ with $\hat{\boldsymbol{y}}(i) := \boldsymbol{x}(i) + z_i$ (without rounding, and without replacing negative values with 0) is already differentially private as desired. Then, since the actual mechanism is a post-processing of this mechanism, it will also be differentially private with the same parameters.

Let $\boldsymbol{x}$ and $\boldsymbol{x}'$ be two neighbouring histograms. For simplicity, for every $i \in \mathcal{G}$, define $x_i := \boldsymbol{x}(i)$ and $x'_i := \boldsymbol{x}'(i)$. Since $\boldsymbol{x} \sim \boldsymbol{x}'$, there exists a $i^* \in \mathcal{G}$ such that $\forall i \ne i^*$, $\boldsymbol{x}(i) = \boldsymbol{x}'(i^*)$, and $|\boldsymbol{x}(i^*) - \boldsymbol{x}'(i^*)| = 1$, which implies that there exists $i^* \in \mathcal{G}$ such that $|x_{i^*} - x'_{i^*}| = 1$ and that $x_i = x'_i$ for every $i \ne \mathcal{G} \setminus \{i^*\}$. Without loss of generality, assume that $x_{i^*} = x'_{i^*} + 1$. Let $|\boldsymbol{x}| = |\boldsymbol{x}'| + 1 = n + 1$. Let $q = \tau(n+1)$ and $q' = \tau n$. For simplicity of notation, we will denote $\mathrm{supp}(\boldsymbol{y})$ by $\mathcal{G}_{\boldsymbol{y}}$ for any $\boldsymbol{y} \in \{\boldsymbol{x}, \boldsymbol{x}'\}$.

In order to prove the lemma, for every subset $S \subseteq \mathbb{H}_{\mathcal{G}}$, we need to show that

$$\Pr[\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\mathrm{trLap}}(\boldsymbol{x}) \in S] \le e^{(1+\nu)\epsilon} \Pr[\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\mathrm{trLap}}(\boldsymbol{x}') \in S] + \delta, \tag{19}$$

$$\Pr[\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\mathrm{trLap}}(\boldsymbol{x}') \in S] \le e^{(1+\nu)\epsilon} \Pr[\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\mathrm{trLap}}(\boldsymbol{x}) \in S] + \delta, \tag{20}$$

where $\delta = \frac{e^{\epsilon}-1}{2(e^{\epsilon q/2}-1)}$. We only prove (20); (19) can be shown similarly.

Fix an arbitrary subset $S \subseteq \mathbb{H}_{\mathcal{G}}$. Since $\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\mathrm{trLap}}$ adds independent noise to each bar of the histogram according to $D_q(z)$, we have that for every $\boldsymbol{s} \in \mathbb{H}_{\mathcal{G}}$, we have $\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\mathrm{trLap}}(\boldsymbol{x})(\boldsymbol{s}) = \prod_{i\in\mathcal{G}_{\boldsymbol{x}}} D_q(s_i - x_i)$ where $s_i = \boldsymbol{s}(i)$. Thus we have

$$\Pr[\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\mathrm{trLap}}(\boldsymbol{x}) \in S] = \int_S \Big[\prod_{i\in\mathcal{G}_{\boldsymbol{x}}} D_q(s_i - x_i)\Big] \,\mathrm{d}\boldsymbol{s} \tag{21}$$

$$\Pr[\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\mathrm{trLap}}(\boldsymbol{x}') \in S] = \int_S \Big[\prod_{i\in\mathcal{G}_{\boldsymbol{x}'}} D_{q'}(s_i - x'_i)\Big] \,\mathrm{d}\boldsymbol{s} \tag{22}$$

Now, using the fact that $\forall k \ne i^*, n_k = n'_k$ and $x_{i^*} = x'_{i^*} + 1$, we partition S into 3 disjoint sets:

1. $S_0 := \{\boldsymbol{s} \in \mathbb{H}_{\mathcal{G}} : s_{i^*} - x'_{i^*} < -q'\} \cup \{\boldsymbol{s} \in \mathbb{H}_{\mathcal{G}} : 0 < s_{i^*} - x'_{i^*}\}$.
2. $S_1 := \{\boldsymbol{s} \in \mathbb{H}_{\mathcal{G}} : -q' \le s_{i^*} - x'_{i^*} < -q' + (1-\tau)\}$.
3. $S_2 := \{\boldsymbol{s} \in \mathbb{H}_{\mathcal{G}} : -q' + (1-\tau) \le s_{i^*} - x'_{i^*} \le 0)\}$.

We first prove the following claim in Appendix C:

Claim 5.1. *Whenever* $\boldsymbol{s} \in S_k, k \in \{0, 2\}$, *we have* $D_{q'}(s_{i^*} - x'_{i^*}) \le e^{(1+\nu)\epsilon} D_q(s_{i^*} - x_{i^*})$, *provided* $n \ge \frac{2}{\epsilon\tau} \ln\left(1 + \frac{1-e^{-\frac{\epsilon\tau}{2}}}{e^{\epsilon(\nu+\frac{\tau}{2})}-1}\right)$.

Construction 1 (Truncated Laplace Mechanism, $\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}$)**.**

Parameters: Threshold $\tau \in [0,1)$, a ground set $\mathcal{G}$, and $\epsilon > 0$.
Input: A histogram, $\boldsymbol{x} : \mathcal{G} \to \mathbb{N}$.
Output: A histogram, $\boldsymbol{y} : \mathcal{G} \to \mathbb{N}$.

for all $g \in \mathcal{G}$ **do**
  $z_g \leftarrow D_q$, where $q := \tau|\boldsymbol{x}|$ and

$$D_q(z) = \begin{cases} \frac{1}{1-e^{-\epsilon\frac{q}{2}}}\mathrm{Lap}(z \mid -\frac{q}{2}, \frac{1}{\epsilon}) & \text{if } z \in [-q, 0], \\ 0 & \text{otherwise.} \end{cases}$$

  $\boldsymbol{y}(g) := \max(0, \lfloor \boldsymbol{x}(g) + z_g \rceil)$ ▷ z_g need not be computed for g s.t. $\boldsymbol{x}(g) = 0$
Return $\boldsymbol{y}$

Mechanism $\mathcal{M}_{\text{trLap}}^{\tau,\mathcal{G}}$ on input $\boldsymbol{x}$ returns $\mathcal{M}_{\text{trLap}}^{\tau,\frac{1}{\sqrt{\tau|\boldsymbol{x}|}},\mathcal{G}}(\boldsymbol{x})$.

Figure 3: Truncated Laplace mechanism for computing Histogram.

Note that this cannot not hold for S_1, as for $\boldsymbol{s} \in S_1$, we have $D_q(s_{i^*} - x_{i^*}) = 0$ and $D_{q'}(s_{i^*} - x'_{i^*}) > 0$, which does not satisfy the above inequality.

Now, define $S_{-1} := S \setminus S_1$. We shall prove the following two claims in Appendix C:

Claim 5.2. $\Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}') \in S_{-1}] \leq e^{(1+\nu)\epsilon} \Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}) \in S_{-1}]$

Claim 5.3. $\Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}') \in S_1] \leq \delta$, *where* $\delta = \frac{e^{\epsilon}-1}{2(e^{\epsilon q/2}-1)}$.

These together imply (19) as follows:

$$\begin{aligned} \Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}') \in S] &= \Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}') \in S_{-1}] + \Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}') \in S_1] \\ &\leq e^{(1+\nu)\epsilon} \Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}) \in S_{-1}] + \delta \\ &\leq e^{(1+\nu)\epsilon} \Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}) \in S] + \delta \end{aligned}$$

This completes the proof of Lemma 5.1. □

5.1.3 Bucketed Truncated Laplace Mechanism

For a given τ, the parameters in Theorem 5.1 improve with the input size n, as long as the support size t remains constant. To handle larger supports, this mechanism can be composed with a simple bucketing mechanism which reduces the support size to t. Here, for simplicity, we present our results for $\mathcal{G} = [0, B)^d$, but they can all be generalized to other metric spaces. Our bucketing mechanism $\mathcal{M}_{\text{buc}}^{t,[0,B)^d}$ (Construction 2) and the final bucketed-histogram mechanism $\mathcal{M}_{\text{bucHist}}^{\alpha,\beta,[0,B)^d}$ (Construction 3) is presented in Figure 4.

Note that $\mathcal{M}_{\text{bucHist}}^{\alpha,\beta,[0,B)^d} = \mathcal{M}_{\text{trLap}}^{\tau,[0,B)^d} \circ \mathcal{M}_{\text{buc}}^{t,[0,B)^d}$ is defined as a composition of two mechanism. So, in order to prove the accuracy and privacy guarantees of $\mathcal{M}_{\text{bucHist}}^{\alpha,\beta,[0,B)^d}$ using our composition theorems Theorem 4.1 and Theorem 4.2, respectively, first we need to show some properties of $\mathcal{M}_{\text{buc}}^{t,[0,B)^d}$ and $\mathcal{M}_{\text{trLap}}^{\tau,[0,B)^d}$, which we show below. Claim 5.4 proves that $\mathcal{M}_{\text{buc}}^{t,[0,B)^d}$ is a neighborhood-preserving mechanism and Claim 5.5 establishes its guarantees for the identity function; Claim 5.6 and Claim 5.7 bound the distortion sensitivity of the

Construction 2 (Bucketing Mechanism, $\mathcal{M}_{\text{buc}}^{t,[0,B)^d}$)**.**

Parameters: Desired number of buckets t, range of inputs $[0, B)^d$.
Input: A histogram $\boldsymbol{x}$ with elements in $[0, B)^d$.
Output: A histogram $\boldsymbol{y}$ with elements in $\{\frac{B}{t}(i - \frac{1}{2}) : i \in [t]\}$ with $|\boldsymbol{y}| = |\boldsymbol{x}|$.

1: **for all** $i = (i_1, \ldots, i_d) \in [t^{\frac{1}{d}}]^d$ **do**
2: $\quad \boldsymbol{y}\left(\frac{B(i_1-\frac{1}{2})}{t^{\frac{1}{d}}}, \ldots, \frac{B(i_d-\frac{1}{2})}{t^{\frac{1}{d}}}\right) := \sum_{(g_1,\ldots,g_d):g_j\in\left[\frac{B(i_j-1)}{t^{\frac{1}{d}}}, \frac{Bi_j}{t^{\frac{1}{d}}}\right), \forall j\in[d]} \boldsymbol{x}(g)$
3: Return $\boldsymbol{y}$ $\qquad\triangleright$ $\boldsymbol{y}(v) = 0$ by default

---

Construction 3 (BucketHist, $\mathcal{M}_{\text{bucHist}}^{\alpha,\beta,[0,B)^d}$)**.**

Parameters: Accuracy parameters α, β; range of inputs $[0, B)^d$.
Input: A histogram $\boldsymbol{x}$ with elements in $[0, B)^d$.
Output: A histogram over $[0, B)^d$.

1: $t := \lceil (\frac{B}{2\beta})^d \rceil$, $\tau := \alpha/t$
2: Return $\mathcal{M}_{\text{trLap}}^{\tau,[0,B)^d} \circ \mathcal{M}_{\text{buc}}^{t,[0,B)^d}(\boldsymbol{x})$

Figure 4: Mechanisms for computing Histogram.

histogram function and the error sensitivity of the truncated Laplace mechanism, respectively. All these claims are proved in Appendix C.

Claim 5.4. *$\mathcal{M}_{\text{buc}}^{t,[0,B)^d}$ is a neighborhood-preserving mechanism, i.e., for any two histograms $\boldsymbol{x}, \boldsymbol{x}'$ such that $\boldsymbol{x} \sim_{\text{hist}} \boldsymbol{x}'$, we have $\mathcal{M}_{\text{buc}}^{t,[0,B)^d}(\boldsymbol{x}) \sim_{\text{hist}} \mathcal{M}_{\text{buc}}^{t,[0,B)^d}(\boldsymbol{x}')$.*

Since $\mathcal{M}_{\text{buc}}^{t,[0,B)^d}$ introduces error in the output space, we need a metric over $\mathbb{H}_{[0,B)^d}$ to analyze its flexible accuracy. We use the following natural metic $\mathfrak{d}_{\text{hist}}$ over $\mathbb{H}_{[0,B)^d}$, which is defined as $\mathfrak{d}_{\text{hist}}(\boldsymbol{y}, \boldsymbol{y}') := W^\infty(\frac{\boldsymbol{y}}{|\boldsymbol{y}|}, \frac{\boldsymbol{y}'}{|\boldsymbol{y}'|})$. Here, $\frac{\boldsymbol{y}}{|\boldsymbol{y}|}$ is treated as a probability distribution.

Claim 5.5. *On all inputs $\boldsymbol{x}$, $\mathcal{M}_{\text{buc}}^{t,[0,B)^d}$ is $\left(0, \frac{B\sqrt{d}}{2t^{\frac{1}{d}}}, 0\right)$-accurate for the identity function w.r.t the metric $\mathfrak{d}_{\text{hist}}$ and any measure of distortion.*

Claim 5.6. *The distortion sensitivity of the histogram identity function w.r.t. $(\partial_{\text{drop}}, \partial_{\text{drop}})$ at $\theta = \omega = 0$ is the identity function.*

Claim 5.7. *For any $\beta \in \mathbb{R}$ and on inputs restricted to t bars, we have $\tau^{\alpha,0}_{\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}, f_{id}}(\beta, 0) \leq \beta$ w.r.t. ∂_{drop} where $\alpha = \tau t$ and f_{id} is the identity function.*

Now we ready to prove the accuracy and privacy guarantees of the $\mathcal{M}_{\text{bucHist}}^{\alpha,\beta,[0,B)^d}$ mechanism.

Theorem 5.3. *On histograms of size at least n, $\mathcal{M}_{\text{bucHist}}^{\alpha,\beta,[0,B)^d}$ is $\left(O(\frac{1}{\sqrt{\sigma n}}), e^{-\Omega(\sqrt{\sigma n})}\right)$-DP w.r.t. $\sim_{\text{hist}}$, where $\sigma = \alpha(\frac{2\beta}{B})^d$, and $(\alpha, \beta, 0)$-accurate for the identity function, w.r.t. the distortion measure ∂_{drop} and metric $\mathfrak{d}_{\text{hist}}$.*

Proof. We have $\mathcal{M}_{\mathrm{bucHist}}^{\alpha,\beta,[0,B)^d} = \mathcal{M}_{\mathrm{trLap}}^{\tau,[0,B)^d} \circ \mathcal{M}_{\mathrm{buc}}^{t,[0,B)^d}$, with $t = \lceil (\frac{B}{2\beta})^d \rceil, \tau = \frac{\alpha}{t}$. Let $\sigma = \tau = \alpha(\frac{2\beta}{B})^d$, and let f_{id} (which is the identity function) denote the underlying function that $\mathcal{M}_{\mathrm{buc}}^{t,[0,B)^d}$ computes. Note that $\mathcal{M}_{\mathrm{bucHist}}^{\alpha,\beta,[0,B)^d}$, $\mathcal{M}_{\mathrm{trLap}}^{\tau,[0,B)^d}$, and $\mathcal{M}_{\mathrm{buc}}^{t,[0,B)^d}$ compute the same underlying function f_{id}. First we prove the privacy guarantee and the show the flexible accuracy guarantee.

- **Differential privacy.** We have from Claim 5.4 that $\mathcal{M}_{\mathrm{buc}}^{t,[0,B)^d}$ is a neighborhood-preserving mechanism w.r.t. the neighborhood relation $\sim_{\mathrm{hist}}$. Note that $\mathcal{M}_{\mathrm{buc}}^{t,[0,B)^d}$ outputs a histogram whose support size is at most t. We also have from Theorem 5.1 that $\mathcal{M}_{\mathrm{trLap}}^{\tau,[0,B)^d}$ on input histograms with support size at most t is differentially-private with the required parameters. Now, it follows from Theorem 4.2 that $\mathcal{M}_{\mathrm{bucHist}}^{\alpha,\beta,[0,B)^d}$ is $\left(O(\frac{1}{\sqrt{\sigma n}}), e^{-\Omega(\sqrt{\sigma n})}\right)$-differentially private w.r.t. $\sim_{\mathrm{hist}}$.
- **Flexible accuracy.** We have from Claim 5.5 that $\mathcal{M}_{\mathrm{buc}}^{t,[0,B)^d}$ is $(0, \beta, 0)$-accurate. We have from Claim 5.6 that $\sigma_{f_{id}}(\alpha) = \alpha$ for all $\alpha \geq 0$, i.e., f_{id} has identity θ-distortion sensitivity for $\theta = 0$. (We are taking $\omega = \theta = 0$ in the definition of $\sigma_{f_{id}}^{\theta,\omega}$ in Definition 11.)

  Note that since $\mathcal{M}_{\mathrm{trLap}}^{\tau,[0,B)^d}$ operates on input histograms with support size at most t, we have from Theorem 5.1 that $\mathcal{M}_{\mathrm{trLap}}^{\tau,[0,B)^d}$ is $(\tau t, 0, 0)$-accurate w.r.t. the distortion measure ∂_{drop} and metric $\mathfrak{d}_{\mathrm{hist}}(,)$, and $(O(\frac{1}{\sqrt{\sigma n}}), e^{-\Omega(\sqrt{\sigma n})})$-differentially private.

  By Claim 5.7, we have, $\tau^{\alpha,0}_{\mathcal{M}_{\mathrm{trLap}}^{\tau,[0,B)^d}, f_{id}}(\beta, 0) \leq \beta$. Also note that, since $\theta = \omega = 0$, we have that $\tau^{0,\theta}_{\mathcal{M}_{\mathrm{trLap}}^{\tau,\epsilon,[0,B)^d}}(\omega, \theta) = 0$.

  Now, applying Theorem 4.1 to $\mathcal{M}_{\mathrm{bucHist}}^{\alpha,\beta,[0,B)^d} = \mathcal{M}_{\mathrm{trLap}}^{\tau,[0,B)^d} \circ \mathcal{M}_{\mathrm{buc}}^{t,[0,B)^d}$, we have that $\mathcal{M}_{\mathrm{bucHist}}^{\alpha,\beta,[0,B)^d}$ is $(\alpha, \beta, 0)$-accurate.

This completes the proof of Theorem 5.3. □

5.1.4 Histogram-Based-Statistics

Theorem 5.3 provides a powerful tool to obtain a DP mechanism for *any deterministic* histogram-based-statistic $f_{\mathrm{HBS}} : \mathbb{H}_{\mathcal{G}} \to \mathcal{A}$, simply by composing $\mathcal{M}_{f_{\mathrm{HBS}}}^{\alpha,\beta,[0,B)^d} = f_{\mathrm{HBS}} \circ \mathcal{M}_{\mathrm{bucHist}}^{\alpha,\beta,[0,B)^d}$, as described in Construction 4 in Figure 5. To analyze the flexible accuracy of $\mathcal{M}_{f_{\mathrm{HBS}}}$, we define the *metric sensitivity* function of f_{HBS}:

Construction 4 (Histogram-Based-Statistics (HBS) Mechanism, $\mathcal{M}_{f_{\mathrm{HBS}}}^{\alpha,\beta,[0,B)^d}$).

Parameters: Accuracy parameters α, β; data range $[0, B)^d$.
Input: A histogram $\boldsymbol{x}$ with elements in $[0, B)^d$.
Output: Output of f_{HBS} on $\boldsymbol{x}$.

1: Return $f_{\mathrm{HBS}} \circ \mathcal{M}_{\mathrm{bucHist}}^{\alpha,\beta,[0,B)^d}(\boldsymbol{x})$

Figure 5: Mechanism for computing a deterministic histogram-based-statistic HBS defined over histograms.

Definition 14. The *metric sensitivity* of a histogram-based-statistic $f_{\mathrm{HBS}} : \mathbb{H}_{[0,B)^d} \to \mathcal{A}$, is given by $\Delta_{f_{\mathrm{HBS}}} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, in terms of a metric $\mathfrak{d}_{\mathcal{A}}$ over $\mathcal{A}$,

$$\Delta_{f_{\mathrm{HBS}}}(\beta) = \sup_{\boldsymbol{x},\boldsymbol{x}' \in \mathbb{H}_{[0,B)^d} : \mathfrak{d}_{\mathrm{hist}}(\boldsymbol{x},\boldsymbol{x}') \leq \beta} \mathfrak{d}_{\mathcal{A}}(f_{\mathrm{HBS}}(\boldsymbol{x}), f_{\mathrm{HBS}}(\boldsymbol{x}')). \tag{23}$$

We need the following lemma to establish the accuracy guarantees of our HBS mechanism in the proof of Theorem 5.4 below.

Lemma 5.2. *Let $f : \mathcal{A} \to \mathcal{B}$ be a deterministic function with $\mathfrak{d}_{\mathcal{A}}, \mathfrak{d}_{\mathcal{B}}$ metrics defined in $\mathcal{A}, \mathcal{B}$, respectively. Then we have,*

$$\sup_{\substack{X,X':\\ W^\infty(X,X')\leq\beta}} W^\infty(f(X), f(X')) = \sup_{\substack{\boldsymbol{y},\boldsymbol{y}'\in\mathcal{A}:\\ \mathfrak{d}_{\mathcal{A}}(\boldsymbol{y},\boldsymbol{y}')\leq\beta}} \mathfrak{d}_{\mathcal{B}}(f(\boldsymbol{y}), f(\boldsymbol{y}'))$$

Theorem 5.4. *On inputs of size at least n, $\mathcal{M}_{f_{\mathrm{HBS}}}^{\alpha,\beta,[0,B)^d}$ is $\left(O(\frac{1}{\sqrt{\sigma n}}), e^{-\Omega(\sqrt{\sigma n})}\right)$-DP, where $\sigma = \alpha(\frac{2\beta}{B})^d$, and $(\alpha, \Delta_{f_{\mathrm{HBS}}}(\beta), 0)$-accurate for f_{HBS} w.r.t. distortion ∂_{drop} and metric $\mathfrak{d}_{\mathcal{A}}$.*

Proof. Since $\mathcal{M}_{f_{\mathrm{HBS}}}^{\alpha,\beta,[0,B)^d} = f_{\mathrm{HBS}} \circ \mathcal{M}_{\mathrm{bucHist}}^{\alpha,\beta,[0,B)^d}$ is a composed mechanism, we will use our composition theorems to establish its accuracy and privacy guarantees.

