Extending the Foundations of Differential Privacy: Robustness and Flexibility

Aman Bansal\textsuperscript{1}, Rahul Chunduru\textsuperscript{1}, Deepesh Data\textsuperscript{2}, and Manoj Prabhakaran\textsuperscript{1}

\textsuperscript{1}Indian Institute of Technology Bombay, India
aman005bansal003@gmail.com, \{chrahul, mp\}@cse.iitb.ac.in

\textsuperscript{2}University of California, Los Angeles, USA
deepe\texthyph{}sh.data@gmail.com

Abstract

In this work, we make foundational contributions to the area of Differential Privacy (DP).

Our first contribution is definitional. We define two complementary concepts that greatly enhance the applicability of DP, namely, robust privacy and flexible accuracy. Robust privacy requires that a mechanism provides the “best possible privacy” without further degrading accuracy guarantees, even if such privacy is not \textit{a priori} anticipated based on input neighborhoods alone. Flexible accuracy allows small distortions in the input (e.g., dropping outliers) before measuring accuracy of the output. Along the way, we also extend the notion of DP to sampling (i.e., computation of randomized functions).

Our second contribution is in establishing versatile composition theorems that relate these notions.

Our third contribution is constructive: We present mechanisms that can help in achieving these notions, where previously no meaningful differentially private mechanisms were possible. 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.

1 Introduction

In this work, we make foundational contributions to the area of Differential Privacy (DP), greatly extending its applicability. Apart from the immediate implications to private machine learning, due to their fundamental nature, we expect the new notions to throw more light on the connections of differential privacy with machine learning and other areas.

Our main contribution is to identify and address two limitations of the DP framework that seem to have evaded attention. 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 start by elaborating on the limitations of DP addressed in this work.
**Limits Set by Sensitivity.** Consider a simplistic learning task which tries to learn an upper bound on integer valued observations – say, ages of patients who recovered from a certain disease – presented to it. For the sake of privacy, one may wish to apply a DP mechanism, rather than output the maximum in the sample 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 mechanism, 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 composition theorem enables us to transfer the accuracy guarantees on histogram to accuracy guarantees on the maximum function.

**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 \(D_2\), where the two datasets are disjoint. 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 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 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 output from the model trained in \(D_1\) has an even value for the least significant digit, and the other has an odd value, the total variation distance between the two output distributions is maximum, while the Wasserstein distance can be very small.

In short, DP only guarantees indistinguishability between datasets which are close to each other in the input space, whereas 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 defined for a mechanism that addresses this.

\(^1\)Indeed, 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 \([NRS07]\) also do not fare any better.
1.1 Our Contributions

Our contributions are in three parts:

- **Definitions:** We define *flexible accuracy* and *robustness* to address the above limitations. Flexible accuracy considers error after applying a small *distortion of the input*. Robustness formalizes the notion of mechanisms which offer *best possible privacy* for the utility they offer.

- **Composition Theorems:** We prove composition theorems that yield guarantees of *flexible accuracy* and *robustness* for *function composition* (rather than for multiple queries). In particular, we show that given a mechanism $M$ for a function $f$ which has zero error after input distortion, then one can derive mechanism for any function $g \circ f$ simply as $g \circ M$, and retain similar flexible accuracy guarantees. We also give a robustness composition theorem to incorporate the robustness guarantee from a mechanism for the identity function into another mechanism, while retaining the latter’s differential privacy guarantees, and the combined flexible accuracy guarantees.

- **Mechanisms:** We give a new differentially private mechanism for histograms, that has zero error after input distortion. This allows one to apply our composition theorems and derive differentially private mechanism for functions of the histogram (like support, maximum, range etc.). Previously, no non-trivial privacy and accuracy guarantees could be achieved for these functions, despite them being very natural statistics required to build models for the data. We also give a robust mechanism for the identity function, that can be used to impart robustness to any mechanism via our robustness composition.

We elaborate on these below.

**Flexible Accuracy.** The high-level idea is to allow for some *distortion of the input* when measuring accuracy. We shall require distortion to be defined using a *quasi-metric* over the input space (a quasi-metric is akin to a metric, but is not required to be symmetric). A typical form of distortion is to *drop a few items* from the dataset; in this case, adding a data item is not considered low distortion. Referring back to the example of reporting maximum, given a dataset with a single elderly patient and many young patients, flexible accuracy with respect to this distortion allows a mechanism for maximum to report the maximum age of the younger group.

Quantitatively, flexible accuracy is presented in two parts – it separately accounts for errors that can be attributed to distortion of the input (input error), and to inaccuracies in the output (output error). The output error in itself is more general than the conventional “probabilistically approximately correct” (PAC) measure of accuracy, as we shall allow the correct output to be randomized (i.e., defined by a distribution). We use a variant of Wasserstein distance to measure the output error. While one may consider the standard Wasserstein distance to already account for the probability and the approximation parts together, we find it helpful to define a 2-parameter notion of accuracy here, using a new measure that we introduce, called *lossy Wasserstein distance.*

As described below (see “Composition Theorems” below), flexible accuracy also provides us with a means for transferring *accuracy guarantees* when composed with other functions or mechanisms.

**Robustness.** We define a mechanism whose output is in a metric space to be robust 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 in the sense of

\[ \text{lossy Wasserstein distance}. \]

Wasserstein distance is also called the Earth Mover Distance, referring to the 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.
differential privacy. 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 robustness, imagine being given a mechanism, which has a certain level of utility. We consider utility as being unaffected by small perturbations of the output. However, the output of the mechanism may contain hints as to the input, which do not contribute to its utility. Robustness 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 fixed point of this iteration of removing non-utile signals.

To contrast DP and robustness, one may view DP mechanisms as trying to provide the best accuracy possible given an a priori privacy specification, whereas robust mechanisms must guarantee the best privacy that is possible given its accuracy. We recommend using a DP mechanism that is also robust.

Composition Theorems. An important part of our contributions is a new class of composition theorems for DP. Traditionally, the DP framework has provided composition theorems which extend DP guarantees of a single function evaluation query to DP guarantees for several (possibly adaptively chosen) queries. Here composition refers to forming a composite query whose answer consists of the answers of all its constituent queries. However, we provide an entirely different class of composition theorems, where multiple mechanisms are composed in the sense of function composition: i.e., the output of the one query is fed as the input to the next one, and only the output from the last query is given to the adversary.

While the input and output errors tend to get amplified in such composition, they do so differently. In a composed mechanism $M_0 \circ M_1$, only the output error of $M_0$ and the input error of $M_1$ are amplified, with the other errors simply adding up. In particular, if these two errors are absent, there is no amplification at all. Note that the ability to separate the error into input and output components is crucial in exploiting this composition theorem (and is perhaps why such composition was not considered in prior literature when such separation did not exist).

Applications and New Mechanisms. We illustrate the power of our extensions by giving new robust and differentially private mechanisms for support and maximum functions, with much higher (but flexible) accuracy than previously possible. We emphasize that for previous DP mechanisms, due to the high-sensitivity of these functions, no non-trivial accuracy guarantees were available.

Towards this, we present a new mechanism for the histogram function (which already had a DP mechanism with good accuracy), with only input error and zero output error. This mechanism can then be composed with simple deterministic mechanisms to learn support and maximum of the data. Recall that even a very low (but non-zero) output error for histograms can result in maximum computed from the histogram (as the maximum-valued bar with a non-zero height) to be wildly inaccurate. But if the inaccuracy in the histogram can be entirely attributed to a distortion in the input, computing maximum on this histogram does not amplify the inaccuracy at all.