We have from Theorem 5.3 that $\mathcal{M}_{\mathrm{bucHist}}^{\alpha,\beta,[0,B)^d}$ is $(O(\frac{1}{\sqrt{\sigma n}}), e^{-\Omega(\sqrt{\sigma n})})$-DP, where $\sigma = \alpha(\frac{2\beta}{B})^d$, and $(\alpha, \beta, 0)$-accurate w.r.t. the distortion measure ∂_{drop} and the metric $\mathfrak{d}_{\mathrm{hist}}(\cdot,\cdot)$. Trivially, by definition, f_{HBS} is $(0, 0, 0)$-accurate for computing f_{HBS}.

Now we compute the parameters required for composition. Since $\alpha_2 = 0$ for composition, we trivially have $\sigma_{f_{id}}^0(0) = 0$, where f_{id} is the identity function. (We are taking $\omega = \theta = 0$ in the definition of $\sigma_{f_1}^{\theta,\omega}$ in Definition 11.)

The error-sensitivity of f_{HBS} at required parameters is $\tau_{f_{\mathrm{HBS}}}^{0,0}(\beta, 0)$. We have,

$$\begin{aligned}
\tau_{f_{\mathrm{HBS}}}^{0,0}(\beta, 0) &= \sup_{\substack{X,X':\\ W^\infty(X,X')\leq\beta}} \inf_{\substack{Y:\\ \partial(X',Y)\leq 0}} W^\infty(f_{\mathrm{HBS}}(X), f_{\mathrm{HBS}}(Y)) \\
&= \sup_{\substack{X,X':\\ W^\infty(X,X')\leq\beta}} W^\infty(f_{\mathrm{HBS}}(X), f_{\mathrm{HBS}}(X')) \\
&\leq \sup_{\substack{\boldsymbol{x},\boldsymbol{x}'\in\mathbb{H}_{\mathcal{G}}:\\ \mathfrak{d}_{\mathrm{hist}}(\boldsymbol{x},\boldsymbol{x}')\leq\beta}} \mathfrak{d}_{\mathcal{A}}(f_{\mathrm{HBS}}(\boldsymbol{x}), f_{\mathrm{HBS}}(\boldsymbol{x}')) && \text{(Using Lemma 5.2)} \\
&= \Delta_{f_{\mathrm{HBS}}}(\beta)
\end{aligned}$$

Thus we have $\tau_{f_{\mathrm{HBS}}}^{0,0}(\beta, 0) \leq \Delta_{f_{\mathrm{HBS}}}(\beta)$. Also note that, since $\theta = \omega = 0$, we have that $\tau_{\mathcal{M}_{\mathrm{trLap}}^{\tau,\epsilon,\mathcal{G}}}^{0,\theta}(\omega, \theta) = 0$.

Hence, by Theorem 4.1, the composed mechanism $\mathcal{M}_{f_{\mathrm{HBS}}}^{\alpha,\beta,B}$ is $(\alpha, \Delta_{f_{\mathrm{HBS}}}(\beta), 0)$-accurate. For the privacy guarantee, note that this composition can be viewed as a post processing over the histogram mechanism, thus $\mathcal{M}_{f_{\mathrm{HBS}}}^{\alpha,\beta,B}$ is also a $(O(\frac{1}{\sqrt{\sigma n}}), e^{-\Omega(\sqrt{\sigma n})})$-differentially private mechanism for f_{HBS}, where $\sigma = \alpha(\frac{2\beta}{B})^d$. □

Examples of New Applications. Theorem 5.4 has direct applications to functions which have high sensitivity (defined w.r.t. the neighborhood relation $\sim$), but low metric sensitivity. As examples of functions for which no previous solutions offered non-trivial guarantees, consider $f_{\max}$ defined as $f_{\max}(\boldsymbol{x}) := \max\{g : g \in \mathrm{supp}(\boldsymbol{x})\}$, and f_{supp} defined to be the same as supp. For $f_{\max}$ the metric over its range is the absolute difference metric over $\mathbb{R}$; for f_{supp}, we use a metric $\mathfrak{d}_{\mathrm{supp}}$ over the set of (finite) subsets of $\mathbb{R}$, defined by $\mathfrak{d}_{\mathrm{supp}}(X, Y) := \max\left\{\max_{x\in X} \min_{y\in Y} |x - y|, \max_{y\in Y} \min_{x\in X} |x - y|\right\}$. ($\mathfrak{d}_{\mathrm{supp}}$ measures the farthest that a point in one of the sets is from any point on the other set.) Hence, when f_{HBS} is either of these functions, we state the results in Corollary 5.1 and Corollary 5.2, respectively. Using the results, we can see that, e.g., for any (small) constant $0 < \alpha < 1$, $\mathcal{M}_{f_{\mathrm{HBS}}}^{\alpha,\alpha B,B}$ is a $(O(\frac{1}{\sqrt{n}}), e^{-\Omega(\sqrt{n})})$-DP mechanism with $(\alpha, \alpha B, 0)$ accuracy w.r.t. ∂_{drop} and the metric from above.

Corollary 5.1. *On inputs of size at least n, $\mathcal{M}_{\max}^{\alpha,\beta,[0,B)^d}$ is $\left(O(\frac{1}{\sqrt{\sigma n}}), e^{-\Omega(\sqrt{\sigma n})}\right)$-DP, where $\sigma = \alpha(\frac{2\beta}{B})^d$, and $(\alpha, \beta, 0)$-accurate w.r.t. the distortion measure ∂_{drop} and the standard metric for $\mathbb{R}$, for the function $f_{\max}$.*

Proof. For any two histograms $\boldsymbol{y}$, $\boldsymbol{y}'$, by definition of $\mathfrak{d}_{\text{hist}}(,)$ and $f_{\max}$, we have $|f_{\max}(\boldsymbol{y}) - f_{\max}(\boldsymbol{y}')| \leq \mathfrak{d}_{\text{hist}}(\boldsymbol{y}, \boldsymbol{y}')$. Using this in (23) implies that $\Delta_{f_{\max}}(\beta) = \beta$ for every $\beta \in \mathbb{R}$, i.e., $\Delta_{f_{\max}}$ is upper-bounded by the identity function. Then, the corollary follows from Theorem 5.4, with $f_{\text{HBS}} = f_{\max}$. ☐

Remark 5.1. *By taking $\sigma = \frac{1}{n\epsilon^2}$ and $\frac{\beta}{B}$ and d to be constants (which means that the output error β is a constant fraction of the entire range B), we have that $\mathcal{M}_{\max}^{\alpha,\beta,[0,B)^d}$ is an $(\epsilon, e^{-\Omega(\frac{1}{2\epsilon})})$-differentially private mechanism that works with a input distortion of $\alpha = O(\frac{1}{n\epsilon^2})$, i.e., it drops $O(\frac{1}{\epsilon^2})$ data points.*

Corollary 5.2. *On inputs of size at least n, $\mathcal{M}_{\text{supp}}^{\alpha,\beta,[0,B)^d}$ is $\left(O(\frac{1}{\sqrt{\sigma n}}), e^{-\Omega(\sqrt{\sigma n})}\right)$-DP, where $\sigma = \alpha(\frac{2\beta}{B})^d$, and $(\alpha, \beta, 0)$-accurate w.r.t. the distortion measure ∂_{drop} and metric $\mathfrak{d}_{\text{supp}}(\cdot,\cdot)$ for the function f_{supp}.*

Proof. For any two histograms $\boldsymbol{y}$, $\boldsymbol{y}'$, by definition of $\mathfrak{d}_{\text{hist}}(,)$ and f_{supp}, we have, $\mathfrak{d}_{\text{supp}}(f_{\text{supp}}(\boldsymbol{y}), f_{\text{supp}}(\boldsymbol{y}')) \leq \mathfrak{d}_{\text{hist}}(\boldsymbol{y}, \boldsymbol{y}')$. Using this in (23) implies that $\Delta_{f_{\text{supp}}}(\beta) = \beta$ for every $\beta \in \mathbb{R}$, i.e., $\Delta_{f_{\text{supp}}}$ is upper-bounded by the identity function. Then, the corollary follows from Theorem 5.4, with $f_{\text{HBS}} = f_{\text{supp}}$. ☐

5.1.5 Further Applications: Beyond ∂_{drop}

Useful variants of Theorem 5.4 can be obtained with measures of distortion other than ∂_{drop}. In particular, we define ∂_{mv} (which is a metric, see Claim E.1) to allow moving the data points, and $\partial_{\text{drmv}}^{\eta}$ (which is a quasi-metric, see Claim E.3) to allow both dropping and moving, as follows:

$$\partial_{\text{mv}}(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} W^{\infty}(\frac{\boldsymbol{x}}{|\boldsymbol{x}|}, \frac{\boldsymbol{y}}{|\boldsymbol{y}|}) & \text{if } |\boldsymbol{x}| = |\boldsymbol{y}| \\ \infty & \text{otherwise} \end{cases} \qquad \partial_{\text{drmv}}^{\eta}(\boldsymbol{x}, \boldsymbol{y}) = \inf_{\boldsymbol{z}} \left(\partial_{\text{drop}}(\boldsymbol{x}, \boldsymbol{z}) + \eta \cdot \partial_{\text{mv}}(\boldsymbol{z}, \boldsymbol{y})\right).$$

The following theorem provides the guarantees obtained by using $\partial_{\text{drmv}}^{\eta}$ as the distortion measure in any deterministic histogram-based-statistic, and we prove it in Appendix E.

Theorem 5.5. *For any deterministic histogram-based-statistic $f_{\text{HBS}} : \mathbb{H}_{[0,B)^d} \to \mathcal{A}$, there exists a mechanism which, on inputs of size at least n, is $\left(O(\frac{1}{\sqrt{\sigma n}}), e^{-\Omega(\sqrt{\sigma n})}\right)$-DP, where $\sigma = \alpha \left(\frac{2\beta}{B}\right)^d$, and $(\alpha + \eta\beta, 0, 0)$-accurate for the identity function, w.r.t. the distortion measure $\partial_{\text{drmv}}^{\eta}$.*

This is analogous to Theorem 5.4, but with the important difference that it does not refer to the metric sensitivity of the function f_{HBS}, and does not even require a metric over its codomain $\mathcal{A}$. This makes this result applicable to complex function families like maximum-margin separators or neural net classifiers. However, the accuracy notion uses a measure of distortion that allows dropping a (small) fraction of the data and (slightly) moving all data points, which may or may not be acceptable to all applications.

5.1.6 Usefulness [BLR13] vs. Flexible Accuracy

To express accuracy guarantees of their mechanisms, Blum et al. [BLR13] introduced a notion of (β, γ, ψ)-*usefulness* that parallels (α, β, γ)-accuracy, except that ψ measures perturbation of the function rather than input distortion. Indeed, mechanisms which are $(\beta, \gamma, 0)$-useful are $(0, \beta, \gamma)$-accurate (in [BLR13], such mechanisms were given for interval queries). But even general usefulness can be translated to flexible accuracy generically, by redefining the function to have an extra input parameter that specifies perturbation. Further, the specific (β, γ, ψ)-useful DP mechanism of [BLR13] for half-space counting queries – with data points on a unit sphere, and the perturbation of the function corresponded to rotating the half-space by ψ radians – is (ψ, β, γ)-accurate for the same functions, w.r.t. the distortion ∂_{mv}. This is because, the rotation of the half-space can be modeled as moving all the points on the unit sphere by a distance of at most ψ.

5.2 Mechanism for Robust Privacy

In this section, first we give a compiler that transforms any mechanism (whose range is real numbers) into a *robustly private* mechanism that preserves its differential privacy and does not degrade the accuracy by much. For simplicity, we consider the output space of the mechanisms to be $\mathbb{R}$.

Theorem 5.6 (Robust privacy compiler). *Suppose a mechanism $\mathcal{M} : A \to \mathbb{R}$ is (ϵ, δ)-DP and (α, β, γ)-accurate for some function $f : A \to \mathbb{R}$, w.r.t. distortion ∂ and metric $\mathfrak{d}$. Then, for any $\rho \geq 0, \theta \in [0, 1], \hat{\epsilon} > \ln 2$ and $\gamma' \in (\gamma, 1]$, there exists an (ϵ, δ)-DP mechanism $\mathcal{M}'$ for f with $(\rho, \theta, \hat{\epsilon}, \hat{\delta})$-robust privacy w.r.t. $\mathfrak{d}$, and $(\alpha, \beta + \beta', \gamma')$-accuracy w.r.t. distortion ∂ and metric $\mathfrak{d}$, where $\hat{\delta} = \theta(1 + \frac{\hat{\epsilon}}{2})$ and $\beta' = \frac{2\rho}{\hat{\epsilon} - \ln 2} \ln \frac{1}{(\gamma' - \gamma)}$.*

In order to prove Theorem 5.6, we first give a construction of a robust mechanism for the identity function and state its guarantees in Theorem 5.7. We then give the proof of Theorem 5.6.

5.2.1 A Robustly Private Mechanism for the Identity Function over $\mathbb{R}$

We now show that the Laplace mechanism stated in Section 2.1.1 for the identity function is also robust and flexibly accurate. Specifically, we define $\mathcal{M}^{\mathrm{id},b}_{\mathrm{Lap}} : \mathbb{R} \to \mathbb{R}$ as follows: for any input $y \in \mathbb{R}$, output $\mathcal{M}^{\mathrm{id},b}_{\mathrm{Lap}}(y) = y + \mathrm{Lap}(b)$.

Theorem 5.7. *For any $\theta, \gamma \in [0, 1]$ and $\rho \geq 0$, $\mathcal{M}^{\mathrm{id},b}_{\mathrm{Lap}}$ is $(\theta, \rho, \epsilon, \delta)$-robust and $(0, \beta, \gamma)$-accurate w.r.t. the absolute difference metric in $\mathbb{R}$, and for all $\beta_1 \geq 0, \gamma, \gamma_1 > 0$, has error sensitivity upper bounded as, $\tau^{0,\gamma+\gamma_1}_{\mathcal{M}^b_{\mathrm{Lap}}}(\beta_1, \gamma_1) \leq \beta_1 + \beta$, where*

$$\epsilon = \frac{2\rho}{b} + \ln(2); \qquad \delta = \frac{\rho\theta}{b(1 - e^{-\frac{\rho}{b}})}; \qquad \beta = b \ln(1/\gamma).$$

Theorem 5.7 follows from Lemma 5.3, Lemma 5.4, and Lemma 5.5 below.

Lemma 5.3. *For any $\theta \in [0, 1]$ and $\rho \geq 0$, $\mathcal{M}^{\mathrm{id},b}_{\mathrm{Lap}}$ achieves $\left(\theta, \rho, 2\frac{\rho}{b} + \ln(2), \frac{\rho\theta}{b(1-e^{-\frac{\rho}{b}})}\right)$-robustness.*

Proof. For convenience, we prove this lemma for discrete input distributions only; the arguments can be readily extended to continuous distributions.

Fix a $\theta \in [0, 1]$. Suppose Q and Q' denote two input distributions defined over a discrete subset $\mathcal{U} \subseteq \mathbb{R}$, taking the value $u \in \mathcal{U}$ with probability $Q(u)$ and $Q'(u)$ respectively. Let P and P' denote the corresponding output distributions of $\mathcal{M}^b_{\mathrm{Lap}}$. Also, let $W^\infty_\theta(P, P') \leq \rho$.

Note that $P(x) = \sum_{u \in \mathcal{U}} Q(u)\mathrm{Lap}(x|u, b)$. That is, P is a convex combination of Laplace distributions. The same holds for P'. Claim 5.8 below shows that such distributions are log-Lipschitz, and then Claim 5.9 establishes the robustness claimed in Lemma 5.3.

Claim 5.8. *Suppose P is a distribution that is a convex combination of Laplace distributions $\{\mathrm{Lap}(u, b) : u \in \mathcal{U}\}$ for some (discrete) set $\mathcal{U} \subseteq \mathbb{R}$. Then, for every $d \in \mathbb{R}$, we have $\frac{P(x)}{P(x+d)} \in [e^{\frac{-|d|}{b}}, e^{\frac{|d|}{b}}]$. Equivalently, $\ln(P)$ is a $\frac{1}{b}$-Lipschitz function.*

Claim 5.9. *Suppose two distributions P, Q (defined over the same alphabet) both satisfy the log-Lipschitz condition given in Claim 5.8, and $W^\infty_\theta(P, Q) \leq \rho$. Then P, Q satisfy the (ϵ, δ)-DP condition (i.e., for every $\mathcal{S} \subseteq \mathbb{R}$, we have $\Pr_{x \leftarrow P}[x \in \mathcal{S}] \leq e^\epsilon \Pr_{x \leftarrow Q}[x \in \mathcal{S}] + \delta$), where $\epsilon = \frac{2\rho}{b} + \ln(2)$ and $\delta = \frac{\rho\theta}{b(1-e^{-\frac{\rho}{b}})}$.*

Claim 5.8 and Claim 5.9 are proved in Appendix F.1. ◻

Lemma 5.4. *For every $\gamma \in [0, 1]$, $\mathcal{M}^{\mathrm{id},b}_{\mathrm{Lap}}$ is $(0, \beta, \gamma)$-accurate, where $\beta = b \ln(\frac{1}{\gamma})$.*

Lemma 5.5. *The error sensitivity of* $\mathcal{M}^{\mathrm{id},b}_{\mathrm{Lap}}$ *has the following upper-bound for all* $\beta_1, \gamma > \gamma_1$*:*

$$\tau^{0,\gamma}_{\mathcal{M}^b_{\mathrm{Lap}}}(\beta_1,\gamma_1) \le \beta_1 + b\ln\left(\frac{1}{\gamma-\gamma_1}\right).$$

Lemma 5.4 and Lemma 5.5 are proved in Appendix F.2.

We now give the proof of Theorem 5.6.

Proof of Theorem 5.6. The mechanism $\mathcal{M}'$ is given by $\mathcal{M}^{\mathrm{id},b}_{\mathrm{Lap}} \circ \mathcal{M}$. We use our composition theorems from Section 4 and some simplifications to calculate the parameters of $\mathcal{M}'$. Note that all the parameters of $\mathcal{M}^{\mathrm{id},b}_{\mathrm{Lap}}$ required for our composition theorem are given by Theorem 5.7. The privacy and accuracy guarantees of $\mathcal{M}'$ given by the composition theorem are as follows:

- **$\mathcal{M}'$ is (ϵ,δ)-DP:** Since $\mathcal{M}$ is (ϵ,δ)-DP and post-processing preserves differential privacy [DR14, Proposition 2.1], $\mathcal{M}^{\mathrm{id},b}_{\mathrm{Lap}} \circ \mathcal{M}$ is also (ϵ,δ)-DP.
- **$\mathcal{M}'$ is $(\alpha, \beta+\beta', \gamma')$-accurate, where $\beta' = b\ln(\frac{1}{\gamma'-\gamma})$:** We get the result by applying Theorem 4.1 to $\mathcal{M}^{\mathrm{id},b}_{\mathrm{Lap}} \circ \mathcal{M}$, substituting the parameters of $\mathcal{M}^{\mathrm{id},b}_{\mathrm{Lap}}$ and $\mathcal{M}$ and using $\sigma_f(0) = 0$.
- **$\mathcal{M}'$ has $(\rho,\theta,\hat{\epsilon},\hat{\delta})$-robust privacy, where $\hat{\epsilon} = 2\frac{\rho}{b} + \ln(2)$ and $\hat{\delta} = \theta\frac{\rho/b}{1-e^{-\rho/b}}$:** Since $\mathcal{M}^{\mathrm{id},b}_{\mathrm{Lap}}$ is $(\rho,\theta,\hat{\epsilon},\hat{\delta})$-robustly private and pre-processing preserves robust privacy (Theorem 4.3), $\mathcal{M}^{\mathrm{id},b}_{\mathrm{Lap}} \circ \mathcal{M}$ is also $(\rho,\theta,\hat{\epsilon},\hat{\delta})$-robustly private.

We can simplify the expression for δ by upper-bounding it using a better expression. Since $x+1 \ge \frac{x}{1-e^{-x}}$ holds for all $x > 0$,[2] we can write $\hat{\delta} = \theta\frac{\rho/b}{1-e^{-\rho/b}} \le \theta(\frac{\rho}{b}+1)$. Note that while we only prove this for $\frac{\rho}{b} > 0$, this holds for $\frac{\rho}{b} = 0$ too.

Now, we put the value of b in terms of $\hat{\epsilon}$. This gives us the following bound on $\hat{\delta}$:

$$\delta \le \theta(1+\frac{\rho}{b}) = \theta(1+\frac{\hat{\epsilon}-\ln(2)}{2}) \le \theta(1-\frac{\ln(2)}{2}+\frac{\hat{\epsilon}}{2}) \le \theta(1+\frac{\hat{\epsilon}}{2})$$

which is the required value.

Putting the value of b in β' gives $\beta' = \frac{2\rho}{\hat{\epsilon}-\ln 2}\ln\frac{1}{(\gamma'-\gamma)}$, which is the required value. This completes the proof of Theorem 5.6. ☐

5.3 Mechanism for Private Sampling

Now we introduce differentially private sampling. For a given randomized function $f : \mathcal{X} \to \mathcal{Y}$, where, for each $\boldsymbol{x} \in \mathcal{X}$, $f(\boldsymbol{x})$ is a probability distribution over $\mathcal{Y}$, we would like to design a DP mechanism $\mathcal{M}$ s.t., for every $\boldsymbol{x} \in \mathcal{X}$, $\mathcal{M}(\boldsymbol{x})$ is close to the distribution $f(\boldsymbol{x})$. For this we extend the definition of (neighborhood) sensitivity to sampling queries.

Definition 15 (Parameterized Sensitivity of a Randomized Query)**.** For $\theta \in [0,1]$, we define θ-sensitivity of a randomized function f, denoted by $S^\theta(f)$, as:

$$S^\theta(f) := \max_{\substack{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}:\\ \boldsymbol{x}\sim\boldsymbol{x}'}} W^\infty_\theta(f(\boldsymbol{x}), f(\boldsymbol{x}')). \tag{24}$$

Now we give a natural extension of the Laplace mechanism given in Section 2.1.1 that can be applied to obtain differential privacy for a randomized queries. For a randomized query $f : \mathcal{X} \to \mathbb{R}$, we define $\mathcal{M}^{f,b}_{\mathrm{Lap}} : \mathcal{X} \to \mathbb{R}$ as follows: for an input $\boldsymbol{x} \in \mathcal{X}$, sample $y \sim f(\boldsymbol{x})$ and $z \sim \mathrm{Lap}(b)$, and output $y + z$.

[2] Note that $e^x \ge x+1$ holds for every $x > 0$. Now $x+1 \ge \frac{x}{1-e^{-x}}$ holds from the following sequence of equivalent expressions: $e^x \ge x+1 \Leftrightarrow e^{-x} \le \frac{1}{x+1} \Leftrightarrow 1-e^{-x} \ge \frac{x}{x+1} \Leftrightarrow x+1 \ge \frac{x}{1-e^{-x}}$

To make the notation different from that of deterministic functions, for convenience, we denote the above Laplace mechanism for a randomized query f by $\mathcal{M}_{\text{rand}}^{f,b}$, which, for any input $\boldsymbol{x}$, first samples $y \sim f(\boldsymbol{x})$ and $z \sim \text{Lap}(b)$, and then outputs $y + z$.

The following theorem establishes the robustness, privacy, and accuracy guarantees of $\mathcal{M}_{\text{rand}}^{f,b}$.

Theorem 5.8. *For a randomized query f, $\mathcal{M}_{\text{rand}}^{f,b}$ is*

1. $\forall \theta' \in [0,1], \rho \geq 0$, *$(\theta', \rho, \hat{\epsilon}, \hat{\delta})$-robust where $\hat{\epsilon} = 2\frac{\rho}{b} + \ln(2), \hat{\delta} = \theta'(1 + \frac{\rho}{b})$,*
2. *(ϵ, δ)-differentially private for every $\delta \geq 0$, where $\epsilon = \frac{S^{\delta}(f)}{b}$, and*
3. *$(0, \beta, \gamma)$-accurate where $\beta = b \ln(\frac{1}{\gamma})$.*

Proof. Observe that f can be treated as a mechanism for f with $(0,0,0)$-accuracy and $(\infty, 0)$-privacy and also that $\mathcal{M}_{\text{rand}}^{f,b} := \mathcal{M}_{\text{Lap}}^{\text{id},b} \circ f$. The robustness and accuracy bounds are obtained from Theorem 5.6. To show that $\mathcal{M}_{\text{rand}}^{f,b}$ is (ϵ, δ)-differentially private, consider any two neighbouring databases $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^n$. Since $\boldsymbol{x} \sim \boldsymbol{x}'$, it follows from (24) in Definition 15 that $W_{\theta}^{\infty}(f(\boldsymbol{x}), f(\boldsymbol{x}')) \leq S^{\theta}(f)$. Let P and Q denote $f(\boldsymbol{x}), f(\boldsymbol{x}')$ respectively. By Corollary A.1, there is a P' such that $\Delta(P, P') \leq \theta$ and $W^{\infty}(P', Q) = S^{\theta}(f)$. Let ϕ be the optimal joint distribution for W^{∞}. Let $R(x) = P(x) - P'(x)$. Let $L(P)$, $L(P')$ and $L(Q)$ denote the distributions after adding laplacian noise. We show one direction of proving DP in Appendix G; the other direction (i.e., $\Pr[L(Q) \in S] \leq e^{\epsilon} \Pr[L(P) \in S] + \delta$) can be shown by switching the role of P and Q. □

Remark 5.2. *Note that the Laplace mechanism stated in Section 2.1.1 is known to give $(\epsilon, 0)$-DP for a deterministic function f. Here, we use it for achieving (ϵ, δ)-DP for a potentially randomized function. This is made possible by our new definition of parameterized sensitivity of a (randomized) function, which extends the existing notion of sensitivity of a deterministic function (by taking $\theta = 0$).*

6 Empirical Evaluation

We empirically compare our basic mechanism $\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}$ (Construction 1) on a ground set $\mathcal{G} = \{1, \cdots, B\}$, against various competing mechanisms, for accuracy on a few histogram-based statistics computed on it. We plot average errors (actual and flexible), on different histograms for functions $\max_k(\boldsymbol{x}) := \max\{i \mid \boldsymbol{x}(i) \geq k\}$, $\max := \max_1$, and $\text{mode}(\boldsymbol{x}) := \arg\max_i \boldsymbol{x}(i)$. The parameters for $\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}$ that we will use in the section are given in Theorem 5.2.

We stress that the plots are only indicative of performance on the specific histograms, and do not suggest *worst-case accuracy guarantees*. On the other hand, our theorems do provide worst-case accuracy guarantees (Theorem 4, w.r.t. ∂_{drop} apply to max and $\max_k$, and Theorem 5 w.r.t. $\partial_{\text{drmv}}^{\eta}$ applies to mode).

Next, we shall briefly describe the other mechanisms we compare against.

Exponential Mechanism. The Exponential Mechanism [MT07] can be tailored for an abstract utility function. We consider the negative of the error, $-\text{err}(\boldsymbol{x}, y)$ as the utility of a response y on input histogram $\boldsymbol{x}$. However, for both $\max_k$ and mode, error has high sensitivity – changing a single element in the histogram can change the error by as much as the number of bars in the histogram. Since the mechanism produces an output r with probability proportional to $e^{\frac{\epsilon \text{err}(\boldsymbol{x},r)}{2\Delta_{\text{err}}}}$, where Δ_{err} is the sensitivity of err, having a large sensitivity has the effect of moving the output distribution close to a uniform distribution. This is reflected in the performance of this mechanism in all our plots.

Propose-Test-Release Mechanism (PTR). We consider the commonly used form of the PTR mechanism of Dwork and Lei [DL09], namely, "releasing stable values" (see [Vad17]). On input $\boldsymbol{x}$, the mechanism either releases the correct result $f(\boldsymbol{x})$ or refuses to do so (replacing it with a random output value), depending on whether the radius of the neighborhood of $\boldsymbol{x}$ where it remains constant is sufficiently large (after adding some noise). For computing a function f and a setting of parameter $\beta = 0$, privacy parameters ϵ, δ, the mechanism calculates this radius for a input $\boldsymbol{x}$ as, $r = d(\boldsymbol{x}, \{\boldsymbol{x}' : \text{LS}_f(\boldsymbol{x}') > 0\}) + \text{Lap}(1/\epsilon)$, where $d(,)$ is the closest hamming distance between a point, set in input histogram space and LS_f is local sensitivity of function f. If

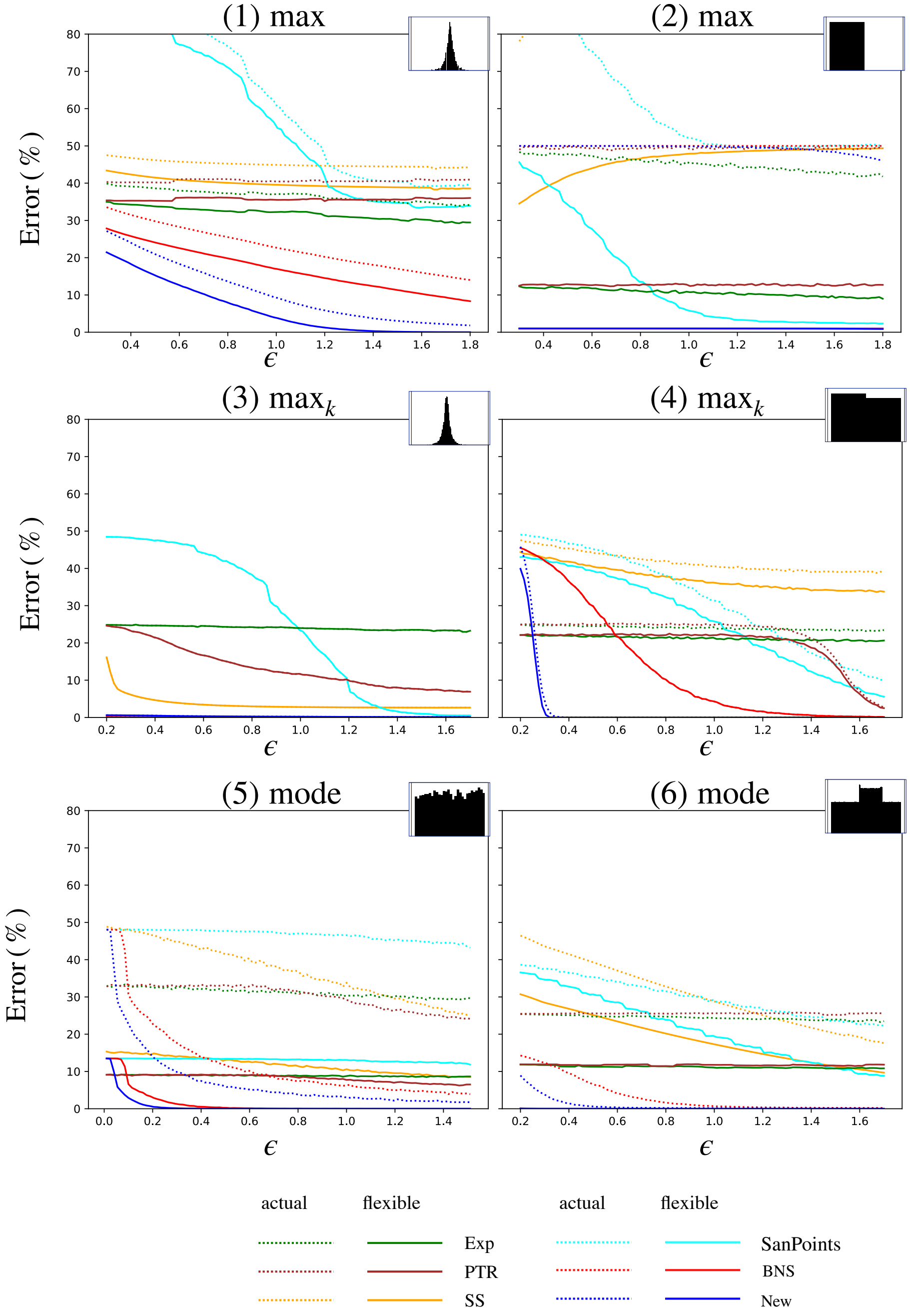

Figure 6: For each evaluation, a typical histogram used is shown in inset. The different data distributions elicit a variety of behaviors of the different mechanisms. Experiment (2) shows an instance which is hard for all the mechanisms without considering flexible accuracy; on the other hand, in Experiment (3), flexible accuracy makes no difference (the plots overlap). In these two experiments BNS and the new mechanism match each other. In all the other experiments, the new mechanism dominates the others, with or without considering flexible accuracy.

this radius r is greater than ${}^{\ln(1/\delta)}/\epsilon$, the mechanism will output the exact answer $f(\boldsymbol{x})$, otherwise it outputs a random value from the domain. For the functions we consider, this radius of stable region can be computed efficiently and is typically small or even empty for input distributions considered which is reflected in our plots.

Smooth-sensitivity Mechanism (SS). This mechanism, due to Nissim et al. [NRS07], uses the smooth sensitivity of query f as $SS^{\epsilon}_f(\boldsymbol{x}) = \max\{LS_f(\boldsymbol{x}')e^{-\epsilon d(\boldsymbol{x},\boldsymbol{x}')} | \boldsymbol{x}' \in \mathbb{H}_{\mathcal{G}}\}$, where $LS_f(\boldsymbol{x}')$ denotes the *local sensitivity* of f at $\boldsymbol{x}'$, and $d(\cdot,\cdot)$ is the hamming distance. Given an input histogram $\boldsymbol{x}$, the mechanism adds noise roughly $O(SS^{\epsilon}_f(\boldsymbol{x})/\epsilon)$ to $f(x)$ for (ϵ,δ)-privacy. For functions like $\max_k$ and mode, again, the local sensitivity tends to be large on many histograms, and this affects the performance of this mechanism on such inputs.

Stability-Based Sanitized Histogram of Bun et al. (BNS). Bun et al. [BNS19] (also see [Vad17]) gave a mechanism for releasing histograms, with provable worst-case guarantees. However, these guarantees are in terms of the errors in the individual bar heights of the histogram, and doesn't necessarily translate to the histogram based functions we consider. Nevertheless, this mechanism provides a potential candidate for a mechanism for any histogram based statistic.

The mechanism adds Laplace noise to each non-zero bar of the histogram, and the resulting value is reported only if it is more than $2\ln(2/\delta)/\epsilon n + 1/n$, otherwise sets it to 0. While this mechanism does not degenerate to providing random answers, in comparison with the Truncated Laplace mechanism, it creates larger room for error by adding (large) positive noise to (sufficiently large) bars; while this happens with a small probability for each bar, the probability of the union of this event over a large number of bars can be substantial.

Choosing-Based Histogram of Beimel et al. (SanPoints). In [BNS16], Beimel et al. presented a mechanism SanPoints for producing a sanitized histogram, with formal (α,β) PAC-guarantees for approximating the height of each bar of the histogram. For a given α,β and privacy paramters ϵ,δ, on an input $\boldsymbol{x}$ with m elements, the mechanism adds little noise, $\mathrm{Lap}(1/m\epsilon)$ to heights of bars which are iteratively chosen without repetition as per a choosing mechanism. The choosing mechanism privately picks the index with maximum bar height. It either chooses the max height index if its bar height is sufficiently large after adding some Laplace noise or otherwise chooses as per an exponential mechanism with bar height as an index's score.

For a given input $\boldsymbol{x}$ with m elements and parameters β,ϵ,δ, the mechanism guarantees $\alpha = O\Big(\big(\frac{\sqrt{\ln(1/\delta)}\ln(\ln(1/\delta))}{\epsilon m}\big)^{2/3}\Big)$. Again though the mechanism doesn't give formal guarantees for the functions we consider, such a histogram-release mechanism can be heuristically used for any histogram based statistic. In our evaluations, SanPoints yields mixed results, but is dominated by BNS and our new mechanism.

6.1 Evaluations Carried Out

In each of the following empirical evaluations, a histogram distribution and one of the following functions was fixed: $\max_k(\boldsymbol{x}) := \max\{i \mid \boldsymbol{x}(i) \geq k\}$, $\max := \max_1$, and $\mathrm{mode}(\boldsymbol{x}) := \arg\max_i \boldsymbol{x}(i)$.

(1) Function max. Histogram of about 10,000 items drawn from a Cauchy distribution with median 45 and scale 4, restricted to 100 bars, with the last 10 set to empty bars.

(2) Function max. Step histogram with two steps (height $\times$ width) : $[1000 \times 50, 1 \times 50]$.

(3) Function $\max_{500}$. Same histogram distribution as in (1) above, but without zeroing out the right-most bars.

(4) Function $\max_{500}$. Step histogram with 100 bars, with two steps (height $\times$ width) : $[540 \times 50, 490 \times 50]$.

(5) Function mode. Histogram of 30 bars, each bar has height drawn from i.i.d Poisson with mean 250.

(6) Function mode. Noisy step histogram, with steps $[130 \times 120, 200 \times 5, 185 \times 85, 190 \times 10, 130 \times 80]$.

The results are shown in Figure 6. In each experiment, a range of values for ϵ are chosen, while we fixed $\delta = 2^{-20}$. Errors are shown in the y-axis as a percentage of the full range $[0, B)$. In all experiments, for each mechanism we also compute flexible accuracy allowing distortion of $\partial_{\mathrm{drop}} = 0.005$.

7 Conclusion

DP has been a highly successful approach to modeling and solving privacy issues arising in statistical databases. However, there remain several avenues for improvement in DP, and more generally in the area of privacy.

The new notions of flexible accuracy and robustness introduced in this work greatly increase the applicability of the DP framework. Towards defining them formally, we introduced lossy Wasserstein distances (which may be of independent interest). Our definitions naturally handle mechanisms for randomized functions, as well as deterministic functions.

We illustrated the usefulness of flexible accuracy by giving new DP mechanisms for support and maximum functions, with worst-case guarantees of (flexible) accuracy. While the basic idea of dropping outliers used in these mechanisms is not new, flexible accuracy allows deriving quantitative guarantees within the DP framework, and without assuming a distribution on the data.

Our composition theorems open up a new avenue for DP. The quantities developed for framing the composition theorems – namely, distortion sensitivity and error sensitivity – provide new gauges in the dashboard when designing mechanisms for simple functions that are to be composed into more complex functions (e.g., layers of a deep neural network).

Finally, our results on robustness could be seen as a step towards privacy *beyond* DP. We leave it for future work to further pursue this line of investigation, and also to build applications that exploit our current extensions.

References

[BLR13] Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to noninteractive database privacy. *J. ACM*, 60(2):12:1–12:25, 2013. Preliminary version published in STOC 2008.

[BNS16] Amos Beimel, Kobbi Nissim, and Uri Stemmer. Private learning and sanitization: Pure vs. approximate differential privacy. *Theory of Computing*, 12(1):1–61, 2016.

[BNS19] Mark Bun, Kobbi Nissim, and Uri Stemmer. Simultaneous private learning of multiple concepts. *Journal of Machine Learning Research*, 20:94:1–94:34, 2019.

[BSU16] Mark Bun, Thomas Steinke, and Jonathan Ullman. Make up your mind: The price of online queries in differential privacy. In *SODA*, 2016.

[CDM+05] Shuchi Chawla, Cynthia Dwork, Frank McSherry, Adam D. Smith, and Hoeteck Wee. Toward privacy in public databases. In *Theory of Cryptography, Second Theory of Cryptography Conference, TCC 2005, Cambridge, MA, USA, February 10-12, 2005, Proceedings*, pages 363–385, 2005.

[CDMT05] Shuchi Chawla, Cynthia Dwork, Frank McSherry, and Kunal Talwar. On privacy-preserving histograms. In *UAI '05, Proceedings of the 21st Conference in Uncertainty in Artificial Intelligence, Edinburgh, Scotland, July 26-29, 2005*, 2005.

[DKM+06] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology - EUROCRYPT 2006, 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28 - June 1, 2006, Proceedings*, pages 486–503, 2006.

[DL09] Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 371–380, 2009.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*, pages 265–284, 2006.

[DR14] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, Aug 2014.

[KM14] Daniel Kifer and Ashwin Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Trans. Database Syst.*, 39(1):3:1–3:36, 2014.

[MT07] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), October 20-23, 2007, Providence, RI, USA, Proceedings*, pages 94–103. IEEE Computer Society, 2007.

[NRS07] Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 75–84, 2007.

[SWC17] Shuang Song, Yizhen Wang, and Kamalika Chaudhuri. Pufferfish privacy mechanisms for correlated data. In *Proceedings of the 2017 ACM International Conference on Management of Data, SIGMOD Conference 2017, Chicago, IL, USA, May 14-19, 2017*, pages 1291–1306, 2017.

[TS13] Abhradeep Thakurta and Adam D. Smith. Differentially private feature selection via stability arguments, and the robustness of the lasso. In *COLT 2013 - The 26th Annual Conference on Learning Theory, June 12-14, 2013, Princeton University, NJ, USA*, volume 30 of *JMLR Workshop and Conference Proceedings*, pages 819–850. JMLR.org, 2013.

[Vad17] Salil P. Vadhan. The complexity of differential privacy. In *Tutorials on the Foundations of Cryptography*, pages 347–450. Springer International Publishing, 2017.

[Vil08] Cedric Villani. *Optimal transport: old and new.* Springer Verlag, 2008.

[ZCP+15] Jun Zhang, Graham Cormode, Cecilia M. Procopiuc, Divesh Srivastava, and Xiaokui Xiao. Private release of graph statistics using ladder functions. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, pages 731–745. ACM, 2015.

A Omitted Details from Section 3.1 (Lossy Wasserstein Distance)

In this section we collect a few results on the lossy Wasserstein distance that will be used in the subsequent proofs. For simplicity, we assume that the infimum in the definition of lossy Wasserstein distance is always achieved; all our proofs can be easily extended to work without this assumption by taking appropriate limits when working with infinitesimal quantities.