The new mechanism for histogram is a simple variation on the standard mechanism (though, as sketched in Section 1.2, it was arrived at by solving for the optimum mechanism, ab initio): to each bar we add noise that is distributed according to the Laplace distribution, but (a) with a negative mean (rather than 0 mean), (b) bounded so that it is never positive, and (c) rounded (and hinged) so that value after adding the noise is a non-negative integer. (The standard mechanism uses a 0-mean Laplace distribution.)

This mechanism works well when the number of bars is small. If that is not the case, then this is composed with a simple bucketing mechanism, to reduce the number of bars. Support and maximum can then be directly computed from the resulting histogram.
The above mechanism is not robust. But it can be made robust (adding bounded output error and no additional input error) using a general compiler we give. The compiler corresponds to simply adding Laplace noise with appropriate parameters. This result could be seen as showing that the standard Laplace mechanism used to achieve DP also achieves robustness, provided the parameters are chosen appropriately.

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. A work leading to the development of DP by Chawla et al. [CDM+05] used a definition that explicitly attempts to address the privacy of outliers. They also gave a mechanism for privacy-preserving histograms, by showing that if the data is sampled from a high-dimensional hypercube, then recursive histogram sanitization – in which the cube is recursively partitioned into smaller regions until no region contains a certain number of data points, and then the exact count of each region is revealed – preserves privacy with high probability. Later, Chawla et al. [CDMT05], among other things, extended this approach to include a broad class of rich and “nice” distributions, including the mixture of high-dimensional Gaussians, balls, and spheres. A subsequent work in the DP framework by Dwork and Lei [DL09] revisited the presence of outliers from the perspective of “robust statistics.” However, formulating their accuracy guarantees requires a distribution over the data, whereas flexible accuracy is a worst-case guarantee. Incidentally, Wasserstein distance has been used in privacy mechanisms in the Pufferfish framework [KM14, SWC17], which also relies on a distribution over data.

1.2 Truncated Laplacian Mechanism

Our new differentially private histogram mechanism was designed to optimize privacy parameters while minimizing the input distortion (and has no further output error). We briefly motivate this construction.

Consider the 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 $(\epsilon, \delta)$privacy, we require

$$p_k \leq p_{k+1} e^\epsilon + \delta,$$

$$(1 - p_k) \leq (1 - p_{k+1}) e^\epsilon + \delta,$$

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

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

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

In Figure 1, on the left, we plot the probabilities $p_k$ against $k/q$ for this choice of $(\epsilon, \delta)$. 

Figure 1: 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).

To generalize this boolean mechanism to a full-fledged application like histogram, 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 \( \nu \leq -k \) should be \( 1 - p_k \). That is, if the noise distribution is given by the density function \( \sigma \), we require that

\[
\int_{-q}^{-k} \sigma(t) \cdot dt = 1 - p_k
\]

Substituting the expression for \( p_k \) from (1), and then differentiating this identity with respect to \( k \), we obtain the following expression for \( \sigma(t) \), for \( t \in [-q,0] \):

\[
\sigma(t) = \frac{1}{1 - e^{-eq/2}} \text{Lap}(t \mid -\frac{q}{2}, \frac{1}{\epsilon}),
\]

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

Our final histogram mechanism in Section 3.1 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).

2 Preliminaries

**Notation.** We denote by \( \mathbb{N} \) all the non-negative integers (including zero). For \( i, j \in \mathbb{N} \), such that \( i \leq j \), we write \([i:j]\) to denote the set \( \{i,i+1,\ldots,j\} \). For elements \( x, y \in \mathbb{R}^n \), we write \(|x-y|\) to denote the distance between them, according to a norm of choice, say, the Euclidean norm. We shall use probability density functions to describe probability distributions, as typically, the distributions of interest to us are continuous in nature (and in other cases, the density functions can be understood to involve the Dirac delta function).

It will be convenient for us to denote probability distributions by random variables, as it would let us conveniently define operations on distributions. For instance, the random variable \( M(X) \) corresponds to the distribution of the output of \( M \) on an input drawn from the distribution associated with the random variable \( X \). We shall write \( p_X \) to denote the probability distribution of a random variable \( X \), so that \( p_X(\omega) = \Pr[X = \omega] \).
Definition 1 (Total Variation Distance). Let $p$ and $q$ be two probability distributions on a sample space $\Omega$. The total variation distance between $p$ and $q$, denoted by $\Delta(p, q)$, is defined as

$$\Delta(p, q) = \frac{1}{2} \int_{\omega \in \Omega} |p(\omega) - q(\omega)| d\omega.$$  

2.1 Differential Privacy

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

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

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

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

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

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

2.2 Wasserstein Distance and Related Notions

Wasserstein distance is defined for distributions on a sample space $\Omega$ which is equipped with a metric $\mathfrak{d}$. Concretely, in the following, we may consider $\Omega = \mathbb{R}^n$ and the metric $\mathfrak{d}$ being an $\ell_p$-metric.

For $\theta \in [0, 1]$, and distribution $p, q$ over a metric space $(\Omega, \mathfrak{d})$, we define $\Phi_\theta(p, q)$ as the set of joint distributions $\phi$ over $\Omega^2$ with marginals $\phi_1$ and $\phi_2$ such that $\Delta(\phi_1, p) + \Delta(\phi_2, q) \leq \theta$ hold. Note that the joint distributions in $\Phi_\theta(p, q)$ have marginals exactly equal to $p$ and $q$.

Definition 4 ($\theta$-Lossy Wasserstein Distance). Let $p$ and $q$ be two probability distributions over a metric space $(\Omega, \mathfrak{d})$, and let $\theta \in [0, 1]$. The $\theta$-lossy Wasserstein distance between $p$ and $q$ is defined as:

$$W_\theta(p, q) = \inf_{\phi \in \Phi_\theta(p, q)} \mathbb{E}_{(x,y) \sim \phi} [\mathfrak{d}(x, y)].$$  

(3)

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

$$W_\theta^\infty(p, q) = \inf_{\phi \in \Phi_\theta(p, q)} \max \mathfrak{d}(x, y).$$  

(4)

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

Lossy $\infty$-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 $\beta$, within an approximation radius $\gamma$ of a desired deterministic
quantity $f$: i.e., $\Pr[|g - f| > \gamma] \leq \beta$ (the probability being over the choice of $g$). Considering $f$ as a point distribution, this can be equivalently written as $W_\beta^\infty(f, g) \leq \gamma$.

The following lemma relates lossy Wasserstein and lossy $\infty$-Wasserstein distances and is proven in Appendix A.

**Lemma 1.** For any two distributions $p, q$, and $0 \leq \beta' < \beta \leq 1$,

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

The following two lemmas are used in our proofs and proven in Appendix A. The first one establishes a version of the triangle inequality:

**Lemma 2.** For distributions $p, q, r$ over a metric space $(\mathbb{R}^n, d)$, and $\gamma_1, \gamma_2 \geq 0$.

$$W_{\gamma_1 + \gamma_2}(p, r) \leq W_{\gamma_1}(p, q) + W_{\gamma_2}(q, r), \quad (5)$$

$$W_{\gamma_1 + \gamma_2}^\infty(p, r) \leq W_{\gamma_1}^\infty(p, q) + W_{\gamma_2}^\infty(q, r). \quad (6)$$

The next lemma bounds the effect that adding independent noise to two distributions has on their (lossy) Wasserstein distances. The noise distribution is characterized in terms of its Wasserstein distance from the point distribution $p_0$, that has all its mass at 0.

**Lemma 3.** Let $X, Y$ and $Z$ be three random variables over a metric space $(\mathbb{R}^n, d)$ with $Z$ being independent of $X$ and $Y$, and let $p_0$ denote the distribution with all its mass at 0. Then $\forall \gamma, \gamma_1$ such that $\gamma, \gamma_1 \geq 0$ and $\gamma_1 \leq \gamma/2$, we have

$$W_{\gamma}(p_{x + z}, p_{y + z}) \leq W_{\gamma}(p_x, p_y) \leq W_{\gamma - \gamma_1}(p_{x + z}, p_{y + z}) + 2W_{\gamma_1}(p_0, p_z), \quad (7)$$

$$W_{\gamma}^\infty(p_{x + z}, p_{y + z}) \leq W_{\gamma}^\infty(p_x, p_y) \leq W_{\gamma - \gamma_1}^\infty(p_{x + z}, p_{y + z}) + 2W_{\gamma_1}^\infty(p_0, p_z). \quad (8)$$

### 3 Flexible Accuracy

Now we present our idea behind flexible accuracy. We would like to define “natural” distortions of a database, that are meaningful for the function to be evaluated on the database. 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 should be defined not using a metric over databases, but a quasi-metric. A quasi-metric is the same as a metric, except that it does not have to be symmetric. We shall use a quasi-metric whose range is normalized to the interval $[0, 1]$, but also includes a special value $\infty$, to indicate 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 $\partial : \mathcal{X} \times \mathcal{X} \to [0, 1] \cup \{\infty\}$ which forms a quasi-metric over $\mathcal{X}$.

Our main example of a measure of distortion is $\partial_{\text{drop}}$, which is defined when each element in $\mathcal{X}$ is a finite multiset over a ground set $\mathcal{G}$. Formally, $x \in \mathcal{X}$ is a function $x : \mathcal{G} \to \mathbb{N}$ that outputs the multiplicity of each element of $\mathcal{G}$. Then, for finite $x, x' \in \mathcal{X}$, we define

$$\partial_{\text{drop}}(x, x') := \left\{ \begin{array}{ll} \sum_{g \in \mathcal{G}} x(g) - x'(g) & \text{if } \forall g \in \mathcal{G}, x(g) \geq x'(g) \\
\infty & \text{otherwise.} \end{array} \right. \quad (9)$$

Informally, flexible accuracy with a distortion bound $\alpha$ guarantees that on input $x$, a mechanism shall produce an output that corresponds to $f(x')$ for some $x'$ such that $\partial(x, x') \leq \alpha$. In addition to such input distortion, we may allow the output to be only probabilistically approximately correct, with an approximation error parameter $\beta$ and an error probability parameter $\gamma$. Formally, the probabilistic approximation guarantee of the output is given as a bound of $\beta$ on a $\gamma$-lossy Wasserstein distance.
Definition 7 ((α, β, γ)-accuracy). Let ∂ be a measure of distortion on a set X and f : X → Y be a randomized function. A mechanism M is said to be (α, β, γ)-accurate for f with respect to ∂, if for each x ∈ X, there is a random variable X′ with support contained in \{x′|∂(x, x′) ≤ α\} such that \(W_γ(f(X′), M(x)) ≤ β\).

3.1 Histogram Mechanism

In this section we propose a new differentially private mechanism for the histogram function, and show that it achieves (α, 0, 0)-accuracy with respect to ∂drop.

We define the histogram function hist with respect to a partition \(P = \{G_1, \cdots, G_t\}\) of a ground set \(G\). It takes a multiset \(x\) with elements in \(G\), and produces a table indexed by \([t]\) with the \(i^{th}\) row showing the number of elements in \(x\) that fall into the part \(G_i\). Formally, writing \(x\) as the multiplicity function, we define hist\(_P\)(\(x\)) : \([t]\) → \(\mathbb{N}\) as a function such that hist\(_P\)(\(x\)) : \(i \mapsto \sum_{g \in G_i} x(g)\).

Recall that ∂drop was defined as a measure of distortion on such multisets. To define (ε, δ)-differential privacy, we use the standard notion of neighborhood: \(x \sim x'\) if \(\exists g^* \in G\) such that \(\forall g \neq g^*, x(g) = x'(g^*) \) and \(|x(g^*) - x'(g^*)| ≤ 1\).

The desired accuracy guarantee of the mechanism is easy to see:

Lemma 4. On inputs which have non-empty intersection with at most \(t\) parts of a partition \(P\), \(\mathcal{M}_{trLap}^\tau\) (defined in Construction 1) is a \((\tau t, 0, 0)\)-accurate mechanism for hist\(_P\) with respect to distortion ∂drop, where \(\tau\) is the parameter of the mechanism and \(t\) is the number of parts in the partition.

Construction 1. Truncated Laplacian Mechanism, \(\mathcal{M}_{trLap}^\tau\)

Parameters: Threshold \(\tau\), partition \(P\) of a ground set as \(G = G_1 \cup \cdots \cup G_t\).

Input: A multiset, \(x \subseteq G\).

Output: A histogram, \(\tilde{y} : [t] \rightarrow \mathbb{N}\).

1: \(y := \text{hist}_P(x)\)
2: for all \(i \in [t]\) do
3: \(\text{sample } z_i \leftarrow \pi_q, \text{ where } q = \tau |x|\) and
4: \(\pi_q(z) = \begin{cases} \frac{1}{1-e^{-\frac{xq}{2}}} \text{Lap}(z \mid -\frac{q}{2}, \sqrt{q}) & \text{if } -q < z < 0, \\ 0 & \text{otherwise,} \end{cases}\)
5: \(\tilde{y}(i) := \max(0, |y(i) + z_i|)\)
6: end for
7: Return \(\tilde{y}\)

Proof: Note that for any \(i\), with probability 1, \(z_i > -q\). Thus \(x'\) such that \(\text{hist}_P(x') = \tilde{y}\) can be obtained by dropping less than \(tq = t\tau|x|\) elements from \(x\). This implies that the mechanism is \((\tau t, 0, 0)\)-accurate.

Now we prove that the mechanism is also differentially private, with parameters that improve with the size of the input.

Lemma 5. On inputs of size at least \(n\), \(\mathcal{M}_{trLap}^\tau\) (defined in Construction 1) is an \((\epsilon, \delta)\)-differentially private mechanism, where \(\epsilon = \frac{1}{\sqrt{\tau n}}\) and \(\delta = \frac{4}{\sqrt{\tau n}}e^{-\frac{\sqrt{\tau}}{2}}\).
Proof sketch: We shall in fact prove that a mechanism which outputs \( \hat{y} \) with \( \hat{y}(i) := y(i) + z_i \) (without rounding, and without replacing negative values with 0) is already differentially private as desired. Then, since the actual mechanism is a postprocessing of this mechanism, it will also be \((\epsilon, \delta)\)-differentially private.

Consider any two neighbouring databases \( x \) and \( x' \). This means that \( x \) and \( x' \) only differ in one entry which implies that histograms of the two databases differ in only one value and by a difference of one. Let us denote this value as \( S \).

\[
\Pr_{y \leftarrow M(\mathcal{D}_1 \cup \mathcal{D}_2)}[y \in S] = e^\epsilon \Pr_{y \leftarrow M(\mathcal{D}_2)}[y \in S] + \delta,
\]

for any subset \( S \) of histograms. To prove this for a particular \( S \), we divide \( S \) into two parts, \( S_1 \) and \( S_2 \) such that \( \Pr_{y \leftarrow M(\mathcal{D}_1 \cup \mathcal{D}_2)}[y \in S_1] > 0 \) and \( \Pr_{y \leftarrow M(\mathcal{D}_1 \cup \mathcal{D}_2)}[y \in S_2] = 0 \).

For \( S_1 \), it can be shown that \( \Pr_{y \leftarrow M(\mathcal{D}_1 \cup \mathcal{D}_2)}[y \in S_1] \leq e^\epsilon \Pr_{y \leftarrow M(\mathcal{D}_2)}[y \in S_1] \). For \( S_2 \), it can be shown that \( \Pr_{y \leftarrow M(\mathcal{D}_1 \cup \mathcal{D}_2)}[y \in S_2] \leq \delta \). Combining these two results will complete the proof. Complete proof is given in Appendix B. \( \square \)

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 accuracy and privacy guarantees for \( \mathcal{M} \) in terms of those for \( \mathcal{M}_1 \) and \( \mathcal{M}_2 \).