Lemma A.1. *Let P and Q be any two distributions over a metric space $(\Omega, \mathfrak{d})$. If $W_\theta^\infty(P,Q) = \beta$, then for all $\theta_1 \in [0,\theta]$, there exist distributions P' and Q' s.t. $\Delta(P,P') \leq \theta_1$, $\Delta(Q,Q') \leq \theta - \theta_1$, and $W^\infty(P',Q') = \beta$.*

Proof. Let P and Q be any two distributions over a metric space $(\Omega, \mathfrak{d})$. Let us assume that the optimal $W_\theta^\infty(P,Q)$ $(= \beta)$ is obtained at the joint distribution ϕ_{opt}. Let the first and the second marginal distributions of ϕ_{opt} be P_{opt} and Q_{opt}, respectively. Let $\Delta(P, P_{opt}) = \theta_{opt}$, which implies that $\Delta(Q, Q_{opt}) \leq \theta - \theta_{opt}$. Define a function $R_{opt} : \Omega \to \mathbb{R}$ as $R_{opt}(\omega) := P_{opt}(\omega) - P(\omega)$ for all $\omega \in \Omega$. Clearly, $\int_\Omega R_{opt}(\omega)\,\mathrm{d}\omega = 0$ and $\int_\Omega |R_{opt}(\omega)|\,\mathrm{d}\omega = 2\theta_{opt}$.

In the discussion below, we shall take a general $\theta_1 \in [0, \theta_{opt})$ and construct distributions P' and Q' s.t. $\Delta(P,P') \leq \theta_1$, $\Delta(Q,Q') \leq \theta - \theta_1$, and $W^\infty(P',Q') = \beta$, as required in the conclusion of Lemma A.1. We can show a similar result for the other case also when $\theta_1 \in (\theta_{opt}, \theta]$ (by swapping the roles of P and Q in the above as well as in the argument below). This will complete the proof of Lemma A.1.

Define a function $R' : \Omega \to \mathbb{R}$ as $R'(\omega) := \frac{\theta_1}{\theta_{opt}} R_{opt}(\omega)$. For any $\omega \in \Omega$, let $P'(\omega) = P(\omega) + R'(\omega)$. After substituting the value of $R_{opt}(\omega) = P_{opt}(\omega) - P(\omega)$, we get $P'(\omega) = \frac{\theta_1}{\theta_{opt}} P_{opt}(\omega) + \left(1 - \frac{\theta_1}{\theta_{opt}}\right) P(\omega)$. Since P'

is a convex combination of two distributions, it is also a valid distribution. It is easy to see that $\Delta(P, P') = \theta_1$. Define a joint distribution ϕ' as follow: for every $(x, y) \in \Omega \times \Omega$, define

$$\phi'(x,y) := \begin{cases} \phi_{opt}(x,y)\frac{P'(x)}{P_{opt}(x)} & \text{if } P_{opt}(x) > 0 \\ P'(x)\delta(x-y) & \text{otherwise} \end{cases}$$

where $\delta(\cdot)$ is the Dirac delta function. It follows from the definition that $\int_\Omega \phi'(x, y)\,\mathrm{d}y = P(x)$, i.e., the first marginal of ϕ' is $P(\cdot)$. This also implies that ϕ' is a valid joint distribution because (i) $\phi'(x, y) \geq 0$ for all $(x, y) \in \Omega \times \Omega$, and (ii) $\int_{\Omega\times\Omega} \phi'(x, y)\,\mathrm{d}x\,\mathrm{d}y = \int_\Omega P'(x)\,\mathrm{d}x = 1$.

Let the second marginal of ϕ' be Q'. We show below that $\Delta(Q, Q') \leq \theta - \theta_1$:

$$\begin{aligned}
\Delta(Q,Q') &\leq \Delta(Q,Q_{opt}) + \Delta(Q_{opt},Q') \\
&\leq \theta - \theta_{opt} + \frac{1}{2}\int_\Omega |Q_{opt}(y) - Q'(y)|\,\mathrm{d}y && \text{(Since } \Delta(Q,Q_{opt}) \leq \theta - \theta_{opt}\text{)} \\
&= \frac{1}{2}\int_\Omega \left| \int_\Omega \phi_{opt}(x,y)\,\mathrm{d}x - \int_\Omega \phi'(x,y)\,\mathrm{d}x \right| \mathrm{d}y + (\theta - \theta_{opt}) \\
&\leq \frac{1}{2}\int_\Omega \int_\Omega |\phi_{opt}(x,y) - \phi'(x,y)|\,\mathrm{d}x\,\mathrm{d}y + (\theta - \theta_{opt}) \\
&= \frac{1}{2}\int_\Omega \int_{x\in\Omega:P_{opt}(x)>0} |\phi_{opt}(x,y) - \phi'(x,y)|\,\mathrm{d}x\,\mathrm{d}y \\
&\quad + \frac{1}{2}\int_\Omega \int_{x\in\Omega:P_{opt}(x)=0} |\phi_{opt}(x,y) - \phi'(x,y)|\,\mathrm{d}x\,\mathrm{d}y + (\theta - \theta_{opt}) && (25)
\end{aligned}$$

We bound the two integral on the RHS of (25) separately, depending on whether $P_{opt}(x)$ is zero or not. Define $\Omega_1 := \{x \in \Omega : P_{opt}(x) > 0\}$.

Case 1. $P_{opt}(x) > 0.$:

$$\begin{aligned}
\frac{1}{2}\int_\Omega \int_{\Omega_1} |\phi_{opt}(x,y) - \phi'(x,y)|\,\mathrm{d}x\,\mathrm{d}y &= \frac{1}{2}\int_\Omega \int_{\Omega_1} \phi_{opt}(x,y)\left|1 - \frac{P'(x)}{P_{opt}(x)}\right| \mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{2}\int_{\Omega_1} \left|1 - \frac{P'(x)}{P_{opt}(x)}\right| \mathrm{d}x \int_\Omega \phi_{opt}(x,y)\,\mathrm{d}y \\
&= \frac{1}{2}\int_{\Omega_1} |P_{opt}(x) - P'(x)|\,\mathrm{d}x && \text{(Since } \textstyle\int_\Omega \phi_{opt}(x,y)\,\mathrm{d}y = P_{opt}(x)\text{)} \\
&\overset{\text{(a)}}{=} \frac{1}{2}\int_{\Omega_1} \left|\left(1 - \frac{\theta_1}{\theta_{opt}}\right) R_{opt}(x)\right| \mathrm{d}x \\
&= \frac{(\theta_{opt} - \theta_1)}{2\theta_{opt}} \int_{\Omega_1} |R_{opt}(x)|\,\mathrm{d}x \\
&\leq \frac{(\theta_{opt} - \theta_1)}{2\theta_{opt}} \int_\Omega |R_{opt}(x)|\,\mathrm{d}x && \text{(Since } \Omega_1 \subseteq \Omega\text{)} \\
&= \theta_{opt} - \theta_1 && \text{(Since } \textstyle\int_\Omega |R_{opt}(\omega)|\,\mathrm{d}\omega = 2\theta_{opt}\text{)}
\end{aligned}$$

Here (a) follows because for every $x \in \Omega$, we have $P_{opt}(x) - P'(x) = R_{opt}(x) + P(x) - P'(x) = R_{opt}(x) - R'(x) = R_{opt}(x) - \frac{\theta_1}{\theta_{opt}} R_{opt}(x)$.

Case 2. $P_{opt}(x) = 0$. Note that this also implies $\phi_{opt}(x, y) = 0$ for all $y \in \Omega$:

$$\frac{1}{2}\int_\Omega \int_{\Omega\setminus\Omega_1} |\phi_{opt}(x,y) - \phi'(x,y)|\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{2}\int_\Omega \int_{\Omega\setminus\Omega_1} |\phi'(x,y)|\,\mathrm{d}x\,\mathrm{d}y$$

$$
\begin{aligned}
&= \frac{1}{2}\int_{\Omega}\int_{\Omega\setminus\Omega_1} P'(x)\delta(x-y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{2}\int_{\Omega\setminus\Omega_1} P'(x)\,\mathrm{d}x \\
&= \frac{1}{2}\int_{\Omega\setminus\Omega_1} |P_{opt}(x) - P'(x)|\,\mathrm{d}x && \text{(Since } P_{opt}(x)=0 \text{ whenever } x\in\Omega\setminus\Omega_1\text{)} \\
&\le \frac{1}{2}\int_{\Omega} |P_{opt}(x) - P'(x)|\,\mathrm{d}x && \text{(Since } \Omega\setminus\Omega_1\subseteq\Omega\text{)} \\
&= \theta_{opt} - \theta_1
\end{aligned}
$$

The last equality follows from the steps that we took in the first case.

Substituting these bounds in (25) gives $\Delta(Q,Q') \le \theta - \theta_1$.

The only thing left to prove is to show that $W^{\infty}(P',Q') = \beta$ for the above constructed P' and Q'. First, note that, since $W^{\infty}_{\theta}(P,Q) = \beta$, we have $W^{\infty}(P',Q') \ge \beta$. This follows because

$$
\begin{aligned}
W^{\infty}_{\theta}(P,Q) &\overset{(a)}{=} \inf_{\phi\in\Phi^{\theta}(P,Q)} \max_{(x,y)\leftarrow\phi} \mathfrak{d}(x,y) \\
&\overset{(b)}{=} \inf_{\substack{\hat{P},\hat{Q}:\\ \Delta(P,\hat{P})+\Delta(Q,\hat{Q})\le\theta}} \inf_{\phi\in\Phi^{0}(\hat{P},\hat{Q})} \max_{(x,y)\leftarrow\phi} \mathfrak{d}(x,y) \\
&\overset{(c)}{=} \inf_{\substack{\hat{P},\hat{Q}:\\ \Delta(P,\hat{P})+\Delta(Q,\hat{Q})\le\theta}} W^{\infty}(\hat{P},\hat{Q}), && (26)\\
&\le W^{\infty}(P',Q'), && (27)
\end{aligned}
$$

where (a) follows from the definition of γ-Lossy ∞-Wasserstein distance; (b) trivially holds by viewing the infimum set differently; in (c) we substituted the definition of W_{∞}; and (d) follows because P',Q' satisfies $\Delta(P,P') + \Delta(Q,Q') \le \theta$.

Therefore, showing $W^{\infty}(P',Q') \le \beta$ suffices.

For the sake of contradiction, let us assume that $W^{\infty}(P',Q') > \beta$. Then there is a pair $(x,y)\in\Omega^2$ such that $\phi'(x,y) > 0$ and $\mathfrak{d}(x,y) > \beta$. This implies that $\phi_{opt}(x,y) = 0$, because, otherwise, we would have $W^{\infty}(P_{opt},Q_{opt}) > \beta$, which contradicts our hypothesis that $W^{\infty}(P_{opt},Q_{opt}) = \beta$. So, we know that $\phi'(x,y) > 0$ and $\phi_{opt}(x,y) = 0$. From the definition of ϕ', this is only possible if $P_{opt}(x) = 0$ and $P'(x)\delta(x-y) > 0$. This can happen only if $x = y$, but this implies $\mathfrak{d}(x,y) = 0 \le \beta$, which is a contradiction. Hence $W^{\infty}(P',Q') \le \beta$.

This completes the proof. □

Corollary A.1. *Let P and Q be any two distributions over a metric space $(\Omega,\mathfrak{d})$. If $W^{\infty}_{\theta}(P,Q) = \beta$, then there exists a distribution P' s.t. $\Delta(P,P') \le \theta$ and $W^{\infty}(P',Q) = \beta$. Similary, there exists a distribution Q' s.t. $\Delta(Q,Q') \le \theta$ and $W^{\infty}(P,Q') = \beta$*

Proof. Using Lemma A.1 with $\theta_1 = \theta$, there exists distributions P' and Q' such that $\Delta(P,P') \le \theta$, $\Delta(Q,Q') \le 0$ and $W^{\infty}(P',Q') = \beta$. Note that $\Delta(Q,Q') \le 0$ is possible if and only if $Q = Q'$, which gives the result. Similary, the second statement can be proved sing Lemma A.1 with $\theta_1 = 0$ □

We now show that θ-lossy ∞-Wasserstein distance follows triangle-inequality. In the proof, we will use the fact that Wasserstein distance is a metric, which is proved in [Vil08].

Lemma (Restating Lemma 3.1)**.** *For distributions P, Q, and R over a metric space $(\Omega,\mathfrak{d})$, and $\gamma_1,\gamma_2\in[0,1]$.*

$$
W^{\infty}_{\gamma_1+\gamma_2}(P,R) \le W^{\infty}_{\gamma_1}(P,Q) + W^{\infty}_{\gamma_2}(Q,R). \tag{28}
$$

Proof. Let $W^\infty_{\gamma_1}(P,Q)=\beta_1$ and $W^\infty_{\gamma_2}(Q,R)=\beta_2$. Using Corollary A.1, we get a P' such that $\Delta(P,P')\le\gamma_1$ and $W^\infty(P',Q)=\beta_1$. Similarly, we get a R' such that $\Delta(R,R')\le\gamma_2$ and $W^\infty(Q,R')=\beta_2$. Using the fact that the Wasserstein distance is a metric, we get $W^\infty(P',R')\le\beta_1+\beta_2$. Let ϕ be the optimal joint distribution for $W^\infty(P',R')$. Since $\Delta(P,P')\le\gamma_1$ and $\Delta(R,R')\le\gamma_2$, $\phi\in\Phi^{\gamma_1+\gamma_2}(P,R)$. Therefore, $W^\infty_{\gamma_1+\gamma_2}\le\beta_1+\beta_2$. $\square$

The following result, used in Appendix F is proven when the metric satisfies additional conditions. Specifically, we shall require that the metric corresponds to the metric in a normed vector space (e.g., $\mathbb{R}^d$).

Lemma (Restating Lemma 3.2)**.** *Let X,Y and Z be three random variables over a normed vector space, Ω with Z being independent of X and Y, and let $\mathfrak{p}_0$ denote the distribution with all its mass at 0. Then, the lossy ∞-Wasserstein distance defined using the metric induced by the norm of Ω satisfies the following: $\forall\gamma,\gamma_1$ such that $\gamma,\gamma_1\ge 0$ and $\gamma_1\le\gamma/2$, we have*

$$W^\infty_\gamma(X+Z,Y+Z)\overset{\text{(a)}}{\le}W^\infty_\gamma(X,Y)\overset{\text{(b)}}{\le}W^\infty_{\gamma-2\gamma_1}(X+Z,Y+Z)+2W^\infty_{\gamma_1}(\mathfrak{p}_0,Z). \tag{29}$$

Proof. Let $\mathfrak{d}$ denote the metric induced by the norm on the vector space.

First we prove (a) and then we prove (b).

Proof of (a). Let $W^\infty_\gamma(X,Y)=\beta$ and let $\phi\in\Phi^\gamma(X,Y)$ be any joint distribution with $\max_{(x,y)\leftarrow\phi}\mathfrak{d}(x,y)=\beta$. Let ϕ_1,ϕ_2 denote its marginals, i.e., $\phi_1(x)=\int_\Omega\phi(x,y)\,\mathrm{d}y$ and $\phi_2(y)=\int_\Omega\phi(x,y)\,\mathrm{d}x$. Consider the following joint distribution ϕ':

$$\phi'(x,y)=\int_\Omega Z(z)\phi(x-z,y-z)\,\mathrm{d}z. \tag{30}$$

Now we show that $\phi'\in\Phi^\gamma(X+Z,Y+Z)$. For this, let ϕ'_1,ϕ'_2 denote the marginals of ϕ'. We can easily show that $\phi'_1(x)=\int_\Omega Z(z)\phi_1(x-z)\,\mathrm{d}z$ and $\phi'_2(y)=\int_\Omega Z(z)\phi_2(y-z)\,\mathrm{d}z$. Now, we compute the statistical difference between $X+Z$ and ϕ'_1:

$$\begin{aligned}
\Delta(X+Z,\phi'_1)&=\frac{1}{2}\int_\Omega|X+Z(x)-\phi'_1(x)|\,\mathrm{d}x\\
&=\frac{1}{2}\int_\Omega\int_\Omega Z(z)|X(x-z)-\phi_1(x-z)|\,\mathrm{d}z\,\mathrm{d}x\\
&=\int_\Omega Z(z)\left[\frac{1}{2}\int_\Omega|X(x-z)-\phi_1(x-z)|\,\mathrm{d}x\right]\mathrm{d}z\\
&=\int_\Omega Z(z)\Delta(X,\phi_1)\,\mathrm{d}z\\
&=\Delta(X,\phi_1).
\end{aligned}$$

Similarly, we can show $\Delta(Y+Z,\phi'_2)=\Delta(Y,\phi_2)$. This, together with $\Delta(X,\phi_1)+\Delta(Y,\phi_2)\le\gamma$ (which follows because $\phi\in\Phi^\gamma(X,Y)$) implies $\Delta(X+Z,\phi'_1)+\Delta(Y+Z,\phi'_2)\le\gamma$. Thus, we have shown that $\phi'\in\Phi^\gamma(X+Z,Y+Z)$.

Now, we show that $\max_{(x,y)\leftarrow\phi'}\mathfrak{d}(x,y)\le\max_{(x,y)\leftarrow\phi}\mathfrak{d}(x,y)$. For the sake of contradiction, suppose $\max_{(x,y)\leftarrow\phi}\mathfrak{d}(x,y)=\beta$ and $\exists(x^*,y^*)$ s.t. $\mathfrak{d}(x^*,y^*)>\beta$ and $\phi'(x^*,y^*)>0$. Since $\mathfrak{d}$ is a metric induced by the norm on the vector space, we have that $\forall z,\mathfrak{d}(x^*-z,y^*-z)=\mathfrak{d}(x^*,y^*)$. This, together with $\mathfrak{d}(x^*,y^*)>\beta$ and $\max_{(x,y)\leftarrow\phi}\mathfrak{d}(x,y)=\beta$ implies that $\phi(x^*-z,y^*-z)=0$ for all z. Therefore, $\phi'(x^*,y^*)=\int_\Omega Z(z)\phi(x^*-z,y^*-z)\,\mathrm{d}z=0$, which is the required contradiction. Now, the inequality (a) of (29) follows from

$$\begin{aligned}
W^\infty_\gamma(X+Z,Y+Z)&=\inf_{\phi'\in\Phi^\gamma(X+Z,Y+Z)}\max_{(x,y)\leftarrow\phi'}\mathfrak{d}(x,y)\\
&\le\max_{(x,y)\leftarrow\phi'}\mathfrak{d}(x,y) &&\text{(for the specific }\phi'\text{ constructed in (30))}
\end{aligned}$$

$$\leq \max_{(x,y)\leftarrow\phi} \mathfrak{d}(x,y)$$
$$= W_{\gamma}^{\infty}(X,Y).$$

Proof of (b). By Lemma 3.1, we can show that

$$W_{\gamma}^{\infty}(X,Y) \leq W_{\gamma-2\gamma_1}^{\infty}(X+Z,Y+Z) + W_{\gamma_1}^{\infty}(X,X+Z) + W_{\gamma_1}^{\infty}(Y,Y+Z).$$

It follows from (a) that

$$W_{\gamma_1}^{\infty}(X,X+Z) \leq W_{\gamma_1}^{\infty}(0,Z)$$
$$W_{\gamma_1}^{\infty}(Y,Y+Z) \leq W_{\gamma_1}^{\infty}(0,Z).$$

Combining the above three inequalities gives the required inequality (b) in (29). This completes the proof of Lemma 3.2. $\square$

A.1 Average Version of Lossy Wasserstein Distance

Lemma (Restating Lemma 3.3)**.** *For any two distributions P, Q, and $0 \leq \beta' < \beta \leq 1$,*

$$W_{\beta}(P,Q) \leq W_{\beta}^{\infty}(P,Q) \leq W_{\beta'}(P,Q)/(\beta-\beta').$$

Proof. Clearly from the definitions, $W_{\beta}(P,Q) \leq W_{\beta}^{\infty}(P,Q)$.

Suppose $W_{\beta'}(P,Q) = \gamma$ and $\phi \in \Phi^{\beta'}(P,Q)$ is an optimal coupling that realizes this. Then, in ϕ, the total mass that is transported more than a distance γ' is at most γ/γ' and the total mass that is lost is at most β'. By choosing to simply not transport this mass at all, one loses $\beta' + \gamma/\gamma'$ mass, but no mass is transported more than a distance γ'. Choosing $\gamma' = \gamma/(\beta-\beta')$ this upper bound on loss is β, and hence this modified coupling shows that $W_{\beta}^{\infty}(P,Q) \leq \gamma'$. $\square$

B Omitted Details from Section 4 (Composition Theorems)

B.1 Flexible Accuracy Under Composition

In this section, we prove Lemma 4.1, Lemma 4.2, and Lemma 4.3.