4.1 Flexible Accuracy Under Composition

Error and Distortion Sensitivity. Intuitively, if a function \( f \) has distortion sensitivity \( \sigma_f \), then distorting \( f(x) \) arbitrarily within an \( \alpha \) radius can be modeled as distorting \( x \) within a \( \sigma_f(\alpha) \) radius. For generality, we shall allow \( x \) to be distorted into a distribution. To formalize this we need a measure of distortion that applies to distributions rather than to a pair of elements. We start by defining this quantity, \( \hat{\partial} \), which upper bounds the distortion that is incurred when transporting one distribution to another. It is defined in the same way as \( W^\infty \), but with respect to a quasi-metric rather than a metric.

**Definition 8** (Extension of a Measure of Distortion to Distributions). For a measure of distortion \( \partial \) over a set \( \mathcal{X} \), the extension of \( \partial \) to distributions, \( \hat{\partial} \) maps a pair of distributions \( p, q \) over \( \mathcal{X} \) to a real number as

\[
\hat{\partial}(p, q) = \inf_{\phi \in \Phi(\mathcal{X}, \mathcal{Y})} \sup_{(x, y) : \phi(x, y) 
eq 0} \partial(x, y)
\]

If \( p \) is a point distribution with all its mass on a point \( x \), we denote \( \hat{\partial}(p, q) \) as \( \hat{\partial}(x, q) \).

Now, we define distortion sensitivity of \( f \). It measures how much any input \( x \) needs to be distorted into a random variable \( X \) so that \( f(X) \) is close (in Wasserstein distance) to any random variable \( Y \) which is at most \( \alpha \) distorted from \( f(x) \).

**Definition 9** (Distortion sensitivity). Let \( f : A \rightarrow B \) be a randomized function where \( B \) admits Wasserstein distances. Let \( \partial_1, \partial_2 \) be measures of distortion on \( A, B \) respectively. Then, the distortion-sensitivity of \( f \) w.r.t. \( (\partial_1, \partial_2) \) and \( \theta \in [0, 1] \) and \( \omega \geq 0 \), is defined as the function \( \sigma^\theta_\omega : [0, 1] \rightarrow [0, 1] \) given by

\[
\sigma^\theta_\omega(\alpha) = \sup_{x, y} \inf_{X : \hat{\partial}_1(x, p_X) \leq \omega} \mathcal{W}(f(x), p_Y) \leq \omega
\]
where $x \in A$, and random variables $X, Y$ are over domains $A, B$ respectively. We write $\sigma_f^\theta$ to denote $\sigma_f^{\theta,0}$ and $\sigma_f^\theta$ to denote $\sigma_f^{0,0}$.

**Remark 1.** The distortion sensitivity of the identity function $f : A \rightarrow A$ with respect to $(\hat{\partial}, \hat{\partial})$ for any distortion measure $\hat{\partial}$ over $A$ satisfies $\sigma_f^\theta(\alpha) \leq \alpha$.

Error sensitivity of a mechanism $\mathcal{M}$ measures the extent to which an error in its input is amplified in its output. More precisely, we would like to bound how far the distributions $\mathcal{M}(X)$ and $\mathcal{M}(Y)$ can be when $X$ and $Y$ are close to each other, with both distances measured in terms of Wasserstein distance.

**Definition 10 (Error sensitivity).** Let $\mathcal{M} : A \rightarrow B$ be a randomized mechanism where $A$ and $B$ both admit Wasserstein distances. For any $\theta \in [0, 1]$, we define the error-sensitivity of $\mathcal{M}$ as the function $\tau^\theta_{\mathcal{M}} : [0, 1] \rightarrow [0, 1]$ given by

$$\tau^\theta_{\mathcal{M}}(\beta) = \sup_{X, Y : W_\theta(\mathcal{M}(X), \mathcal{M}(Y)) \leq \beta} \tau^\theta_{\mathcal{M}}(\beta).$$

(11)

The following lemmas are proven in Appendix C.

**Lemma 6.** If $\hat{\partial}$ is a measure of distortion over $A$, then $\hat{\partial}$ is a quasi-metric.

**Lemma 7.** If $\mathcal{M} : A \rightarrow B$ is an $(\alpha, \beta, \gamma)$-accurate mechanism for $f$ w.r.t. $\hat{\partial}$, then for any random variable $X$ over $A$, there is a random variable $X^*$ such that

$$\hat{\partial}(p_X, p_{X^*}) \leq \alpha, \quad W_\gamma(f(X^*), \mathcal{M}(X)) \leq \beta.$$

The next lemma translates the definition of distortion sensitivity (Definition 9) to apply to distortion of input distributions. For simplicity, we state this lemma assuming that the infima in the definition of $\sigma$ are achieved. (This is the case, e.g., when the measure of distortion is discrete, as would be in our applications.) However, even without this assumption, the lemma continues to hold by considering $\sigma_f^\theta(\alpha) + \nu$ for an infinitesimal $\nu$.

**Lemma 8.** Suppose $f : A \rightarrow B$ has distortion sensitivity $\sigma_f^\theta$, w.r.t. $(\partial_1, \partial_2)$. Then, for random variables $X_0$ over $A$, and $Y$ over $B$ such that $\partial_2(f(X_0), p_Y) \leq \alpha$, there exists a distribution $X$ over $A$ such that $W_\theta(f(X), p_Y) = 0$ and $\partial_1(p_{X_0}, p_X) \leq \sigma_f^\theta(\alpha)$.

**Theorem 1.** Let $M_1 : A \rightarrow B$ and $M_2 : B \rightarrow C$ be mechanisms which are $(\alpha_1, \beta_1, \gamma_1)$ and $(\alpha_2, \beta_2, \gamma_2)$ accurate for functions $f_1 : A \rightarrow B$ and $f_2 : B \rightarrow C$ with respect to measures of distortion $\partial_1$, $\partial_2$ defined on $A, B$ and metrics $\mathcal{d}_1, \mathcal{d}_2$ defined on $B, C$ respectively. Then the mechanism $M_2 \circ M_1 : A \rightarrow C$ is, for any $\theta \in [0, 1]$, $(\alpha, \beta, \gamma)$-accurate for the function $f_2 \circ f_1$ w.r.t distortion $\partial_1$ and metric $\mathcal{d}_2$, where $\alpha = \alpha_1 + \sigma_{f_1}(\alpha_2)$, $\beta = \tau_{M_2}(\beta_1) + \beta_2$, $\gamma = \gamma_1 + \gamma_2 + \theta$.

**Proof:** To compare $f_2 \circ f_1$ and $M_2 \circ M_1$, we consider the hybrid mechanism $M_2 \circ f_1$. For a given element $x \in A$, since $M_1$ is $(\alpha_1, \beta_1, \gamma_1)$-accurate mechanism for $f_1$, there exists a random variable $X'$ with distribution $p_{X'}$ such that

$$\hat{\partial}_1(x, p_{X'}) \leq \alpha \quad \text{(12)}$$

$$W_{\gamma_1}(f_1(X'), M_1(x)) \leq \beta_1 \quad \text{(13)}$$

Now, applying the mechanism $M_2$ over the distributions $M_1(x), f_1(X')$ may amplify the error to at most $\tau_{M_2}(\beta_1)$ (see Definition 10), which gives

$$W_{\gamma_1}(M_2(f_1(X')), M_2(M_1(x))) \leq \tau_{M_2}(\beta_1). \quad \text{(14)}$$
To bound the effect of applying $\mathcal{M}_2$ on distortion, we note that by Lemma 7, there exists a random variable $Y^*$ such that,

$$\hat{d}_2(f_1(X'), p_{Y^*}) \leq \alpha_2$$  \hspace{1cm} (15)

$$W_{\gamma_2}(f_2(Y^*), \mathcal{M}_2(f_1(X'))) \leq \beta_2$$  \hspace{1cm} (16)

From (15) and Lemma 8, there exists $X$ over $A$ such that

$$\hat{d}_1(p_{X'}, p_X) \leq \sigma^0_{f_1}(\alpha_2) \hspace{1cm} W_\theta(f(X), p_{Y^*}) = 0.$$  

Now by Lemma 6, since $\hat{d}_1$ being a quasi-metric, so is $\hat{d}_1$. Hence, combined with (12), we have

$$\hat{d}_1(x, p_X) \leq \alpha_1 + \sigma^0_{f_1}(\alpha_2).$$

On the other hand, combined with (14) and (16), this gives

$$W_{\gamma_1 + \gamma_2 + \theta}(f_2(f_1(X)), \mathcal{M}_2(\mathcal{M}_1(x))) \leq W_{\gamma_1}(\mathcal{M}_2(f_1(X')), \mathcal{M}_2(\mathcal{M}_1(x)))$$

$$+ W_{\gamma_2}(f_2(Y^*), \mathcal{M}_2(f_1(X')))$$

$$+ W_\theta(f(X), p_{Y^*})$$

$$\leq \tau^\mathcal{M}_2_1(\beta_1) + \beta_2.$$  

This completes the proof of Theorem 1. \hfill \square

### 4.2 Privacy Under Composition

We prove a “pre-processing theorem” complementing the “post-processing theorem” for differential privacy (see [DR14]). The post-processing theorem states that if $\mathcal{M}_1$ is $(\epsilon, \delta)$ differentially private, then for any arbitrary mechanism $\mathcal{M}_2$, the mechanism $\mathcal{M}_2 \circ \mathcal{M}_1$ would remain $(\epsilon, \delta)$ differentially private. Our pre-processing theorem roughly states that if $\mathcal{M}_2$ is differentially private, then so would $\mathcal{M}_2 \circ \mathcal{M}_1$ be (i.e., pre-processing by $\mathcal{M}_1$ does not hurt privacy); but this can hold only if $\mathcal{M}_1$ is well-behaved. Following is a notion of being well-behaved that suffices for our purposes.

**Definition 11** (Neighborhood preserving Mechanism). A mechanism $\mathcal{M} : A \rightarrow B$ 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. $p_X = \mathcal{M}(x)$ and $p_Y = \mathcal{M}(y)$, and $\Pr[X \sim_B Y] = 1$.

**Theorem 2.** Let $\mathcal{M}_1 : A \rightarrow B$ be a neighborhood preserving mechanism w.r.t. neighborhood relations $\sim_A$ and $\sim_B$ over $A$ and $B$ respectively, and let $\mathcal{M}_2 : B \rightarrow C$ be an $(\epsilon, \delta)$-differentially private mechanism w.r.t. $\sim_B$. Then the mechanism $\mathcal{M}_2 \circ \mathcal{M}_1 : A \rightarrow C$ is $(\epsilon, \delta)$-differentially private 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. $p_{X_1} = \mathcal{M}_1(x)$, $p_{X_2} = \mathcal{M}_1(x')$ and $\Pr[X_1 \sim_B X_2] = 1$. So, for all $(x_1, x_2)$ such that $p_{X_1, X_2}(x_1, x_2) > 0$, we have $x_1 \sim_B x_2$ and hence, by the $(\epsilon, \delta)$-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,

\[
\Pr[\mathcal{M}_2(\mathcal{M}_1(x)) \in S] = \sum_{x_1} p_{x_1}(x_1) \Pr[\mathcal{M}_2(x_1) \in S] = \sum_{(x_1, x_2)} p_{x_1, x_2}(x_1, x_2) \Pr[\mathcal{M}_2(x_1) \in S] \\
\leq \sum_{(x_1, x_2)} p_{x_1, x_2}(x_1, x_2) (e^\varepsilon \Pr[\mathcal{M}_2(x_2) \in S] + \delta) \\
= e^\varepsilon \left( \sum_{(x_1, x_2)} p_{x_1, x_2}(x_1, x_2) \Pr[\mathcal{M}_2(x_2) \in S] \right) + \delta \\
= e^\varepsilon \Pr[\mathcal{M}_2(\mathcal{M}_1(x')) \in S] + \delta
\]

\( \square \)

### 4.3 Applications

Now we give applications of our composition theorems. We note that Construction 1 for histograms works well when the number of parts in partition \( \mathcal{P} \) is small. If that is not the case, then this mechanism can be composed with a simple bucketing mechanism (Construction 2), to reduce the number of bars; see Figure 2.

Once we have obtained the histogram in a differentially private and accurate manner, we can do post-processing and compute deterministic functions (e.g., max and support) over it, and the resulting mechanism satisfies similar accuracy guarantees (without degrading the privacy).

To this end, we define deterministic mechanisms \( f_{\text{max}}, f_{\text{supp}} \) on histograms which returns the maximum element with non-zero occurrence and the set of elements with non-zero occurrences, respectively; see Figure 3.

For two histograms \( y, y' : [0, B) \rightarrow \mathbb{N} \), we define the distance metric \( \mathfrak{d}_{\text{hist}}(y, y') \) as

\[
\mathfrak{d}_{\text{hist}}(y, y') := W^\infty(\text{norm}(y), \text{norm}(y')) \tag{17}
\]

where \( \text{norm}(y) \) returns the histogram with each bar in \( y \) scaled by a constant factor so that sum of all the bars equals 1 and can be thought of as a discrete probability distribution.

We define a neighborhood relation \( \sim_{\text{mset}} \) over multisets as follows: for two multisets \( x, x' \), we say that \( x \sim_{\text{mset}} x' \), if \( x \) can be obtained from \( x' \) by adding or dropping at most one element.

Now we show the following results for the mechanisms constructed in Figure 2.

**Lemma 9.** The Bucketing Mechanism \( \mathcal{M}_{\text{buc}}^{t,B} \) (from Construction 2) is \((0, B/2t, 0)\)-accurate for the identity function, w.r.t. distortion measure \( \mathfrak{d}_{\text{drop}} \) and the absolute difference metric over \( \mathbb{R} \).

**Proof:** Note that \( \mathcal{M}_{\text{buc}}^{t,B} \) edits every element by at most \( B/2t \), which implies that for any input \( x \), we have \( W(\text{norm}(x), \text{norm}(\mathcal{M}_{\text{buc}}^{t,B}(x))) \leq B/2t \), where the underlying metric for \( W \) is the absolute difference metric over \( \mathbb{R} \) – distance between two real numbers. Therefore, \( \mathcal{M}_{\text{buc}}^{t,B} \) is a \((0, B/2t, 0)\)-accurate mechanism for the identity function. \( \square \)

The following theorems are proven in Appendix D.