Lemma (Restating Lemma 4.1)**.** *If ∂ is a measure of distortion over A, then $\widehat{\partial}$ is a quasi-metric.*

Proof. We need to show that for any three distributions P, Q, and R over the same space A, we have (i) $\widehat{\partial}(P,Q) \geq 0$, where the equality holds if and only if $P = Q$, and (ii) $\widehat{\partial}$ satisfies the triangle inequality: $\widehat{\partial}(P,Q) \leq \widehat{\partial}(P,R) + \widehat{\partial}(R,Q)$. We show them one by one below:

1. The first property follows from the definition of $\widehat{\partial}$ (see Definition 10): If $\widehat{\partial}(P,Q) = 0$, then the optimal $\phi \in \Phi(P,Q)$ is a diagonal distribution, which means that $P = Q$. On the other hand, if $P = Q$, then there exists a coupling ϕ in $\Phi(P,Q)$, which is a diagonal distribution and hence $\widehat{\partial}(P,Q) = 0$.
2. For the second property, let $\phi_2 \in \Phi(P,R)$ and $\phi_3 \in \Phi(R,Q)$ denote the optimal couplings for $\widehat{\partial}(P,R)$ and $\widehat{\partial}(R,Q)$, respectively, i.e., $\widehat{\partial}(P,R) = \sup_{\substack{(x,y):\\ \phi_2(x,y)\neq 0}} \partial(x,y)$ and $\widehat{\partial}(R,Q) = \sup_{\substack{(y,z):\\ \phi_3(y,z)\neq 0}} \partial(y,z)$. It follows from the Gluing Lemma [Vil08] that we can find a coupling ϕ' over $A \times A \times A$ such that the projection of ϕ' onto its first two coordinates is equal to ϕ_2 and its last two coordinates is equal to ϕ_3. Let ϕ_1 denote the projection of ϕ' onto its first and the third coordinates. Note that $\phi_1 \in \Phi(P,Q)$, but it may not be an optimal coupling for $\widehat{\partial}(P,Q)$. Now the triangle inequality follows from the following set of inequalities:

$$\widehat{\partial}(P,Q) = \inf_{\phi\in\Phi(p,q)} \sup_{\substack{(x,z):\\ \phi(x,z)\neq 0}} \partial(x,z) \quad \leq \quad \sup_{\substack{(x,z):\\ \phi_1(x,z)\neq 0}} \partial(x,z) \quad = \quad \sup_{\substack{(x,y,z):\\ \phi'(x,y,z)\neq 0}} \partial(x,z)$$

$$
\begin{aligned}
&\overset{(a)}{\leq} \sup_{\substack{(x,y,z):\\ \phi'(x,y,z)\neq 0}} \partial(x,y) + \partial(y,z) \\
&= \sup_{\substack{(x,y,z):\\ \phi'(x,y,z)\neq 0}} \partial(x,y) \quad + \sup_{\substack{(x,y,z):\\ \phi'(x,y,z)\neq 0}} \partial(y,z) \\
&= \sup_{\substack{(x,y):\\ \phi_2(x,y)\neq 0}} \partial(x,y) \quad + \sup_{\substack{(y,z):\\ \phi_3(y,z)\neq 0}} \partial(y,z) \\
&= \widehat{\partial}(P,R) + \widehat{\partial}(R,Q),
\end{aligned}
$$

where (a) follows from the fact that ∂ is a measure of distortion, which is a quasi-metric. □

Lemma (Restating Lemma 4.2)**.** *If $\mathcal{M} : A \to B$ is an (α, β, γ)-accurate mechanism for f w.r.t. ∂, then for any random variable X over A, there is a random variable X^* such that*

$$\widehat{\partial}(X, X^*) \leq \alpha, \qquad W_\gamma^\infty(f(X^*), \mathcal{M}(X)) \leq \beta.$$

Proof. From the accuracy guarantee of $\mathcal{M}$, we have for each x, there is a random variable X_x such that $\mathrm{supp}(X_x) \subseteq \{x' \mid \partial(x,x') \leq \alpha\}$ and $W_\gamma^\infty(f(X_x), \mathcal{M}(x)) \leq \beta$, where $\mathrm{supp}(Y)$ denotes the support of Y.

The first condition in the statement of the lemma follows by considering ϕ in the definition of $\widehat{\partial}$ to be the distribution of the pairs (x, x') where x is sampled according to X, and then x' is sampled according to X_x. The random variable X^* is defined by the distribution of x' in the above experiment. Then, note that $\phi \in \Phi^0(X, X^*)$ and for all (x, x') in its support, $\partial(x, x') \leq \alpha$.

To see the second part, let ϕ_x denote the optimal distribution for $W_\gamma^\infty(f(X_x), \mathcal{M}(x))$. That is, for each x, $\phi_x \in \Phi^\gamma(f(X_x), \mathcal{M}(x))$ and $W_\gamma^\infty(f(X_x), \mathcal{M}(x)) = \max_{(a,b)\leftarrow\phi_x} \mathfrak{d}(a,b)$. Let ϕ be defined by $\phi(a,b) = X(x)\phi_x(a,b)$. It is easy to verify that $\phi \in \Phi^\gamma(f(X^*), \mathcal{M}(X))$. Further,

$$
\begin{aligned}
W_\gamma^\infty(f(X^*), \mathcal{M}(X)) &\leq \max_{(a,b)\leftarrow\phi} \mathfrak{d}(a,b) = \max_{x\leftarrow X} \max_{(a,b)\leftarrow\phi_x} \mathfrak{d}(a,b) \\
&= \max_{x\leftarrow X} W_\gamma^\infty(f(X_x), \mathcal{M}(x)) \leq \beta.
\end{aligned}
$$

This completes the proof of Lemma 4.2. □

The following lemma translates the definition of distortion sensitivity (Definition 11) to apply to distortion of input distributions. It is stated assuming that all infima in the definition of σ are achieved. This is the case, e.g., when the measure of distortion is discrete.

Lemma (Restating Lemma 4.3)**.** *Suppose $f : A \to B$ has distortion sensitivity $\sigma_f^{\theta,\omega}$, w.r.t. (∂_1, ∂_2). Then, for random variables X_0 over A, and Y over B such that $\widehat{\partial}_2(f(X_0), Y) \leq \alpha$, there exists a distribution X over A such that $W_\theta^\infty(f(X), Y) \leq \omega$ and $\widehat{\partial}_1(X_0, X) \leq \sigma_f^{\theta,\omega}(\alpha)$.*

Proof. Fix random variables X_0 over A, and Y over B such that $\widehat{\partial}_2(f(X_0), Y) \leq \alpha$. Let ϕ be an optimal coupling that achieves the infimum in the definition of $\widehat{\partial}_2(f(X_0), Y)$, i.e.,

$$\widehat{\partial}_2(f(X_0), Y) = \sup_{(u,y)\leftarrow\phi} \partial_2(u, y). \tag{31}$$

For each $x_0 \in \mathrm{supp}(X_0)$, consider the conditional distribution $\phi_{x_0} = \phi|\{X_0 = x_0\}$. Clearly, the first marginal of ϕ_{x_0} is $f(x_0)$. Let its second marginal be denoted by Y_{x_0}. First we show that for each $x_0 \in \mathrm{supp}(X_0)$, we have $\widehat{\partial}_2(f(x_0), Y_{x_0}) \leq \alpha$.

$$\widehat{\partial}_2(f(x_0), Y_{x_0}) = \inf_{\phi\in\Phi^0(f(x_0),Y_{x_0})} \sup_{(u,y)\leftarrow\phi} \partial_2(u, y)$$

$$
\begin{aligned}
&\leq \sup_{(u,y)\leftarrow\phi_{x_0}} \partial_2(u,y) \\
&\overset{\text{(a)}}{\leq} \sup_{(u,y)\leftarrow\phi} \partial_2(u,y) \\
&\overset{\text{(b)}}{=} \widehat{\partial}_2(f(X_0),Y) \\
&\leq \alpha.
\end{aligned}
$$

Here (a) follows from the fact that $\mathrm{supp}(\phi_{x_0}) \subseteq \mathrm{supp}(\phi)$ and (b) follows from (31). Thus for each $x_0 \in \mathrm{supp}(X_0)$, we have $\widehat{\partial}_2(f(x_0), Y_{x_0}) \leq \alpha$. By the definition of $\sigma_f^{\theta,\omega}$, there exist X_{x_0} such that,

$$
W_\theta^\infty(f(X_{x_0}), Y_{x_0}) \leq \omega, \quad \widehat{\partial}_1(x_0, X_{x_0}) \leq \sigma_f^{\theta,\omega}(\alpha)
$$

Now, define $X = \sum_{x_0\in\mathrm{supp}(X_0)} X_0(x_0)X_{x_0}$. Now, using a similar argument as used to prove the second part of Lemma 4.2, in the following we show that $W_\theta^\infty(f(X),Y) \leq \omega$ and $\widehat{\partial}_1(X_0,X) \leq \sigma_f^{\theta,\omega}(\alpha)$.

- Showing $W_\theta^\infty(f(X),Y) \leq \omega$: For each $x_0 \in \mathrm{supp}(X_0)$, let ψ_{x_0} be the optimal coupling that achieves the infimum in the definition of $W_\theta^\infty(f(X_{x_0}), Y_{x_0})$. That is, for each x_0, $\psi_{x_0} \in \Phi^\theta(f(X_{x_0}), Y_{x_0})$ and $W_\theta^\infty(f(X_{x_0}), Y_{x_0}) = \mathbb{E}_{(a,b)\leftarrow\psi_{x_0}}[\mathfrak{d}(a,b)]$. Let ψ be defined by $\psi(a,b) = X_0(x_0)\psi_{x_0}(a,b)$. It is easy to verify that $\psi \in \Phi^\theta(f(X),Y)$. Further,

$$
\begin{aligned}
W_\theta^\infty(f(X),Y) &\leq \max_{(a,b)\leftarrow\psi} \mathfrak{d}(a,b) \\
&= \max_{x_0\leftarrow X_0} \max_{(a,b)\leftarrow\psi_{x_0}} [\mathfrak{d}(a,b) \\
&= \max_{x_0\leftarrow X_0} [W_\theta^\infty(f(X_{x_0}), Y_{x_0})] \\
&\leq \omega.
\end{aligned}
$$

- Showing $\widehat{\partial}_1(X_0,X) \leq \sigma_f^{\theta,\omega}(\alpha)$: For each $x_0 \in \mathrm{supp}(X_0)$, let ψ_{x_0} be the optimal coupling that achieves the infimum in the definition of $\widehat{\partial}_1(x_0, X_{x_0})$. That is, for each x_0, $\psi_{x_0} \in \Phi^0(x_0, X_{x_0})$ and $\widehat{\partial}_1(x_0, X_{x_0}) = \sup_{(a,b)\leftarrow\psi_{x_0}} \partial_1(a,b)$. Let ψ be defined by $\psi(a,b) = X_0(x_0)\psi_{x_0}(a,b)$. It is easy to verify that $\psi \in \Phi^0(X_0,X)$. Further,

$$
\begin{aligned}
\widehat{\partial}_1(X_0,X) &\leq \sup_{(a,b)\leftarrow\psi} \partial_1(a,b) \\
&= \sup_{x_0\leftarrow X_0} \sup_{(a,b)\leftarrow\psi_{x_0}} \partial_1(a,b) \\
&= \sup_{x_0\leftarrow X_0} \widehat{\partial}_1(x_0, X_{x_0}) \\
&\leq \sigma_f^{\theta,\omega}(\alpha).
\end{aligned}
$$

This completes the proof of Lemma 4.3. $\square$

B.2 Differential Privacy Under Composition

Theorem (Restating Theorem 4.2)**.** *Let $\mathcal{M}_1 : A \to B$ and $\mathcal{M}_2 : B \to C$ be any two mechanisms. If $\mathcal{M}_1$ is neighborhood-preserving w.r.t. neighborhood relations $\sim_A$ and $\sim_B$ over A and B, respectively, and $\mathcal{M}_2$ is (ϵ,δ)-DP w.r.t. $\sim_B$, then $\mathcal{M}_2 \circ \mathcal{M}_1 : A \to C$ is (ϵ,δ)-DP w.r.t. $\sim_A$.*

Proof. For simplicity, we consider the case when B is discrete. The proof can be generalized to the continuous setting.

Since the mechanism $\mathcal{M}_1$ is neighborhood preserving, for $x, x' \in A$ s.t. $x_1 \sim_A x_2$, there exists a pair of jointly distributed random variables (X_1, X_2) over $B \times B$ s.t, $X_1 = \mathcal{M}_1(x)$, $X_2 = \mathcal{M}_1(x')$ and $\Pr[X_1 \sim_B X_2] = 1$. So, for all (x_1, x_2) such that $X_1, X_2(x_1, x_2) > 0$, we have $x_1 \sim_B x_2$ and hence, by the (ϵ, δ)-differential privacy of the mechanism $\mathcal{M}_2$, for all subsets $S \subseteq C$, we have,

$$\Pr(\mathcal{M}_2(x_1) \in S) \leq e^{\epsilon} \Pr(\mathcal{M}_2(x_2) \in S) + \delta.$$

Thus, if $x \sim_A x'$, then for any subset $S \subseteq C$, we have,

$$\begin{aligned}
\Pr[\mathcal{M}_2(\mathcal{M}_1(x)) \in S] &= \sum_{x_1} X_1(x_1) \Pr[\mathcal{M}_2(x_1) \in S] \\
&= \sum_{(x_1,x_2)} X_1, X_2(x_1, x_2) \Pr[\mathcal{M}_2(x_1) \in S] \\
&\leq \sum_{(x_1,x_2)} X_1, X_2(x_1, x_2) \left(e^{\epsilon} \Pr[\mathcal{M}_2(x_2) \in S] + \delta\right) \\
&= e^{\epsilon} \left(\sum_{(x_1,x_2)} X_1, X_2(x_1, x_2) \Pr[\mathcal{M}_2(x_2) \in S] \right) + \delta \\
&= e^{\epsilon} \left(\sum_{x_2} X_2(x_2) \Pr[\mathcal{M}_2(x_2) \in S] \right) + \delta \\
&= e^{\epsilon} \Pr[\mathcal{M}_2(\mathcal{M}_1(x')) \in S] + \delta
\end{aligned}$$

□

C Omitted Details from Section 5.1.2 (Histogram Mechanism)

Claim (Restating Claim 5.1)**.** *Whenever $\boldsymbol{s} \in S_k, k \in \{0, 2\}$, we have, $D_{q'}(s_{i^*} - x'_{i^*}) \leq e^{(1+\nu)\epsilon} D_q(s_{i^*} - x_{i^*})$, provided $n \geq \frac{2}{\epsilon\tau} \ln\left(1 + \frac{1 - e^{-\frac{\epsilon\tau}{2}}}{e^{\epsilon(\nu + \frac{\tau}{2})} - 1}\right)$.*

Proof. For $\boldsymbol{s} \in S_0$, $D_{q'}(s_{i^*} - x'_{i^*}) = 0$ so the inequality trivially holds. For $\boldsymbol{s} \in S_2$, both $D_{q'}(s_{i^*} - x'_{i^*}) > 0$ and $D_q(s_{i^*} - x_{i^*}) > 0$; hence, we will be done if we show that $\frac{D_{q'}(s_{i^*} - x'_{i^*})}{D_q(s_{i^*} - x_{i^*})} \leq e^{(1+\nu)\epsilon}$. Note that we are given the following inequality:

$$n \geq \frac{2}{\epsilon\tau} \ln\left(1 + \frac{1 - e^{-\epsilon\frac{\tau}{2}}}{e^{\epsilon(\nu + \frac{\tau}{2})} - 1}\right),$$

which can be rewritten as

$$\ln\left(\frac{1 - e^{-\epsilon\frac{\tau(n+1)}{2}}}{1 - e^{-\epsilon\frac{\tau n}{2}}}\right) \leq \epsilon(\nu + \frac{\tau}{2}). \tag{32}$$

By substituting $q = \tau(n+1)$ and $q' = \tau n$, (32) is equivalent to

$$\frac{1}{\epsilon} \ln\left(\frac{1 - e^{-\epsilon\frac{q}{2}}}{1 - e^{-\epsilon\frac{q'}{2}}}\right) + (1 - \frac{\tau}{2}) \leq 1 + \nu.$$

This, using the triangle inequality, implies that

$$\frac{1}{\epsilon} \ln\left(\frac{1 - e^{-\epsilon\frac{q}{2}}}{1 - e^{-\epsilon\frac{q'}{2}}}\right) + \left|s_{i^*} - x_{i^*} + \frac{q}{2}\right| - \left|s_{i^*} - x_{i^*} + \frac{q}{2} + (1 - \frac{\tau}{2})\right| \leq 1 + \nu.$$

Putting $q' = q - \tau$ and $x'_{i^*} = x_{i^*} - 1$, we get

$$\frac{1}{\epsilon}\ln\left(\frac{1-e^{-\epsilon\frac{q}{2}}}{1-e^{-\epsilon\frac{q'}{2}}}\right) + \left|s_{i^*} - x_{i^*} + \frac{q}{2}\right| - \left|s_{i^*} - x'_{i^*} + \frac{q'}{2}\right| \leq 1+\nu.$$

By taking exponents of both sides, this is equivalent to showing

$$\frac{(1-e^{-\epsilon\frac{q}{2}})}{(1-e^{-\epsilon\frac{q'}{2}})}\frac{e^{-\epsilon|s_{i^*}-x'_{i^*}+\frac{q'}{2}|}}{e^{-\epsilon|s_{i^*}-x_{i^*}+\frac{q}{2}|}} \leq e^{(1+\nu)\epsilon}$$

By substituting the values of $D_q(s_{i^*} - x_{i^*})$ and $D_{q'}(s_{i^*} - x'_{i^*})$, this can be equivalently written as $\frac{D_{q'}(s_{i^*}-x'_{i^*})}{D_q(s_{i^*}-x_{i^*})} \leq e^{(1+\nu)\epsilon}$, which concludes the proof of Claim 5.1. $\square$

Claim (Restating Claim 5.2)**.** $\Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}') \in S_{-1}] \leq e^{(1+\nu)\epsilon}\Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}) \in S_{-1}]$.

Proof.

$$\begin{aligned}
\Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}') \in S_{-1}] &= \int_{S_{-1}} \Big[\prod_{i\in\mathcal{G}_{\boldsymbol{x}'}} D_{q'}(s_i - x'_i)\Big]\,\mathrm{d}\boldsymbol{s} \\
&= \int_{S_{-1}} \Big[\prod_{i\in\mathcal{G}_{\boldsymbol{x}'}:i\neq i^*} D_{q'}(s_i - x'_i)\Big] D_{q'}(s_{i^*} - x'_{i^*})\,\mathrm{d}\boldsymbol{s} \\
&\leq \int_{S_{-1}} \Big[\prod_{i\in\mathcal{G}_{\boldsymbol{x}}:i\neq i^*} D_{q}(s_i - x_i)\Big] e^{(1+\nu)\epsilon} D_{q}(s_{i^*} - x_{i^*})\,\mathrm{d}\boldsymbol{s} \\
&\qquad\qquad \text{(Using Claim 5.1 and that } x_i = x'_i, \forall i \neq i^*) \\
&= e^{(1+\nu)\epsilon}\int_{S_{-1}} \Big[\prod_{i\in\mathcal{G}_{\boldsymbol{x}}} D_{q}(s_i - x_i)\Big]\,\mathrm{d}\boldsymbol{s} \\
&= e^{(1+\nu)\epsilon}\Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}) \in S_{-1}]
\end{aligned}$$

$\square$

Claim (Restating Claim 5.3)**.** $\Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}') \in S_{1}] \leq \frac{e^{\epsilon}-1}{2(e^{\epsilon q/2}-1)}$.

Proof. Observe that, for every $\boldsymbol{s} \in S_1$, we have $-q' \leq s_{i^*} - x'_{i^*} < -q' + (1-\tau)$. Recall that $\mathcal{G}_{\boldsymbol{x}'} = \text{supp}(\boldsymbol{x}')$ and $|\boldsymbol{x}'| = n$. Let $|\mathcal{G}_{\boldsymbol{x}'}| = t$ for some $t \leq n$, and, for simplicity, assume that $\mathcal{G}_{\boldsymbol{x}'} = \{1, 2, \ldots, t\}$. For $i \in [t]$, define $S_1(i) := \{\hat{s}_i : \exists \boldsymbol{s} \in S_1 \text{ s.t. } \hat{s}_i = s_i\}$.

$$\begin{aligned}
\Pr[\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}(\boldsymbol{x}') \in S_{1}] &= \int_{S_1} \Big[\prod_{i=1}^{t} D_{q'}(s_i - x'_i)\Big]\,\mathrm{d}\boldsymbol{s} \\
&= \int_{S_1(1)} \cdots \int_{S_1(i^*)} \cdots \int_{S_1(t)} \Big[\prod_{i=1}^{t} D_{q'}(s_i - x'_i)\Big]\,\mathrm{d}s_t \ldots \mathrm{d}s_{i^*} \ldots \mathrm{d}s_1 \\
&= \int_{S_1(i^*)} D_{q'}(s_{i^*} - x'_{i^*}) \underbrace{\left(\int_{S_1(1)} \cdots \int_{S_1(t)} \Big[\prod_{i=1:i\neq i^*}^{t} D_{q'}(s_i - x'_i)\Big]\,\mathrm{d}s_t \ldots \mathrm{d}s_1\right)}_{\leq 1} \mathrm{d}s_{i^*} \\
&\leq \int_{S_1(i^*)} D_{q'}(s_{i^*} - x'_{i^*})\,\mathrm{d}s_{i^*} \\
&= \int_{q'}^{q'+(1-\tau)} D_{q'}(z)\,\mathrm{d}z \qquad\qquad (\forall \boldsymbol{s} \in S_1,\ (s_{i^*} - x'_{i^*}) \in [-q', -q' + (1-\tau)))
\end{aligned}$$

$$= \frac{e^{(1-\tau)\epsilon} - 1}{2(1 - e^{-\epsilon q/2})} e^{-\epsilon q/2}$$
$$\leq \frac{e^{\epsilon} - 1}{2(e^{\epsilon q/2} - 1)}. \qquad \text{(Since } \tau > 0\text{)}$$

□

Theorem (Restating Theorem 5.1)**.** *On histograms of size at least n, i.e.,* $|\boldsymbol{x}| > n$*,* $\mathcal{M}^{\tau,\mathcal{G}}_{\text{trLap}}$ *from Construction 1 satisfies the following guarantees:*

- $\mathcal{M}^{\tau,\mathcal{G}}_{\text{trLap}}$ *is* $\left(O(\frac{1}{\sqrt{\tau n}}), e^{-\Omega(\sqrt{\tau n})}\right)$*-differentially private w.r.t.* $\sim_{\text{hist}}$*, and*
- *If* $|\text{supp}(\boldsymbol{x})| \leq t$*, then* $\mathcal{M}^{\tau,\mathcal{G}}_{\text{trLap}}$ *is* $(\tau t, 0, 0)$*-accurate for the identity function, w.r.t. the distortion measure* ∂_{drop}*.*

Proof. First we show the privacy part and then show the accuracy part. Note that the requirement of $|\text{supp}(\boldsymbol{x})| \leq t$ is only needed for accuracy.

- **Differential privacy.** We use Lemma 5.1 and put a restriction that ν should be > 0. We will analyze the effect of this restriction on the bound of $|\boldsymbol{x}|$. We restate the bound on $|\boldsymbol{x}|$ here again for convenience:

$$|\boldsymbol{x}| \geq \frac{2}{\epsilon\tau} \ln\left(1 + \frac{1 - e^{-\epsilon\frac{\tau}{2}}}{e^{\epsilon(\nu+\frac{\tau}{2})} - 1}\right)$$

It can be easily checked that for any fixed $\epsilon, \nu > 0$, the RHS is a decreasing function of τ. Hence, if we set τ to its minimum value, we get a lower bound on $|\boldsymbol{x}|$ which is independent of τ. Since this expression is not defined at $\tau = 0$, we will take its one-sided limit as $\tau \to 0^+$, i.e.,

$$\lim_{\tau\to0^+} \frac{2}{\epsilon\tau} \ln\left(1 + \frac{1 - e^{-\epsilon\frac{\tau}{2}}}{e^{\epsilon(\nu+\frac{\tau}{2})} - 1}\right)$$

We will replace $\frac{\epsilon\tau}{2}$ with l. As $\tau \to 0^+$, $l \to 0^+$, and we get

$$\begin{aligned}
\lim_{\tau\to0^+} \frac{2}{\epsilon\tau} \ln\left(1 + \frac{1 - e^{-\epsilon\frac{\tau}{2}}}{e^{\epsilon(\nu+\frac{\tau}{2})} - 1}\right) &= \lim_{l\to0^+} \frac{1}{l} \ln\left(1 + \frac{1 - e^{-l}}{e^{\epsilon\nu+l} - 1}\right) \\
&= \lim_{l\to0^+} \frac{1}{l} \ln\left(1 + \frac{1 - e^{-l}}{e^{\epsilon\nu+l} - 1}\right)\left(\frac{1 - e^{-l}}{e^{\epsilon\nu+l} - 1}\right)\left(\frac{e^{\epsilon\nu+l} - 1}{1 - e^{-l}}\right) \\
&= \lim_{l\to0^+} \left(\frac{1}{e^{\epsilon\nu+l} - 1}\right)\left(\frac{1 - e^{-l}}{l}\right)\left(\frac{\ln\left(1 + \frac{1-e^{-l}}{e^{\epsilon\nu+l}-1}\right)}{\frac{1-e^{-l}}{e^{\epsilon\nu+l}-1}}\right) \\
&= \frac{1}{e^{\epsilon\nu} - 1} \qquad (\lim_{x\to0^+} \tfrac{1-e^{-x}}{x} = 1;\ \lim_{x\to0^+} \tfrac{\ln(1+x)}{x} = 1)
\end{aligned}$$

We have proved that on inputs $\boldsymbol{x}$ s.t. $|\boldsymbol{x}| > \frac{1}{e^{\epsilon\nu}-1}$ and $\nu > 0$, $\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\text{trLap}}$ is $\left((1+\nu)\epsilon, \frac{e^{\epsilon}-1}{2(e^{\epsilon n\tau/2}-1)}\right)$-DP w.r.t. $\sim_{\text{hist}}$. However, we are given $|\boldsymbol{x}| > n$, not $|\boldsymbol{x}| > \frac{1}{e^{\epsilon\nu}-1}$. So, in order to prove the result, we will set the values of ϵ, ν such that $\frac{1}{e^{\epsilon\nu}-1} \leq n$ holds. We can take any ϵ, ν that satisfy $\epsilon\nu \geq \ln\left(1 + \frac{1}{n}\right)$. In particular, we can take $\epsilon = O(\frac{1}{\sqrt{\tau n}})$ and $\nu = 1$. This would imply that $\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\text{trLap}}$ is $\left(O(\frac{1}{\sqrt{\tau n}}), e^{-\Omega(\sqrt{\tau n})}\right)$-differentially private.
- **Flexible accuracy.** Note that noise added by $\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\text{trLap}}$ in each bar of the histogram is bounded by $-\tau|\boldsymbol{x}|$ which can lead to a drop of at most τ fraction of elements. Combined with the fact that $|\text{supp}(\boldsymbol{x})| \leq t$, the maximum number of elements that can be dropped are τt. Hence, $\mathcal{M}^{\tau,\mathcal{G}}_{\text{trLap}}$ is $(\tau t, 0, 0)$-accurate.

This completes the proof of Theorem 5.1. $\square$

Theorem (Restating Theorem 5.2)**.** *On inputs* $\boldsymbol{x}$ *s.t.* $|\boldsymbol{x}| \geq \frac{2}{\epsilon\tau} e^{-\frac{\epsilon\tau}{2}}$, $\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}$ *is* $\left(\epsilon, \frac{e^{\epsilon}-1}{2(e^{\epsilon n\tau/2}-1)}\right)$*-DP w.r.t.* $\sim_{\text{hist}}$.

Proof. Substituting $\nu = 0$ in Lemma 5.1 gives that when $\boldsymbol{x}$ satisfies $|\boldsymbol{x}| \geq \frac{2}{\epsilon\tau} \ln\left(1 + \frac{1-e^{-\frac{\epsilon\tau}{2}}}{e^{\frac{\epsilon\tau}{2}}-1}\right)$, we have that $\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}$ is $\left(\epsilon, \frac{e^{\epsilon}-1}{2(e^{\epsilon n\tau/2}-1)}\right)$-DP w.r.t. $\sim_{\text{hist}}$. Now, the corollary follows because

$$\frac{2}{\epsilon\tau} e^{-\frac{\epsilon\tau}{2}} \geq \frac{2}{\epsilon\tau} \ln\left(1 + e^{-\frac{\epsilon\tau}{2}}\right) = \frac{2}{\epsilon\tau} \ln\left(1 + \frac{1 - e^{-\epsilon\frac{\tau}{2}}}{e^{\epsilon\frac{\tau}{2}} - 1}\right),$$

where the first inequality uses $x \geq \ln(1+x)$. $\square$

Claim (Restating Claim 5.4)**.** $\mathcal{M}_{\text{buc}}^{t,[0,B)^d}$ *is a neighborhood-preserving mechanism, i.e., for any two histograms* $\boldsymbol{x}, \boldsymbol{x}'$ *such that* $\boldsymbol{x} \sim_{\text{hist}} \boldsymbol{x}'$, *we have* $\mathcal{M}_{\text{buc}}^{t,[0,B)^d}(\boldsymbol{x}) \sim_{\text{hist}} \mathcal{M}_{\text{buc}}^{t,[0,B)^d}(\boldsymbol{x}')$.

Proof. Removing any one element changes the output of bucketing by at most one element. Hence, neighbours remain neighbours after bucketing. $\square$

Claim (Restating Claim 5.5)**.** *On all inputs* $\boldsymbol{x}$, $\mathcal{M}_{\text{buc}}^{t,[0,B)^d}$ *is* $\left(0, \frac{B\sqrt{d}}{2t^{\frac{1}{d}}}, 0\right)$*-accurate for the identity function w.r.t the metric* $\mathfrak{d}_{\text{hist}}$ *and any measure of distortion.*

Proof. We are using $\mathfrak{d}_{\text{hist}}$ as the metric, which is basically the Wasserstein distance between histograms treated as probability distributions after normalization. Since the target function is the identity function, the accuracy is essentially the Wasserstein distance between the input and output histograms. Since every point in the input histogram is moved to the center of its bucket, which means that the maximum distance any point can move is the maximum distance of the center of a bucket from any point in that bucket, which is $\frac{B\sqrt{d}}{2t^{\frac{1}{d}}}$. $\square$

Claim (Restating Claim 5.6)**.** *The distortion sensitivity of the histogram identity function w.r.t.* $(\partial_{\text{drop}}, \partial_{\text{drop}})$ *at* $\theta = \omega = 0$ *is the identity function.*

Proof. Setting the function f in Definition 11 as the identity function and putting $\theta = \omega = 0$, we get

$$\begin{aligned}
\sigma_f(\alpha) &= \sup_{x,Y:\widehat{\partial}_{\text{drop}}(x,Y)\leq\alpha} \inf_{X:X=Y} \widehat{\partial}_{\text{drop}}(x,X) \\
&= \sup_{x,Y:\widehat{\partial}_{\text{drop}}(x,Y)\leq\alpha} \widehat{\partial}_{\text{drop}}(x,Y) \\
&= \alpha
\end{aligned}$$

$\square$

Claim (Restating Claim 5.7)**.** *For any* $\beta \in \mathbb{R}$ *and on inputs restricted to* t *bars, we have* $\tau^{\alpha,0}_{\mathcal{M}_{\text{trLap}}^{\tau,\epsilon,\mathcal{G}}, f_{id}}(\beta, 0) \leq \beta$ *w.r.t.* ∂_{drop} *where* $\alpha = \tau t$ *and* f_{id} *is the identity function.*

Proof. Consider any two histograms $\boldsymbol{x}$, $\boldsymbol{y}$ over $\mathcal{G}$ such that $\mathfrak{d}_{\text{hist}}(\boldsymbol{x}, \boldsymbol{y}) \leq \beta$. Let $\mathfrak{d}_{\mathcal{G}}(\cdot,\cdot)$ be the underlying metric over $\mathcal{G}$(consists of t elements) and $|\boldsymbol{x}|$ denote number of elements in the histogram $\boldsymbol{x}$. By definition of $\mathfrak{d}_{\text{hist}}(\cdot,\cdot)$, we have $\mathfrak{d}_{\text{hist}}(\boldsymbol{x}, \boldsymbol{y}) = W^{\infty}(\mathsf{norm}(\boldsymbol{x}), \mathsf{norm}(\boldsymbol{y}))$. Let ϕ be an optimal coupling of $\mathsf{norm}(\boldsymbol{x})$ and $\mathsf{norm}(\boldsymbol{y})$ such that

$$\sup_{(a,b)\leftarrow\phi} \mathfrak{d}_{\mathcal{G}}(a,b) = \beta. \tag{33}$$

From a coupling ϕ, we define transform f_ϕ as, given a histogram $\boldsymbol{z}$ that is α-distorted from $\boldsymbol{x}$, returns $f_\phi(\boldsymbol{z})$, an α-distorted histogram from $\boldsymbol{y}$ as below. Recall that for a histogram $\boldsymbol{x}$ and $a \in \mathcal{G}$, we denote by $\boldsymbol{x}(a)$ the multiplicity of a in $\boldsymbol{x}$.

$$f_\phi(\boldsymbol{z})(b) := |\boldsymbol{y}| \sum_{a \in \mathcal{G}} \frac{\boldsymbol{z}(a)\phi(a,b)}{\boldsymbol{x}(a)}.$$

To see that $f_\phi(\boldsymbol{z})$ is α-distorted from $\boldsymbol{y}$, we need to show two things: (i) $f_\phi(\boldsymbol{z})(b) \leq \boldsymbol{y}(b)$ holds for every $b \in \mathcal{G}$, and (ii) $\sum_{b \in \mathcal{G}} f_\phi(\boldsymbol{z})(b) = (1-\alpha)\sum_{b \in \mathcal{G}} \boldsymbol{y}(b)$. The first condition holds because $\boldsymbol{z}(a) \leq \boldsymbol{x}(a), \forall a \in \mathcal{G}$ and $\sum_{a \in \mathcal{G}} \phi(a,b) = \frac{\boldsymbol{y}(b)}{|\boldsymbol{y}|}$. For the second condition,

$$\begin{aligned}
\sum_{b \in \mathcal{G}} f_\phi(\boldsymbol{z})(b) &= \sum_{b \in \mathcal{G}} |\boldsymbol{y}| \sum_{a \in \mathcal{G}} \frac{\boldsymbol{z}(a)\phi(a,b)}{\boldsymbol{x}(a)} \\
&= |\boldsymbol{y}| \sum_{a \in \mathcal{G}} \frac{\boldsymbol{z}(a)}{\boldsymbol{x}(a)} \phi(a) && \left(\text{where } \phi(a) := \textstyle\sum_{b \in \mathcal{G}} \phi(a,b)\right) \\
&= |\boldsymbol{y}| \sum_{a \in \mathcal{G}} \frac{\boldsymbol{z}(a)}{\boldsymbol{x}(a)} \frac{\boldsymbol{x}(a)}{|\boldsymbol{x}|} && \left(\text{Since } \phi(a) = \tfrac{\boldsymbol{x}(a)}{|\boldsymbol{x}|}\right) \\
&= \frac{|\boldsymbol{y}|}{|\boldsymbol{x}|} \sum_{a \in \mathcal{G}} \boldsymbol{z}(a) \\
&= (1-\alpha)|\boldsymbol{y}| && \left(\text{Since } \textstyle\sum_{a \in \mathcal{G}} \boldsymbol{z}(a) = |\boldsymbol{z}| = (1-\alpha)|\boldsymbol{x}|\right)
\end{aligned}$$

Therfore, $f_\phi(\boldsymbol{z})$ is α-distorted from $\boldsymbol{y}$.

Infact, the transform f_ϕ doesn't amplify $\mathfrak{d}_{\text{hist}}(,)$ metric. To see this, define $\phi'(a,b) = \frac{\boldsymbol{z}(a)|\boldsymbol{x}|\phi(a,b)}{\boldsymbol{x}(a)|\boldsymbol{z}|}$. It can be easily verified that the first marginal, $\sum_{b \in \mathcal{G}} \phi'(a,b) = \frac{\boldsymbol{z}(a)}{|\boldsymbol{z}|}$ for all $a \in \mathcal{G}$ and the second marginal, $\sum_{a \in \mathcal{G}} \phi'(a,b) = \frac{f_\phi(\boldsymbol{z})(b)}{|f_\phi(\boldsymbol{z})|}$ for all $b \in \mathcal{G}$. This means that $\phi'(a,b)$ is a valid coupling of $\boldsymbol{z}, f_\phi(\boldsymbol{z})$. Now, by definition of ϕ', we have that $\text{support}(\phi') \subseteq \text{support}(\phi)$, and hence, $\mathfrak{d}_{\text{hist}}(\boldsymbol{z}, f_\phi(\boldsymbol{z})) \leq \mathfrak{d}_{\text{hist}}(\boldsymbol{x}, \boldsymbol{y}) = \beta$.

Finally, in order to prove the required result, with f_{phi} defined as above, consider the random variable $Y = f_\phi\left(\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\text{trLap}}(\boldsymbol{x})\right)$, i.e.,

$$\Pr(Y = f_\phi(\boldsymbol{z})) = \Pr(\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\text{trLap}}(\boldsymbol{x}) = \boldsymbol{z}).$$

The support of $\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\text{trLap}}(\boldsymbol{x})$ consists of α-distorted histograms from $\boldsymbol{x}$ where $\alpha = \tau t$ (see Theorem 5.1). From previous arguments, this implies that Y is also a distribution over α-distorted histograms from $\boldsymbol{y}$. Therefore, $W^\infty(Y, \mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\text{trLap}}(\boldsymbol{x})) \leq \beta$. Since this holds for any two histograms $\boldsymbol{x}, \boldsymbol{y}$, we have,

$$\sup_{\substack{\boldsymbol{x},\boldsymbol{y}: \\ \mathfrak{d}_{\text{hist}}(\boldsymbol{x},\boldsymbol{y}) \leq \beta}} \inf_{\substack{Y: \\ \widehat{\partial}(\boldsymbol{y},Y) \leq \alpha}} W^\infty(Y, \mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\text{trLap}}) \leq \beta$$

We can extend this construction of Y for distributions X, X' on histograms with the same guarantee. Therefore,

$$\tau^{\alpha,0}_{\mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\text{trLap}}}(\beta, 0) = \sup_{\substack{X,X': \\ W^\infty(X,X') \leq \beta}} \inf_{\substack{Y: \\ \widehat{\partial}(X',Y) \leq \alpha}} W^\infty(Y, \mathcal{M}^{\tau,\epsilon,\mathcal{G}}_{\text{trLap}}) \leq \beta.$$

This completes the proof of Claim 5.7. □

D Omitted Details from Section 5.1.4 (Histogram-Based-Statistics)

Lemma (Restating Lemma 5.2)**.** *Let $f : \mathcal{A} \to \mathcal{B}$ be a deterministic function with $\mathfrak{d}_\mathcal{A}, \mathfrak{d}_\mathcal{B}$ metrics defined in $\mathcal{A}, \mathcal{B}$, respectively. Then we have,*

$$\sup_{\substack{X,X': \\ W^\infty(X,X') \leq \beta}} W^\infty(f(X), f(X')) = \sup_{\substack{\boldsymbol{y},\boldsymbol{y}' \in \mathcal{A}: \\ \mathfrak{d}_\mathcal{A}(\boldsymbol{y},\boldsymbol{y}') \leq \beta}} \mathfrak{d}_\mathcal{B}(f(\boldsymbol{y}), f(\boldsymbol{y}'))$$

Proof. We first prove that L.H.S. $\geq$ R.H.S.:

$$\sup_{\substack{X,X':\\ W^\infty(X,X')\leq\beta}} W^\infty(f(X),f(X')) \geq \sup_{\substack{x,x':\\ W^\infty(x,x')\leq\beta}} W^\infty(f(x),f(x'))$$

(considering only point distributions restricts the set of supremum)

$$= \sup_{\substack{x,x':\\ \mathfrak{d}_{\mathcal{A}}(x,x')\leq\beta}} \mathfrak{d}_{\mathcal{B}}(f(x),f(x'))$$

(Wasserstein distance for point distributions is just the distance between the points)

$$= \sup_{\substack{y,y':\\ \mathfrak{d}_{\mathcal{A}}(y,y')\leq\beta}} \mathfrak{d}_{\mathcal{B}}(f(y),f(y')) \qquad \text{(renaming)}$$

Now we prove the other direction. Consider two distributions X, X' over $\mathcal{A}$ s.t, $W^\infty(X,X') \leq \beta$. Let ϕ_1 be the optimal coupling between X, X' for which this is true. Therefore, we have

$$\sup_{(\boldsymbol{x},\boldsymbol{x}')\leftarrow\phi_1} \mathfrak{d}_{\mathcal{A}}(\boldsymbol{x},\boldsymbol{x}') \leq \beta.$$

We can define a coupling ϕ_2 between $f(X), f(X')$ as follows,

$$\Pr[\phi_2(\mathbf{a},\mathbf{b})] = \sum_{\substack{\boldsymbol{x},\boldsymbol{x}':\\ f(\boldsymbol{x})=\mathbf{a},f(\boldsymbol{x}')=\mathbf{b}}} \Pr[\phi_1(\boldsymbol{x},\boldsymbol{x}')]$$

It can be verified that ϕ_2 is indeed a valid coupling between $f(X), f(X')$. Now

$$\begin{aligned}
W^\infty(f(X),f(X')) &= \inf_{\phi\in\Phi(f(X),f(X'))} \sup_{(\mathbf{a},\mathbf{b})\leftarrow\phi} \mathfrak{d}_{\mathcal{B}}(\mathbf{a},\mathbf{b})\\
&\leq \sup_{(\mathbf{a},\mathbf{b})\leftarrow\phi_2} \mathfrak{d}_{\mathcal{B}}(\mathbf{a},\mathbf{b})\\
&= \sup_{(\boldsymbol{x},\boldsymbol{x}')\leftarrow\phi_1} \mathfrak{d}_{\mathcal{B}}(f(\boldsymbol{x}),f(\boldsymbol{x}')))\\
&\leq \sup_{\substack{\boldsymbol{x},\boldsymbol{x}'\in\mathcal{A}:\\ \mathfrak{d}_{\mathcal{A}}(\boldsymbol{x},\boldsymbol{x}')\leq\beta}} \mathfrak{d}_{\mathcal{B}}(f(\boldsymbol{x}),f(\boldsymbol{x}')))
\end{aligned}$$

Note that the RHS of the last inequality does not depend on X, X'. So, taking supremum over all distributions X, X' such that $W^\infty(X,X') \leq \beta$ gives the required result. $\square$

E Omitted Details from Section 5.1.5 (Beyond ∂_{drop})

We introduced two new distortions: $\partial_{\text{mv}}(\boldsymbol{x},\boldsymbol{y})$ and $\partial^\eta_{\text{drmv}}(\boldsymbol{x},\boldsymbol{y})$. We first prove the following results before giving the applications. We present the definitions again here for convenience:

$$\partial_{\text{mv}}(\boldsymbol{x},\boldsymbol{y}) = \begin{cases} W^\infty(\frac{\boldsymbol{x}}{|\boldsymbol{x}|},\frac{\boldsymbol{y}}{|\boldsymbol{y}|}) & \text{if } |\boldsymbol{x}|=|\boldsymbol{y}|\\ \infty & \text{otherwise}\end{cases} \qquad \partial^\eta_{\text{drmv}}(\boldsymbol{x},\boldsymbol{y}) = \inf_{\boldsymbol{z}}\left(\partial_{\text{drop}}(\boldsymbol{x},\boldsymbol{z}) + \eta\cdot\partial_{\text{mv}}(\boldsymbol{z},\boldsymbol{y})\right).$$

Note that in the definition of ∂_{mv}, when $|\boldsymbol{x}| = |\boldsymbol{y}| = 0$, we define $\partial_{\text{mv}}(\boldsymbol{x},\boldsymbol{y}) = 0$.