**Theorem 3.** On an \( n \)-element database, \( \mathcal{M}_{1\text{D}-\text{hist}}^{\alpha,\beta,B} \) (from Construction 3) is \((\alpha, \beta, 0)\)-accurate w.r.t. the distortion measure \( \mathfrak{d}_{\text{drop}} \) and metric \( \mathfrak{d}_{\text{hist}}(\cdot) \), and \( (\frac{1}{\sqrt{q}}, \frac{4}{\sqrt{q}} e^{-\frac{\varepsilon}{2t}}) \)-differentially private, where \( q = 2^{naB/B} \), for the histogram function.
Construction 2. Bucketing Mechanism, \( M_{\text{buc}}^{t,B} \)

**Parameters:** Desired number of buckets \( t \), range of inputs \( B \).

**Input:** A multiset \( x \) with elements in \([0, B)\).

**Output:** A multiset \( y \subseteq \{(i - \frac{1}{2})\frac{B}{t} | i \in [t]\} \)

1. for all \( i \in [t] \) do
2. \( I_i := \left[ \frac{B}{t}(i - 1), \frac{B}{t}i \right) \) and \( w_i = \frac{B}{t}(i - \frac{1}{2}) \)
3. \( y(w_i) := \sum_{g \in x \cap I_i} x(g) \)
4. end for

\( y(w) \) is multiplicity of \( w \) in \( y \)

\( y(w) = 0 \) for all other \( w \)

5: Return \( y \)

Construction 3. Histogram Mechanism, \( M_{\text{1D-hist}}^{\alpha,\beta,B} \)

**Parameters:** Accuracy parameters \( \alpha, \beta \); data range \( B \).

**Input:** A multiset \( x \) with elements in \([0, B)\).

**Output:** A histogram for the input.

1: \( t := \lceil \frac{B}{2\beta} \rceil \)
2: \( P := (I_1, \cdots, I_t) \) where \( I_i = \left[ \frac{B}{t}(i - 1), \frac{B}{t}i \right) \).
3: \( \tau := \alpha/t \)
4: Return \( M_{\text{trLap}}^{\tau,P} \circ M_{\text{buc}}^{t,B}(x) \)

---

**Figure 2:** Mechanisms for computing Histogram.

Now we apply the above theorems and get mechanisms for the max and the support functions.

In the output space of support function on histograms, i.e., the power set, \( 2^{[0,B)} \), we define the metric, \( d_{\text{supp}}(\cdot, \cdot) \) as follows. For two elements \( X, Y \in 2^{[0,B)} \), we define,

\[
d_{\text{supp}}(X, Y) = \max\left\{ \max_{x \in X} \min_{y \in Y} |x - y|, \max_{y \in Y} \min_{x \in X} |x - y| \right\}
\] (18)

It is easy to see that \( d_{\text{supp}}(\cdot, \cdot) \) is a metric. Intuitively, it is the smallest radius by which either of the supports needs to be "grown" to cover the elements of the other.

**Theorem 4.** On an \( n \)-element database, \( M_{\text{1D-sup}}^{\alpha,\beta,B} \) (from Construction 4) is \( (\frac{1}{\sqrt{q}}, \frac{4}{\sqrt{q}}e^{-\frac{q}{2}}) \)-differentially private, with \( q = 2^{n\alpha\beta}/B \), and \((\alpha, \beta, 0)\)-accurate w.r.t. the distortion measure \( \partial_{\text{drop}} \) and metric \( d_{\text{supp}}(\cdot, \cdot) \) for the function \( f_{\text{supp}} \).

**Theorem 5.** On and \( n \)-element database, \( M_{\text{max}}^{\alpha,\beta,B} \) (from Construction 5) is a \( (\frac{1}{\sqrt{q}}, \frac{4}{\sqrt{q}}e^{-\frac{q}{2}}) \)-differentially private mechanism, with \( q = 2^{n\alpha\beta}/B \), that is \((\alpha, \beta, 0)\)-accurate w.r.t. distortion measure \( \partial_{\text{drop}} \) and the standard metric for \( \mathbb{R} \), for the function \( f_{\text{max}} \).

5 Robust Mechanisms

**Definition 12** (Robustness). Let \( \rho, \theta, \epsilon, \delta \) be non-negative real numbers with \( \delta, \theta \leq 1 \). Let \( M : A \to B \) be a randomized mechanism, and let \( \mathfrak{d} \) be a metric over \( B \). The mechanism \( M \) is said
Construction 4. Support Mechanism, $\mathcal{M}_{1D-\text{supp}}^{\alpha,\beta,B}$

**Parameters:** Accuracy parameters $\alpha, \beta$; data range $B$.

**Input:** A multiset $x$ with elements in $[0, B)$.

**Output:** A set $S \subseteq [t]$.

1: Return $f_{\text{supp}} \circ \mathcal{M}_{1D-\text{hist}}^{\alpha,\beta,B}(x)$

Construction 5. Max Mechanism, $\mathcal{M}_{\text{max}}^{\alpha,\beta,B}$

**Parameters:** Accuracy parameters $\alpha, \beta$; data range $B$.

**Input:** A multiset $x$ with elements in $[0, B)$.

**Output:** A value in $[0, B)$.

1: Return $f_{\text{max}} \circ \mathcal{M}_{1D-\text{hist}}^{\alpha,\beta,B}(x)$

Figure 3: Mechanisms for computing Support and Max.

to be $(\theta, \rho, \epsilon, \delta)$-robust w.r.t. $\mathfrak{d}$ if for all distributions $p, q$ over $A$ such that $W_\theta^\infty(\mathcal{M}(p), \mathcal{M}(q)) \leq \rho$ (where $W_\theta^\infty$ is defined w.r.t. the metric $\mathfrak{d}$),

$$\Pr_{x \sim p} [\mathcal{M}(x) \in S] \leq e^\epsilon \Pr_{x \sim q} [\mathcal{M}(x) \in S] + \delta \quad \forall S \subseteq B \quad (19)$$

5.1 Robustness via Composition

It turns out that, as the robustness property is formulated entirely in terms of the output distributions (with no reference to the input neighborhoods or the function of interest), it is easily transferred via post-processing. The following theorem captures this (with simplified parameters, for clarity).

**Theorem 6.** Suppose a mechanism $\mathcal{M} : A \rightarrow A$ is $(\epsilon, \delta)$-private and $(\alpha, \beta, \gamma)$-accurate for some function $f : A \rightarrow B$, w.r.t. a distortion measure $\mathfrak{d}$ and metric $\mathfrak{d}$. Also, suppose $\mathcal{M}_{\text{rob}} : B \rightarrow B$ is a mechanism for a function $g : B \rightarrow B$ which has $(\theta, \rho, \epsilon_r, \delta_r)$-robustness and $(0, \beta', \gamma')$-accuracy w.r.t. the metric $\mathfrak{d}$, and error sensitivity upper bounded by identity function (i.e., for all $\beta_1$, $\tau_{\mathcal{M}_{\text{rob}}}^\theta(\beta_1) \leq \beta_1$). Then $\mathcal{M}_{\text{rob}} \circ \mathcal{M}$ is a mechanism for $g \circ f$ with

1. $(\theta, \rho, \epsilon_r, \delta_r)$-robustness w.r.t. $\mathfrak{d}$,
2. $(\epsilon, \delta)$-differential privacy, and
3. $(\alpha, \beta + \beta', \gamma + \gamma')$-accuracy w.r.t. distortion measure $\mathfrak{d}$ and metric $\mathfrak{d}$.

**Proof:** We prove the three guarantees one by one below.

1. **Robustness:** The proof of robustness is straightforward. By Definition 12, if any two input distributions result in output distributions that are close in terms of Wasserstein distance, then the output distributions are also close in the sense of differential privacy. Since the outputs of $\mathcal{M}_{\text{rob}} \circ \mathcal{M}$ on two input distributions $X, X'$ are the outputs of $\mathcal{M}_{\text{rob}}$ on two input distributions $\mathcal{M}(X), \mathcal{M}(X')$, the robustness guarantee of $\mathcal{M}_{\text{rob}}$ holds for $\mathcal{M}_{\text{rob}} \circ \mathcal{M}$ as well. That is, $\mathcal{M}_{\text{rob}} \circ \mathcal{M}$ is $(\theta, \rho, \epsilon_r, \delta_r)$-robust.
2. **Privacy:** For the privacy guarantee, since $\mathcal{M}$ is $(\epsilon, \delta)$-differentially private and post-processing preserves differential privacy [DR14, Proposition 2.1], $\mathcal{M}_{rob} \circ \mathcal{M}$ is also $(\epsilon, \delta)$-differentially private.