Claim E.1. *$\partial_{\text{mv}}(\cdot,\cdot)$ is a metric.*

Proof. Since $\partial_{\mathrm{mv}}(\cdot,\cdot)$ is defined as the Wasserstein distance between normalized histograms, we can conclude that $\partial_{\mathrm{mv}}(\cdot,\cdot)$ is a metric from the fact that Wasserstein distance is a metric. Note that when $|\boldsymbol{x}| = |\boldsymbol{y}| = 0$, the Wasserstein distance is undefined, but we have defined $\partial_{\mathrm{mv}}(\boldsymbol{x},\boldsymbol{y})$ in this case separately as 0 which is consistent with the properties of a metric. □

We first give an intermediate result which will be used in proving that $\partial^{\eta}_{\mathrm{drmv}}$ is a quasi-metric.

Claim E.2. *Let $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ be any three histograms over a ground set $\mathcal{G}$,associated with a metric $\mathfrak{d}$, such that $\partial_{\mathrm{mv}}(\boldsymbol{x},\boldsymbol{z}) = \alpha_1$ and $\partial_{\mathrm{drop}}(\boldsymbol{z},\boldsymbol{y}) = \alpha_2$ with $\alpha_1 \geq 0$ and $\alpha_2 < 1$. Then there exists a histogram $\boldsymbol{s}$ such that $\partial_{\mathrm{drop}}(\boldsymbol{x},\boldsymbol{s}) = \alpha_2$ and $\partial_{\mathrm{mv}}(\boldsymbol{s},\boldsymbol{y}) \leq \alpha_1$.*

Proof. Using the definitions of ∂_{drop} and ∂_{mv}, we have the following:

Z.1 $|\boldsymbol{x}| = |\boldsymbol{z}|$

Z.2 $W^{\infty}(\frac{\boldsymbol{x}}{|\boldsymbol{x}|}, \frac{\boldsymbol{z}}{|\boldsymbol{z}|}) \leq \alpha_1$. We will use ϕ_z to denote the optimal joint distribution which achieves the infimum in the definition of $W^{\infty}(\frac{\boldsymbol{x}}{|\boldsymbol{x}|}, \frac{\boldsymbol{z}}{|\boldsymbol{z}|})$.

Z.3 $|\boldsymbol{y}| = (1-\alpha_2)|\boldsymbol{z}|$

Z.4 For all $g \in \mathcal{G}$, $0 \leq \boldsymbol{y}(g) \leq \boldsymbol{z}(g)$

Now we want to prove the existence of a histogram $\boldsymbol{s}$ with the following property:

S.1 $|\boldsymbol{s}| = (1-\alpha_2)|\boldsymbol{x}|$

S.2 For all $g \in \mathcal{G}$, $0 \leq \boldsymbol{s}(g) \leq \boldsymbol{x}(g)$

S.3 $|\boldsymbol{s}| = |\boldsymbol{y}|$

S.4 $W^{\infty}(\frac{\boldsymbol{s}}{|\boldsymbol{s}|}, \frac{\boldsymbol{y}}{|\boldsymbol{y}|}) \leq \alpha_1$.

Consider the following joint distribution ϕ_s:

$$\phi_s(g_x,g_y) = \begin{cases} \frac{1}{1-\alpha_2}\phi_z(g_x,g_y)\frac{\boldsymbol{y}(g_y)}{\boldsymbol{z}(g_y)} & \text{if } \boldsymbol{z}(g_y) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

We denote the first marginal of ϕ_s by $\frac{\boldsymbol{s}}{|\boldsymbol{s}|}$, where $\boldsymbol{s}$ corresponds to the histogram that we want to show exists.

By definition, for all $g_x, g_y \in \mathcal{G}$, we have $\phi_s(g_x,g_y) \geq 0$. Also note that, if $\boldsymbol{z}(g_y) = 0$, then for all $g_x \in \mathcal{G}$, we have $\phi_z(g_x,g_y) = 0$; this is because $\frac{\boldsymbol{z}}{|\boldsymbol{z}|}$ is the second marginal of ϕ_z. Now we show that the above-defined ϕ_s satisfies properties **S.1**-**S.4** – we show these in the sequence of **S.4**, **S.3**, **S.1**, **S.2**.

- **Proof of S.4.** Note that the first marginal of ϕ_s is assumed to be $\frac{\boldsymbol{s}}{|\boldsymbol{s}|}$. Now we show that its second marginal is $\frac{\boldsymbol{y}}{|\boldsymbol{y}|}$ and that $\max_{(g_x,g_y)\leftarrow\phi_s}\mathfrak{d}(g_x,g_y) \leq \alpha_1$. Note that these together imply that $W^{\infty}(\frac{\boldsymbol{s}}{|\boldsymbol{s}|}, \frac{\boldsymbol{y}}{|\boldsymbol{y}|}) \leq \alpha_1$.
  - *Second marginal of ϕ_s is $\frac{\boldsymbol{y}}{|\boldsymbol{y}|}$:* We show it in two parts, first for $g_y \in \mathcal{G}$ for which $\boldsymbol{z}(g_y) = 0$ and then for the rest of the $g_y \in \mathcal{G}$. Note that when $\boldsymbol{z}(g_y) = 0$, we have from **Z.4** that $\boldsymbol{y}(g_y) = 0$. Now we show that $\int_{\mathcal{G}} \phi_s(g_x,g_y)\,\mathrm{d}g_x = 0$. It follows from (34) that for all g_y such that $\boldsymbol{z}(g_y) = 0$, we have $\phi_s(g_x,g_y) = 0, \forall g_x \in \mathcal{G}$, which implies that $\int_{\mathcal{G}} \phi_s(g_x,g_y)\,\mathrm{d}g_x = 0$. Now we analyze the case when $\boldsymbol{z}(g_y) > 0$.

$$\begin{aligned}
\int_{\mathcal{G}} \phi_s(g_x,g_y)\,\mathrm{d}g_x &= \int_{\mathcal{G}} \frac{1}{1-\alpha_2}\phi_z(g_x,g_y)\frac{\boldsymbol{y}(g_y)}{\boldsymbol{z}(g_y)}\,\mathrm{d}g_x && \text{(using (34))} \\
&= \frac{1}{1-\alpha_2}\frac{\boldsymbol{y}(g_y)}{\boldsymbol{z}(g_y)}\int_{\mathcal{G}} \phi_z(g_x,g_y)\,\mathrm{d}g_x \\
&= \frac{1}{1-\alpha_2}\frac{\boldsymbol{y}(g_y)}{\boldsymbol{z}(g_y)}\frac{\boldsymbol{z}(g_y)}{|\boldsymbol{z}|} && \text{(using Z.2)} \\
&= \frac{\boldsymbol{y}(g_y)}{(1-\alpha_2)|\boldsymbol{z}|} \\
&= \frac{\boldsymbol{y}(g_y)}{|\boldsymbol{y}|}. && \text{(Using Z.3)}
\end{aligned}$$

– $W^\infty(\frac{\boldsymbol{s}}{|\boldsymbol{s}|}, \frac{\boldsymbol{y}}{|\boldsymbol{y}|}) \leq \alpha_1$: We have shown that the first and the second marginals of ϕ_s are $\frac{\boldsymbol{s}}{|\boldsymbol{s}|}$ and $\frac{\boldsymbol{y}}{|\boldsymbol{y}|}$, respectively. So, it suffices to show that $\max_{(g_x,g_y)\leftarrow\phi_s} \mathfrak{d}(g_x, g_y) \leq \alpha_1$. Consider any pair $(g_x, g_y) \in \mathcal{G}^2$ s.t. $\phi_s(g_x, g_y) > 0$. This is possible only if $\phi_z(g_x, g_y) > 0$ (see (34)), which, when combined with **Z.2**, gives $\mathfrak{d}(g_x, g_y) \leq \alpha_1$. Hence, for any pair $(g_x, g_y) \in \mathcal{G}^2$ s.t. $\phi_s(g_x, g_y) > 0$, we have $\mathfrak{d}(g_x, g_y) \leq \alpha_1$.

- **Proof of S.3.** Note that (34) gives the normalized $\boldsymbol{s}$, but we still have the freedom to choose $|\boldsymbol{s}|$. To satisfy **S.3**, we set $|\boldsymbol{s}| = |\boldsymbol{y}|$.
- **Proof of S.1.** Note that **S.1** is already satisfied using **Z.1**, **Z.3**, and **S.3**.
- **Proof of S.2.** Let us denote $\{g \in \mathcal{G} \mid \boldsymbol{z}(g) > 0\}$ by $\mathcal{G}_z$. We will show that for any $g \in \mathcal{G}$, we have $\boldsymbol{x}(g) - \boldsymbol{s}(g) \geq 0$:

$$\begin{aligned}
\boldsymbol{x}(g) - \boldsymbol{s}(g) &= |\boldsymbol{x}| \int_{\mathcal{G}} \phi_z(g, g_y)\,\mathrm{d}g_y - |\boldsymbol{s}| \int_{\mathcal{G}} \phi_s(g, g_y)\,\mathrm{d}g_y && \text{(using Z.2 and S.4)} \\
&= |\boldsymbol{x}| \int_{\mathcal{G}_z} \phi_z(g, g_y)\,\mathrm{d}g_y - |\boldsymbol{s}| \int_{\mathcal{G}} \phi_s(g, g_y)\,\mathrm{d}g_y && \text{(Since } \boldsymbol{z}(g_y) = 0 \Rightarrow \phi_z(g, g_y) = 0, \forall g \in \mathcal{G}; \text{ Z.2)} \\
&= |\boldsymbol{x}| \int_{\mathcal{G}_z} \phi_z(g, g_y)\,\mathrm{d}g_y - |\boldsymbol{s}| \int_{\mathcal{G}_z} \frac{1}{1-\alpha_2} \phi_z(g, g_y) \frac{\boldsymbol{y}(g_y)}{\boldsymbol{z}(g_y)}\,\mathrm{d}g_y && \text{(using (34))} \\
&= |\boldsymbol{x}| \int_{\mathcal{G}_z} \phi_z(g, g_y)\,\mathrm{d}g_y - \frac{(1-\alpha_2)|\boldsymbol{x}|}{1-\alpha_2} \int_{\mathcal{G}_z} \phi_z(g, g_y) \frac{\boldsymbol{y}(g_y)}{\boldsymbol{z}(g_y)}\,\mathrm{d}g_y && \text{(using S.1)} \\
&= |\boldsymbol{x}| \int_{\mathcal{G}_z} \phi_z(g, g_y)\,\mathrm{d}g_y - |\boldsymbol{x}| \int_{\mathcal{G}_z} \phi_z(g, g_y) \frac{\boldsymbol{y}(g_y)}{\boldsymbol{z}(g_y)}\,\mathrm{d}g_y \\
&= |\boldsymbol{x}| \int_{\mathcal{G}_z} \phi_z(g, g_y) \left(1 - \frac{\boldsymbol{y}(g_y)}{\boldsymbol{z}(g_y)}\right) \mathrm{d}g_y \\
&\geq 0 && \text{(using Z.4, } \tfrac{\boldsymbol{y}(g_y)}{\boldsymbol{z}(g_y)} \leq 1\text{)}
\end{aligned}$$

Thus, we have shown that the joint distribution ϕ_s defined in (34) satisfies all four properties **S.1**-**S.4**. This completes the proof of Claim E.2. ☐

Claim E.3. *For all $\eta \in \mathbb{R}_{\geq 0}$, $\partial^\eta_{\mathrm{drmv}}(\cdot,\cdot)$ is a quasi metric.*

Proof. Note that both ∂_{drop} and ∂_{mv} are quasi-metrics. Hence, for any $\boldsymbol{x}, \boldsymbol{y}$, $\partial_{\mathrm{drop}}(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ and $\partial_{\mathrm{mv}}(\boldsymbol{x}, \boldsymbol{y}) \geq 0$. This implies that for every $\boldsymbol{x}, \boldsymbol{y}$, $\partial^\eta_{\mathrm{drmv}}(\boldsymbol{x}, \boldsymbol{y}) \geq 0$. Now we one by one prove that $\partial^\eta_{\mathrm{drmv}}$ satisfies the properties of quasi-metric:

Property #1: For all $\boldsymbol{x}$ and $\boldsymbol{y}$, $\boldsymbol{x} = \boldsymbol{y} \Leftrightarrow \partial^\eta_{\mathrm{drmv}}(\boldsymbol{x}, \boldsymbol{y}) = 0$.

1. For all $\boldsymbol{x}$, $\partial^\eta_{\mathrm{drmv}}(\boldsymbol{x}, \boldsymbol{x}) = 0$:

$$\begin{aligned}
\partial^\eta_{\mathrm{drmv}}(\boldsymbol{x}, \boldsymbol{x}) &= \inf_{\boldsymbol{z}} \left(\partial_{\mathrm{drop}}(\boldsymbol{x}, \boldsymbol{z}) + \eta \cdot \partial_{\mathrm{mv}}(\boldsymbol{z}, \boldsymbol{x})\right) \\
&\leq \partial_{\mathrm{drop}}(\boldsymbol{x}, \boldsymbol{x}) + \eta \cdot \partial_{\mathrm{mv}}(\boldsymbol{x}, \boldsymbol{x}) \quad \text{(infimum over a set is} \leq \text{the value at any fixed point in set)} \\
&= 0
\end{aligned}$$

   Since $\partial^\eta_{\mathrm{drmv}}(\boldsymbol{x}, \boldsymbol{x}) \geq 0$ as well as ≤ 0, $\partial^\eta_{\mathrm{drmv}}(\boldsymbol{x}, \boldsymbol{x}) = 0$.
2. For all $\boldsymbol{x}, \boldsymbol{y}$, $\partial^\eta_{\mathrm{drmv}}(\boldsymbol{x}, \boldsymbol{y}) = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{y}$:
   $\partial^\eta_{\mathrm{drmv}}(\boldsymbol{x}, \boldsymbol{y}) = 0$ implies that $\inf_{\boldsymbol{z}} \left(\partial_{\mathrm{drop}}(\boldsymbol{x}, \boldsymbol{z}) + \eta \cdot \partial_{\mathrm{mv}}(\boldsymbol{z}, \boldsymbol{y})\right) = 0$. As both $\partial_{\mathrm{drop}}(\boldsymbol{x}, \boldsymbol{z})$ and $\partial_{\mathrm{mv}}(\boldsymbol{z}, \boldsymbol{y})$ are ≥ 0 for any value of $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$, this is possible only if $\partial_{\mathrm{drop}}(\boldsymbol{x}, \boldsymbol{z}) = \partial_{\mathrm{mv}}(\boldsymbol{z}, \boldsymbol{y}) = 0$ which means that $\boldsymbol{x} = \boldsymbol{z} = \boldsymbol{y}$. Hence $\boldsymbol{x} = \boldsymbol{y}$.

Property #2: For all $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$, $\partial^\eta_{\mathrm{drmv}}(\boldsymbol{x}, \boldsymbol{z}) \leq \partial^\eta_{\mathrm{drmv}}(\boldsymbol{x}, \boldsymbol{y}) + \partial^\eta_{\mathrm{drmv}}(\boldsymbol{y}, \boldsymbol{z})$.

We assume that the infimum in both $\partial^{\eta}_{\text{drmv}}(\boldsymbol{x},\boldsymbol{y})$ and $\partial^{\eta}_{\text{drmv}}(\boldsymbol{y},\boldsymbol{z})$ is achieved by $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, respectively (the proof can be easily extended to the case when the infimum is not achieved). This means that there exists $a,b,c,d \geq 0$, such that

$$\partial_{\text{drop}}(\boldsymbol{x},\boldsymbol{s}_1) = a;\ \partial_{\text{mv}}(\boldsymbol{s}_1,\boldsymbol{y}) = b;\ \partial_{\text{drop}}(\boldsymbol{y},\boldsymbol{s}_2) = c;\ \partial_{\text{mv}}(\boldsymbol{s}_2,\boldsymbol{z}) = d,$$

which implies $\partial^{\eta}_{\text{drmv}}(\boldsymbol{x},\boldsymbol{y}) = a + \eta b$ and $\partial^{\eta}_{\text{drmv}}(\boldsymbol{y},\boldsymbol{z}) = c + \eta d$. We need to show that $\partial^{\eta}_{\text{drmv}}(\boldsymbol{x},\boldsymbol{z}) \leq (a+c) + \eta(b+d)$.

Using Claim E.2 with $\partial_{\text{mv}}(\boldsymbol{s}_1,\boldsymbol{y}) = b$ and $\partial_{\text{drop}}(\boldsymbol{y},\boldsymbol{s}_2) = c$, we get that there is a $\boldsymbol{y}'$ such that $\partial_{\text{drop}}(\boldsymbol{s}_1,\boldsymbol{y}') = c$ and $\partial_{\text{mv}}(\boldsymbol{y}', s_2) \leq b$. This gives the following:

$$\partial_{\text{drop}}(\boldsymbol{x},\boldsymbol{s}_1) = a;\ \partial_{\text{drop}}(\boldsymbol{s}_1,\boldsymbol{y}') = c;\ \partial_{\text{mv}}(\boldsymbol{y}',\boldsymbol{s}_2) \leq b;\ \partial_{\text{mv}}(\boldsymbol{s}_2,\boldsymbol{z}) = d.$$

Now we prove that $\partial^{\eta}_{\text{drmv}}(\boldsymbol{x},\boldsymbol{z}) \leq (a+c) + \eta(b+d)$:

$$\begin{aligned}
\partial^{\eta}_{\text{drmv}}(\boldsymbol{x},\boldsymbol{z}) &= \inf_{\boldsymbol{z}} \left(\partial_{\text{drop}}(\boldsymbol{x},\boldsymbol{y}) + \eta \cdot \partial_{\text{mv}}(\boldsymbol{y},\boldsymbol{z})\right) \\
&\leq \partial_{\text{drop}}(\boldsymbol{x},\boldsymbol{y}') + \eta \cdot \partial_{\text{mv}}(\boldsymbol{y}',\boldsymbol{z}) \\
&\leq \partial_{\text{drop}}(\boldsymbol{x},\boldsymbol{s}_1) + \partial_{\text{drop}}(\boldsymbol{s}_1,\boldsymbol{y}') + \eta \cdot \partial_{\text{mv}}(\boldsymbol{y}',\boldsymbol{z}) && (\partial_{\text{drop}} \text{ is a quasi-metric}) \\
&\leq \partial_{\text{drop}}(\boldsymbol{x},\boldsymbol{s}_1) + \partial_{\text{drop}}(\boldsymbol{s}_1,\boldsymbol{y}') + \eta \cdot \left(\partial_{\text{mv}}(\boldsymbol{y}',\boldsymbol{s}_2) + \partial_{\text{mv}}(\boldsymbol{s}_2,\boldsymbol{z})\right) && (\partial_{\text{mv}} \text{ is a metric}) \\
&\leq (a+c) + \eta(b+d).
\end{aligned}$$

This concludes the proof of Claim E.3 □

Theorem (Restating Theorem 5.5)**.** *For any deterministic histogram-based-statistic $f_{\text{HBS}} : \mathbb{H}_{[0,B)^d} \to \mathcal{A}$, there exists a mechanism which, on inputs of size at least n, is $\left(O(\frac{1}{\sqrt{\sigma n}}), e^{-\Omega(\sqrt{\sigma n})}\right)$-DP, where $\sigma = \alpha\left(\frac{\beta}{B}\right)^d$, and $(\alpha + \eta\beta, 0, 0)$-accurate for the identity function, w.r.t. the distortion measure $\partial^{\eta}_{\text{drmv}}$.*

Proof. As is easy to see, this closely resembles the guarantees of Theorem 5.4 but without the metric sensitivity in the accuracy. To prove this, we first show that we can achieve the exact same guarantees for the identity functions over histograms. After we show such a mechanism, it is easy to see that we can compose it with any f_{HBS} without any change in parameters. This is because the second parameter (which corresponds to the output error) in $(\alpha + \eta\beta, 0, 0)$-accuracy is 0 and that $\Delta_{f_{\text{HBS}}}(0) = 0$ for all deterministic f_{HBS}. Hence, showing a mechanism with same guarantees for the identity function over histogram will complete the proof.

We do this by first recalling Theorem 5.3 which gave the final privacy and accuracy guarantees for identity function for d-dimensional histograms. We will prove that this exact mechanism satisfied our requirements. Since the privacy guarantees remain the same, we focus on the accuracy guarantees. We will now show that the $(\alpha,\beta,0)$-accuracy guarantee provided by Theorem 5.3 w.r.t. ∂_{drop} as the measure of distortion and $\mathfrak{d}_{\text{hist}}$ as the metric is equivalent to $(\alpha + \eta\beta, 0, 0)$-accuracy w.r.t. $\partial^{\eta}_{\text{drmv}}$ measure of distortion and any metric.

We show this by converting the output error into the distortion. Note that the definition of $\mathfrak{d}_{\text{hist}}$ (the metric used for giving accuracy guarantees in Theorem 5.3) is exactly the same as that of ∂_{mv}. Hence, an error of β in the output can be attributed to (or can be handled by) a distortion of β in the ∂_{mv}-part of $\partial^{\eta}_{\text{drmv}}$. Doing this results in the accuracy of $(\alpha + \eta\beta, 0, 0)$-accuracy for the identity function w.r.t. the distortion measure $\partial^{\eta}_{\text{drmv}}$ and any metric. □

F Omitted Details from Section 5.2 (Robust Privacy Mechanism)

In this section, we prove the omitted proofs from Section 5.2. We prove Claim 5.8 and Claim 5.9 in Appendix F.1 and Lemma 5.4 and Lemma 5.5 in Appendix F.2.