3. **Accuracy:** We apply Theorem 1 to $\mathcal{M}_{rob} \circ \mathcal{M}$, taking $\theta = 0$ and using the fact that $\sigma^0(0) = 0$ for any function $f$). Then, substituting the parameters of $\mathcal{M}_{rob}$ and $\mathcal{M}$, we get that $\mathcal{M}_{rob} \circ \mathcal{M}$ is $(\alpha, \beta' + \beta, \gamma' + \gamma)$-accurate for $g \circ f$ (w.r.t. the same distortion and metric as used for $\mathcal{M}$).

This completes the proof of Theorem 6.

5.2 A Robust Mechanism for the Identity Function over $\mathbb{R}$

We now propose a simple Laplace mechanism $\mathcal{M} = \text{Lap}(b) : \mathbb{R} \rightarrow \mathbb{R}$ and show that it is a robust mechanism for the identity function. The mechanism is as follows.

**Construction 6.** $\mathcal{M} = \text{Lap}(b)$: On input $y \in \mathbb{R}$, sample $z$ according to the probability distribution $\text{Lap}(b)$ and output $y + z$.

As shown below, $\mathcal{M} = \text{Lap}(b)$ satisfies the conditions of $M_{rob}$ required in Theorem 6.

**Theorem 7.** For any $\theta, \gamma \in [0, 1]$, $\mathcal{M} = \text{Lap}(b)$ is $(\theta, \rho, \epsilon, \delta)$-robust and $(0, \beta, \gamma)$-accurate w.r.t. the absolute difference metric in $\mathbb{R}$, and has error sensitivity upper bounded by identity function (i.e., for all $\beta, r_{M_{rob}}^{(0)}(\beta_1) \leq \beta_1$), where

$$\epsilon = 2\rho \frac{\rho}{b} + \ln(2) \quad \delta = \frac{\rho \theta}{b(1 - e^{-\frac{\theta}{b}})} \quad \beta = b\left(1 - \frac{\ln\left(\frac{1}{\gamma}\right)}{\frac{1}{\gamma} - 1}\right).$$

Theorem 7 follows from Lemma 10, Lemma 11, and Lemma 12, proven below. Note that this lets one use $\mathcal{M} = \text{Lap}(b)$ as the robust mechanism $\mathcal{M}_{rob}$ in Theorem 6 (with $g$ being the identity function) to compile the mechanism $\mathcal{M}$ for a function $f$ into a robust mechanism for the same function.

**Lemma 10.** For any $\theta \in [0, 1]$, $\mathcal{M} = \text{Lap}(b)$ achieves $(\theta, \rho, 2\rho, b\ln(2), \frac{\rho \theta}{b(1 - e^{-\frac{\theta}{b}})})$-robustness.

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

Fix $\theta \in [0, 1]$. Suppose $q$ and $q'$ denote two input distributions defined over a discrete subset $U \subseteq \mathbb{R}$, taking the value $u \in U$ with probability $q_u$ and $q_u'$ respectively. Let $p$ and $p'$ denote the corresponding output distributions of $\mathcal{M} = \text{Lap}(b)$. Also, let $W_\theta^\infty(p, p') \leq \rho$.

Note that $p(x) = \sum_{u \in U} q_u \text{Lap}(x|u, b)$. That is, $p$ is a convex combination of Laplace distributions. The same holds for $p'$. Claim 1 below shows that such distributions are log-Lipschitz, and then Claim 2 establishes the robustness claimed in Lemma 10. Both the claims are proved in Appendix E.

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

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

16
This concludes the proof of Lemma 10. □

Now we establish an accuracy result for our mechanism \( M_{\text{Lap}}^b \) in the following lemma.

**Lemma 11.** For every constant \( 0 \leq \gamma \leq 1 \), \( M_{\text{Lap}}^b \) is \((0, \beta, \gamma)\)-accurate, where \( \beta = b \left( 1 - \frac{\ln(\frac{b}{\gamma})}{\gamma - 1} \right) \).

*Note that if \( \gamma = 0 \), then \( \beta = b \).*

**Proof:** Fix a constant \( \gamma \geq 0 \) and any input \( x \in \mathbb{R} \). Instead of treating \( x \) as a real, 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 \( M_{\text{Lap}}^b \) when the input is drawn from \( x \). We want to show that \( W_\gamma(x, q) \leq \beta \), for the above-mentioned \( \beta \). By definition of \( W_\gamma \) from (3) we have \( W_\gamma(x, q) = \inf_{\phi \in \Phi^\gamma(x, q)} E_{(x, t) \sim \phi} [|y - t|] \).

Consider the following \( \phi^* \):

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

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 \( W_\gamma(x, q) \leq E_{(x, t) \sim \phi^*} [|y - t|] \). We show in Appendix E that \( E_{(y, t) \sim \phi^*} [|y - t|] \leq \frac{b}{(1-\gamma)} \left( 1 - \gamma \right) (1 + \ln(\frac{1}{\gamma})) \). This will prove Lemma 11. □

**Lemma 12.** \( M_{\text{Lap}}^b \) has error sensitivity upper bounded by identity function, i.e., for all \( \beta_1 \), \( \tau_{M_{\text{Lap}}^b}^\beta(\beta_1) \leq \beta_1 \).

**Proof:**

\[
\tau_{M_{\text{Lap}}^b}^\beta(\beta_1) = \max_{p, q \in \mathcal{P}} W_\theta(M_{\text{Lap}}^b \circ p, M_{\text{Lap}}^b \circ q) \\
= \max_{p, q \in \mathcal{P}} W_\theta(p, q) \leq \beta_1.
\]

Here (a) follows from Lemma 3 together with the fact that \( M_{\text{Lap}}^b \) adds independent Laplacian noise with the same distribution. □

### 6 Differential Privacy for Randomized Functions

First, we generalize the notion of sensitivity to randomized functions. Let \( f : \mathcal{X} \to \mathcal{Y} \) be a randomized function. That is, for each \( x \in \mathcal{X} \), \( f(x) \) is a random variable over \( \mathcal{Y} \) with a probability distribution that depends on \( x \). Recall that \( \mathcal{X} \) has an associated neighborhood relation \( \sim \). In the following, we assume \( \mathcal{X} \) to be a collection of databases. Then, for any two distinct databases \( x, x' \in \mathcal{X} \), we write \( x \sim x' \) to denote that \( x \) and \( x' \) differ in only one element.

**Definition 13** (Parameterized Sensitivity of a Randomized Query). For \( \theta \in [0, 1] \), we define \( \theta\)-sensitivity of a randomized function \( f \), denoted by \( S^\theta(f) \), as:

\[
S^\theta(f) := \max_{x, x' \in \mathcal{X}, x \sim x'} W^\theta_\phi(f(x), f(x')).
\]
Consider the following randomized mechanism for a randomized query $f : \mathcal{X} \rightarrow \mathbb{R}$, followed by a theorem, which establishes its robustness, privacy, and accuracy guarantees.

**Construction 7. Randomized Query Mechanism $M_b$**

**Parameters:** Laplace Parameter $b$.

**Input:** $x \in \mathcal{X}$

**Output:** $y \in \mathbb{R}$

**Algorithm:** Sample $y_1 \leftarrow f(x)$, $y_2 \leftarrow \text{Lap}(b)$, and output $y_1 + y_2$.

**Theorem 8.** For a randomized query $f$, Construction 7 is

(i) $\forall \theta' \in [0, 1], (\theta', S^0(f), \epsilon', \delta')$-robust where $\epsilon' = 2 S^0(f) b + \ln(2)$, $\delta = \theta'$,

(ii) $(\epsilon, \delta)$-differentially private where $\epsilon = 2 S^0(f) b + \ln(2)$, $\delta = \theta$ and

(iii) $(0, \beta, \gamma)$-accurate where $\beta = b \left( 1 - \frac{\ln(\frac{1}{\gamma})}{1-1} \right)$.