F.1 Proofs of Claim 5.8 and Claim 5.9

Claim (Restating Claim 5.8)**.** *Suppose P is a distribution that is a convex combination of Laplace distributions $\{\mathrm{Lap}(u,b) : u \in \mathcal{U}\}$ for some (discrete) set $\mathcal{U} \subseteq \mathbb{R}$. Then, for every $d \in \mathbb{R}$, we have $\frac{P(x)}{P(x+d)} \in [e^{\frac{-|d|}{b}}, e^{\frac{|d|}{b}}]$. Equivalently, $\ln(P)$ is a $\frac{1}{b}$-Lipschitz function.*

Proof. Let $P(x) = \sum_{u\in\mathcal{U}} Q(u)\mathrm{Lap}(x|u,b)$, where $\sum_{u\in\mathcal{U}} Q(u) = 1$ and $Q(u) \geq 0$. Then we have

$$\frac{P(x)}{P(x+d)} = \frac{\sum_{u\in\mathcal{U}} Q(u)\mathrm{Lap}(x|u,b)}{\sum_{u\in\mathcal{U}} Q(u)\mathrm{Lap}((x+d)|u,b)} \tag{35}$$

It follows from the definition of Laplace distribution that

$$\frac{\mathrm{Lap}(x|\mu,b)}{\mathrm{Lap}(x+d|\mu,b)} = e^{(|x-\mu|-|x+d-\mu|)/b} \leq e^{|d|/b}, \tag{36}$$

$$\frac{\mathrm{Lap}(x|\mu,b)}{\mathrm{Lap}(x+d|\mu,b)} = e^{(|x-\mu|-|x+d-\mu|)/b} \geq e^{-|d|/b}. \tag{37}$$

Using (36) in (35) yields $\frac{P(x)}{P(x+d)} \leq e^{|d|/b}$ as follows:

$$\begin{aligned}
\frac{P(x)}{P(x+d)} &= \frac{\sum_{u\in\mathcal{U}} Q(u)\mathrm{Lap}(x|u,b)}{\sum_{u\in\mathcal{U}} Q(u)\mathrm{Lap}((x+d)|u,b)} \\
&\leq \frac{\sum_{u\in\mathcal{U}} Q(u)e^{|d|/b}\mathrm{Lap}((x+d)|u,b)}{\sum_{u\in\mathcal{U}} Q(u)\mathrm{Lap}((x+d)|u,b)} \\
&\leq e^{|d|/b}
\end{aligned}$$

Using (37) in (35) yields $\frac{P(x)}{P(x+d)} \geq e^{-|d|/b}$ as follows:

$$\begin{aligned}
\frac{P(x)}{P(x+d)} &\geq \frac{\sum_{u\in\mathcal{U}} Q(u)e^{-|d|/b}\mathrm{Lap}((x+d)|u,b)}{\sum_{u\in\mathcal{U}} Q(u)\mathrm{Lap}((x+d)|u,b)} \\
&= e^{-|d|/b}
\end{aligned}$$

This proves Claim 5.8. ∎

Claim (Restating Claim 5.9)**.** *Suppose two distributions P, Q (defined over the same alphabet) both satisfy the log-Lipschitz condition given in Claim 5.8, and $W_\theta^\infty(P,Q) \leq \rho$. Then P, Q satisfy the (ϵ,δ)-DP condition (i.e., for every $\mathcal{S} \subseteq \mathbb{R}$, we have $\Pr_{x\leftarrow P}[x \in \mathcal{S}] \leq e^\epsilon \Pr_{x\leftarrow Q}[x \in \mathcal{S}] + \delta$), where $\epsilon = \frac{2\rho}{b} + \ln(2)$ and $\delta = \frac{\rho\theta}{b(1-e^{-\frac{\rho}{b}})}$.*

Proof. Since $W_\theta^\infty(P,Q) \leq \rho$, we have from Corollary A.1 that there is a distribution P' such that $\Delta(P,P') \leq \theta$ and $W^\infty(P',Q) \leq \rho$.

For $x \in \mathbb{R}$, define two functions $a(x)$ and $z(x)$ as follows:

$$a(x) = \max\{0, P(x) - P'(x)\} \tag{38}$$

$$z(x) = \int_{y=x-\rho}^{y=x+\rho} a(y)\,\mathrm{d}y. \tag{39}$$

Note that $a(x)$ is non-negative and $P'(x) \geq P(x) - a(x)$ holds for every $x \in \mathbb{R}$. We shall use the following claim in the proof.

Claim F.1. $\int_{\mathbb{R}} z(x)\,\mathrm{d}x = 2\rho \int_{\mathbb{R}} a(x)\,\mathrm{d}x \leq 2\rho\theta$.

Proof. The first equality follows by exchanging the order of integration in the expansion of $\int_{\mathbb{R}} z(x)\,\mathrm{d}x$ as follows: $\int_{\mathbb{R}} z(x)\,\mathrm{d}x = \int_{\mathbb{R}} \left(\int_{y=x-\rho}^{y=x+\rho} a(y)\,\mathrm{d}y\right)\,\mathrm{d}x = \int_{\mathbb{R}} \left(\int_{x=y-\rho}^{x=y+\rho} a(y)\,\mathrm{d}x\right)\,\mathrm{d}y = \int_{\mathbb{R}} a(y)\left(\int_{x=y-\rho}^{x=y+\rho}\mathrm{d}x\right)\,\mathrm{d}y = 2\rho\int_{\mathbb{R}} a(y)\,\mathrm{d}y$.

For the second inequality, note that $\int_{\mathbb{R}} a(x)\,\mathrm{d}x = \int_{\mathbb{R}} \max\{0, P(x)-P'(x)\}\,\mathrm{d}x \overset{(a)}{=} \frac{1}{2}\int_{\mathbb{R}} |P(x)-P'(x)|\,\mathrm{d}x = \Delta(P,P') \le \theta$. Here (a) follows from the fact that $\int_{x\,:\,P'(x)\ge P(x)}(P'(x)-P(x))\,\mathrm{d}x = \int_{x\,:\,P'(x)<P(x)}(P(x)-P'(x))\,\mathrm{d}x$. ☐

Fix an arbitrary point $x \in \mathbb{R}$. Let $M_P, M_{P'}$ denote the respective probability masses of P, P' in the interval $[x-\rho, x+\rho]$, and let M_Q denote the probability mass of Q in the interval $[x-2\rho, x+2\rho]$. Since $P(x') \ge P(x) - a(x)$ and by the definition of $z(x)$ from (39), we have $M_{P'} \ge M_P - z(x)$. We shall also prove the following inequality and use it in our proof.

Claim F.2. $M_Q \ge M_{P'}$

Proof. Let $\phi \in \Phi(P', Q)$ be the optimal coupling such that $\max_{(u,v):\phi(u,v)\neq 0} |u-v| = W^\infty(P', Q)$. A crucial observation is the following: it follows from $W^\infty(P', Q) \le \rho$ that, for every (u,v) such that $\phi(u,v)\neq 0$, if $u \in [x-\rho, x+\rho]$, then we have that $v \in [x-2\rho, x+2\rho]$. The above observation implies that $M_{P'} \le M_Q$. Note that P' may not satisfy the log-Lipschitz condition of Claim 5.8, but we do not need this in order to prove the claim. ☐

Since P and Q satisfy the log-Lipschitz condition of Claim 5.8, we have that $\frac{P(y)}{P(x)} \ge e^{\frac{-|x-y|}{b}}$ and $\frac{Q(y)}{Q(x)} \le e^{\frac{|x-y|}{b}}$ hold for every $x, y \in \mathbb{R}$. We can lower-bound M_P and upper-bound M_Q as follows:

$$\begin{aligned}
M_p &= \int_{x-\rho}^{x+\rho} P(y)\,\mathrm{d}y \\
&\ge \int_{x-\rho}^{x+\rho} e^{-\frac{|x-y|}{b}} P(x)\,\mathrm{d}y \\
&= P(x)\int_{x-\rho}^{x+\rho} e^{-\frac{|x-y|}{b}}\,\mathrm{d}y \\
&= P(x)\cdot 2b(1-e^{-\frac{\rho}{b}}) \qquad (40)\\
M_Q &= \int_{x-2\rho}^{x+2\rho} Q(y)\,\mathrm{d}y \\
&\le \int_{x-2\rho}^{x+2\rho} e^{\frac{|x-y|}{b}} Q(x)\,\mathrm{d}y \\
&= Q(x)\int_{x-2\rho}^{x+2\rho} e^{\frac{|x-y|}{b}}\,\mathrm{d}y \\
&= Q(x)\cdot 2b(e^{\frac{2\rho}{b}}-1) \qquad (41)
\end{aligned}$$

By substituting the bounds on M_P and M_Q from (40) and (41), respectively, we have that

$$\begin{aligned}
&P(x)\cdot 2b(1-e^{-\frac{\rho}{b}}) \le Q(x)\cdot 2b(e^{\frac{2\rho}{b}}-1) + z(x) \\
\Rightarrow \qquad & P(x) \le Q(x)\Big(\frac{e^{\frac{2\rho}{b}}-1}{1-e^{-\frac{\rho}{b}}}\Big) + \frac{z(x)}{2b(1-e^{-\frac{\rho}{b}})} \\
&\qquad = Q(x)e^{\frac{\rho}{b}}\Big(\frac{e^{\frac{2\rho}{b}}-1}{e^{\frac{\rho}{b}}-1}\Big) + \frac{z(x)}{2b(1-e^{-\frac{\rho}{b}})} \\
&\qquad = Q(x)e^{\frac{\rho}{b}}(e^{\frac{\rho}{b}}+1) + \frac{z(x)}{2b(1-e^{-\frac{\rho}{b}})}
\end{aligned}$$

$$\leq Q(x)2e^{\frac{2\rho}{b}} + \frac{z(x)}{2b(1-e^{-\frac{\rho}{b}})}$$

which implies $P(x) \leq e^{\epsilon}Q(x) + \frac{z(x)}{2b(1-e^{-\frac{\rho}{b}})}$, where $\epsilon = \frac{2\rho}{b} + \ln(2)$. Using this and integrating over an arbitrary subset $\mathcal{S} \subseteq \mathbb{R}$ and using $\int_{\mathcal{S}} z(x)\,\mathrm{d}x \leq \int_{\mathbb{R}} z(x)\,\mathrm{d}x \leq 2\rho\theta$ (from Claim F.1) gives

$$\Pr_{x \leftarrow P}[x \in \mathcal{S}] \leq e^{\epsilon} \Pr_{x \leftarrow Q}[x \in \mathcal{S}] + \frac{\rho\theta}{b(1-e^{-\frac{\rho}{b}})}. \tag{42}$$

Take $\delta = \frac{\rho\theta}{b(1-e^{-\frac{\rho}{b}})}$. Note that $\delta \to 0$ as $\theta \to 0$. This concludes the proof of Claim 5.9. □

F.2 Proofs of Lemma 5.4 and Lemma 5.5

Now we establish an accuracy result for our mechanism $\mathcal{M}^b_{\text{Lap}}$ in the following lemma.

Lemma (Restating Lemma 5.4)**.** *For every $\gamma \in [0,1]$, $\mathcal{M}^b_{\text{Lap}}$ is $(0,\beta,\gamma)$-accurate, where $\beta = b\ln(\frac{1}{\gamma})$.*

Proof. Fix a $\gamma \in [0,1]$ and any input $x \in \mathbb{R}$. Instead of treating x as a real number, for this proof, we will treat x as a point distribution over $\mathbb{R}$. Clearly, this is equivalent to treating x as a deterministic input. Let Q denote the output distribution of $\mathcal{M}^b_{\text{Lap}}$ when the input is drawn from x. We want to show that $W^\infty_\gamma(x,Q) \leq \beta$, for the above-mentioned β. By definition of W^∞_γ, we have $W^\infty_\gamma(x,Q) = \inf_{\phi \in \Phi^\gamma(x,Q)} \max_{(y,t)\leftarrow\phi}[|y-t|]$.

Consider the following ϕ^*:

$$\phi^*(i,t) = \begin{cases} 0 & \text{if } t < -b\ln(\frac{1}{\gamma}) + i \text{ or } t > b\ln(\frac{1}{\gamma}) + i; \\ \frac{1}{1-\gamma}\text{Lap}(t|b,i)x(i) & \text{if } t \in [-b\ln(\frac{1}{\gamma}) + i, b\ln(\frac{1}{\gamma}) + i]. \end{cases}$$

It can be verified that $\Delta(\phi^*_1, x) = 0$ and $\Delta(\phi^*_2, q) \leq \gamma$, which implies that $\phi^* \in \Phi^\gamma(x,Q)$. This in turn implies that $W^\infty_\gamma(x,Q) \leq \max_{(y,t)\leftarrow\phi^*}[|y-t|]$. It follows from the definition of ϕ^* that $\max_{(y,t)\leftarrow\phi^*}[|y-t|] = b\ln(\frac{1}{\gamma})$, which gives $W^\infty_\gamma(x,Q) \leq b\ln(\frac{1}{\gamma})$. This concludes the proof of Lemma 5.4. □

Lemma (Restating Lemma 5.5)**.** *The error sensitivity of $\mathcal{M}^b_{\text{Lap}}$ has the following upper bounded, for all $\beta_1, \gamma > \gamma_1$, $\tau^{0,\gamma}_{\mathcal{M}^b_{\text{Lap}}}(\beta_1,\gamma_1) \leq \beta_1 + b\ln(\frac{1}{\gamma-\gamma_1})$.*

Proof. Since $\mathcal{M}^b_{\text{Lap}}$ adds independent Laplacian noise, using Lemma 3.2 we have,

$$\begin{aligned} \max_{\substack{P,Q:\\ W^\infty_\theta(P,Q)\leq\beta_1}} W^\infty_\theta(\mathcal{M}^b_{\text{Lap}} \circ P, \mathcal{M}^b_{\text{Lap}} \circ Q) &\overset{\text{(a)}}{\leq} \max_{\substack{P,Q:\\ W^\infty_\theta(P,Q)\leq\beta_1}} W^\infty_\theta(P,Q) \\ &= \beta_1. \end{aligned}$$

It follows from Lemma 3.2 (if P and Q are independent, then $W^\infty_\gamma(P+Q,Q) \leq W^\infty_\gamma(P,\mathfrak{p}_0)$) and Lemma 5.4 that for any $\gamma > 0$ and random variable X, we have

$$W^\infty_\gamma(\mathcal{M}^b_{\text{Lap}}(X), X)) \leq b\ln(\frac{1}{\gamma}).$$

Using these inequalities and that $\mathcal{M}^b_{\text{Lap}}$ is a mechanism for the identity function, for $\gamma > \gamma_1$, we can upper-bound $\tau^{0,\gamma}_{\mathcal{M}^b_{\text{Lap}}}(\beta_1,\gamma_1)$ as follows,

$$\tau^{0,\gamma}_{\mathcal{M}^b_{\text{Lap}}}(\beta_1,\gamma_1) = \sup_{\substack{X,X':\\ W^\infty_{\gamma_1}(X,X')\leq\beta_1}} \inf_{\substack{Y:\\ \widehat{\partial}(X',Y)=0}} W^\infty_\gamma(\mathcal{M}^b_{\text{Lap}}(X), Y)$$

$$
\begin{aligned}
&= \sup_{\substack{X,X':\\ W^\infty_{\gamma_1}(X,X')\le\beta_1}} W^\infty_{\gamma}(\mathcal{M}^b_{\text{Lap}}(X), X') \\
&= \sup_{\substack{X,X':\\ W^\infty_{\gamma_1}(X,X')\le\beta_1}} W^\infty_{\gamma_1+\gamma-\gamma_1}(\mathcal{M}^b_{\text{Lap}}(X), X') \\
&\overset{(a)}{\le} \sup_{\substack{X,X':\\ W^\infty_{\gamma_1}(X,X')\le\beta_1}} \left(W^\infty_{\gamma_1}(\mathcal{M}^b_{\text{Lap}}(X), \mathcal{M}^b_{\text{Lap}}(X')) + W^\infty_{\gamma-\gamma_1}(\mathcal{M}^b_{\text{Lap}}(X'), X')\right) \\
&\le \sup_{\substack{X,X':\\ W^\infty_{\gamma_1}(X,X')\le\beta_1}} W^\infty_{\gamma_1}(\mathcal{M}^b_{\text{Lap}}(X), \mathcal{M}^b_{\text{Lap}}(X')) + \sup_X W^\infty_{\gamma-\gamma_1}(\mathcal{M}^b_{\text{Lap}}(X), X) \\
&\overset{(b)}{\le} \sup_{\substack{X,X':\\ W^\infty_{\gamma_1}(X,X')\le\beta_1}} W^\infty_{\gamma_1}(X, X') + \sup_X W^\infty_{\gamma-\gamma_1}(\mathcal{M}^b_{\text{Lap}}(X), X) \\
&\le \beta_1 + b\ln\left(\frac{1}{\gamma-\gamma_1}\right)
\end{aligned}
$$

Here, (a) follows from Lemma 3.1 and (b) follows from Lemma 3.2. □

G Omitted Details from Section 5.3 (Private Sampling)

Claim G.1. *For any $S \subset \mathbb{R}$, we have* $\Pr[L(P) \in S] \le e^{\epsilon} \Pr[L(Q) \in S] + \delta$

Proof. Take an arbitrary (measurable) subset $S \subset \mathbb{R}$.

$$
\begin{aligned}
\Pr[L(P) \in S] &= \frac{1}{2b}\int_S\int_{\mathbb{R}} P(x) e^{-\frac{|s-x|}{b}}\,\mathrm{d}x\,\mathrm{d}s \\
&= \frac{1}{2b}\int_S\int_{\mathbb{R}} (P'(x)+R(x)) e^{-\frac{|s-x|}{b}}\,\mathrm{d}x\,\mathrm{d}s \\
&= \frac{1}{2b}\int_S\int_{\mathbb{R}} P'(x) e^{-\frac{|s-x|}{b}}\,\mathrm{d}x\,\mathrm{d}s + \frac{1}{2b}\int_S\int_{\mathbb{R}} R(x) e^{-\frac{|s-x|}{b}}\,\mathrm{d}x\,\mathrm{d}s \\
&= \frac{1}{2b}\int_S\int_{\mathbb{R}} P'(x) e^{-\frac{|s-x|}{b}}\,\mathrm{d}x\,\mathrm{d}s + \int_{\mathbb{R}} R(x)\int_S \frac{1}{2b} e^{-\frac{|s-x|}{b}}\,\mathrm{d}s\,\mathrm{d}x \\
&\le \frac{1}{2b}\int_S\int_{\mathbb{R}} P'(x) e^{-\frac{|s-x|}{b}}\,\mathrm{d}x\,\mathrm{d}s + \int_{\mathbb{R}:\mathbb{R}(x)>0} R(x)\int_S \frac{1}{2b} e^{-\frac{|s-x|}{b}}\,\mathrm{d}s\,\mathrm{d}x \\
&\le \frac{1}{2b}\int_S\int_{\mathbb{R}} P'(x) e^{-\frac{|s-x|}{b}}\,\mathrm{d}x\,\mathrm{d}s + \int_{\mathbb{R}:\mathbb{R}(x)>0} R(x)\cdot 1 \\
&\le \frac{1}{2b}\int_S\int_{\mathbb{R}} P'(x) e^{-\frac{|s-x|}{b}}\,\mathrm{d}x\,\mathrm{d}s + \theta \\
&= \frac{1}{2b}\int_S\int_{\mathbb{R}}\int_{\mathbb{R}} \phi(x,y) e^{-\frac{|s-x|}{b}}\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}s + \theta \\
&= \frac{1}{2b}\int_S\int_{\mathbb{R}}\int_{\mathbb{R}} \phi(x,y) e^{-\frac{|s-x|}{b}}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s + \theta \\
&\le \frac{1}{2b}\int_S\int_{\mathbb{R}}\int_{\mathbb{R}} \phi(x,y) e^{\frac{|x-y|}{b}} e^{-\frac{|s-y|}{b}}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s + \theta \\
&\qquad \text{(Using the triangle inequality } |s-x| \ge |s-y| - |x-y|) \\
&\le \frac{1}{2b}\int_S\int_{\mathbb{R}}\int_{\mathbb{R}} \phi(x,y) e^{\frac{S^\theta(f)}{b}} e^{-\frac{|s-y|}{b}}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}s + \theta \\
&\qquad \text{(Since for all } (x,y) \text{ such that } \phi(x,y)\ne 0\text{, we have } |x-y| \le S^\theta(f))
\end{aligned}
$$

$$
\begin{aligned}
&= e^{\frac{S^\theta(f)}{b}} \frac{1}{2b} \int_S \int_{\mathbb{R}} \int_{\mathbb{R}} \phi(x,y) e^{-\frac{|s-y|}{b}} \,\mathrm{d}x \,\mathrm{d}y \,\mathrm{d}s + \theta \\
&= e^{\frac{S^\theta(f)}{b}} \frac{1}{2b} \int_S \int_{\mathbb{R}} Q(y) e^{-\frac{|s-y|}{b}} \,\mathrm{d}y \,\mathrm{d}s + \theta \\
&= e^{\frac{S^\theta(f)}{b}} \Pr[L(Q) \in S] + \theta
\end{aligned}
$$

Since the above calculations hold for an arbitrary subset of $\mathbb{R}$, they hold for all subsets of $\mathbb{R}$ too. $\square$