**Proof:** Observe that $f$ can be treated as a mechanism for $f$ with $(0,0,0)$-accuracy and $(\infty,0)$-privacy and that Construction 7 is equivalent to $M_{\text{rand}} := M_{\text{rob}} \circ f$, where in $M_{\text{rob}}$, the Laplace noise has $\rho = S^0(f)$. The robustness and accuracy bounds are obtained from Theorem 6. To show that $M_{\text{rand}}$ is $(\epsilon,\delta)$-differentially private, consider any two neighbouring databases $x, x' \in \mathcal{X}^n$. Since $x \sim x'$, it follows from Definition 13 that $W_{\theta}^{\infty}(f(x), f(x')) \leq S^0(f)$. Since we add independent noise in the mechanism $M_{\text{rob}}$, it follows from Lemma 3 that $W_{\theta}^{\infty}(M_{\text{rand}}(x), M_{\text{rand}}(x')) \leq W_{\theta}^{\infty}(f(x), f(x')) \leq S^0(f)$ This, together with the fact that Construction 7 is $(\theta, S^0(f), \epsilon, \delta)$-robust, implies that $Pr[M_{\text{rand}}(x) \in S] \leq e^\epsilon Pr[M_{\text{rand}}(x') \in S] + \delta$, where $\epsilon = 2 S^0(f) b + \ln(2)$, $\delta = \theta$. \hfill $\square$

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.
Acknowledgement

We thank Satish Kanaparthi for helpful discussions during the early stages of this work.

References


A Omitted Details from Section 2

Lemma (Restating Lemma 1). For any two distributions $p, q$, and $0 \leq \beta' < \beta \leq 1$,

$$W_{\beta}(p, q) \leq W_{\beta'}(p, q) \leq W_{\beta}(p, q)/(\beta - \beta').$$
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 \( \phi \), the total mass that is transported more than a distance \( \gamma' \) is at most \( \gamma/\gamma' \) and the total mass that is lost is at most \( \beta' \). 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 \( \gamma' \). Choosing \( \gamma' = \gamma/(\beta - \beta') \) this upper bound on loss is \( \beta \), and hence this modified coupling shows that \( W_\beta^\infty(p, q) \leq \gamma'. \)

We have the following triangle inequality for lossy Wasserstein distances.

**Lemma (Restating Lemma 2).** Let \( p, q, r \) be any three distributions defined over the same metric space \( (\mathbb{R}^n, \| \cdot \|) \), and let \( \gamma_1, \gamma_2 \geq 0 \). We have

\[
W_{\gamma_1 + \gamma_2}(p, r) \leq W_{\gamma_1}(p, q) + W_{\gamma_2}(q, r),
\]

(21)

\[
W_{\gamma_1 + \gamma_2}^\infty(p, r) \leq W_{\gamma_1}^\infty(p, q) + W_{\gamma_2}^\infty(q, r).
\]

(22)

**Proof:** We only prove (22); and the proof can be extended to show the other inequality (21). First we show the following claim regarding \( W_\infty \), and then using that claim, we show (22).

**Claim 3.** Let \( p, q \) be two given probability distributions over the same space. Let \( p' \) be a distribution such that \( \Delta(p, p') \leq \delta \). Then, there exists another distribution \( q' \), such that we have \( \Delta(q, q') \leq \delta \) and \( W_\infty(p', q) \geq W_\infty(p, q') \).

**Proof:** For brevity, we prove this result only for discrete distributions. The proof can be extended to work with continuous distributions also. Fix any two distributions \( p, q \) over the same space. Let \( p' \) be a distribution such that \( \Delta(p, p') \leq \delta \). Note that the set \( \Phi^0(p', q) \) is a closed set, which implies that the infimum is achieved in the definition of \( W_\infty \). Let \( \phi \in \Phi^0(p', q) \) achieves the infimum, i.e., \( \max_{(x,y) \in \Phi} \| x - y \| = W_\infty(p', q) \). In the following, we convert \( \phi \) into another joint distribution \( \phi' \), whose marginals are \( p \) and \( q' \), such that we have \( \Delta(q, q') \leq \delta \) and \( \max_{(x,y) \in \Phi} \| x - y \| \leq \max_{(x,y) \in \Phi} \| x - y \| \).

Let \( S \) be equal to the union of the support of \( p \) and the support of \( p' \). Since \( \Delta(p, p') \leq \delta \), we have that \( p' \) has extra mass at some points in the support and less mass at some other points in support as compared to \( p \). Let the set of points where \( p' \) has more mass be denoted by \( Y \subset S \), and the set of points where \( p' \) has less mass be denoted by \( X \subset S \). For every \( s \in S \), let \( g(s) \) denote \( |p(s) - p'(s)| \). We have (i) \( g(s) = p(s) - p'(s) \), if \( s \in X \), (ii) \( g(s) = p'(s) - p(s) \), if \( s \in Y \), and (iii) \( g(s) = 0 \), i.e., \( p(s) = p'(s) \), if \( s \in S \setminus \{X \cup Y\} \). Note that, since \( \Delta(p, p') \leq \delta \), we have \( \sum_{x \in X} g(x) = \sum_{y \in Y} g(y) \leq \delta \). Now we construct \( \phi' \) as follows.

Initially, let \( \phi' \) be equal to \( \phi \). Now, \( \forall x \in X \), add \( g(x) \) to \( \phi'(x, x) \). This implies that (i) for every \( s \in X \), we have \( \sum_{s' \in S} \phi'(s, s') = p'(s) + g(s) = p(s) \), (ii) for every \( s \in Y \), we have \( \sum_{s' \in S} \phi'(s, s') = p'(s) - g(s) + p(s) \), and (iii) for every \( s \in S \setminus \{X \cup Y\} \), we have \( \sum_{s' \in S} \phi'(s, s') = p'(s) \). We want to have the first marginal of \( \phi' \) to be equal to \( p \). For that, we subtract some non-negative value from \( \phi'(y, s) \), \( \forall y \in Y, s \in S \), without making any of them negative, such that \( \forall y \in Y \), a total of \( g(y) \) is subtracted from \( \sum_{s \in S} \phi'(y, s) \). Since \( \forall y \in Y \), we have that \( \sum_{s \in S} \phi'(y, s) \geq g(y) \), this can be done. After doing this, the resulting joint distribution is the required \( \phi' \). Observe that the resulting \( \phi' \) has its first marginal equal to \( p \).

It follows from the above construction of obtaining \( \phi' \) from \( \phi \) that

- \( \phi' \) is a valid joint distribution, i.e., all joint probabilities are non-negative and sum up to 1. This follows because initially we had \( \phi' = \phi \), then we added \( \sum_{x \in X} g(x) \) to \( \phi' \), and then subtracted \( \sum_{y \in Y} g(y) = \sum_{x \in X} g(x) \) from \( \phi' \), without making any of the values negative.
- The first marginal of \( \phi' \) is equal to \( p \).
Let the second marginal be $q'$. It is easy to verify that $\Delta(q, q') \leq \delta$. This can be seen from the fact that $\sum_{x \in X} g(x) = \sum_{y \in Y} g(y) \leq \delta$, and that in the construction, first we added $g(x)$ to $\phi'(x, x)$ for every $x \in X$ (a total of $\sum_{x \in X} g(x)$) and then subtracted a total of $g(y)$ from $\sum_{s \in S} \phi'(y, s)$ for every $y \in Y$ (a total of $\sum_{y \in Y} g(y)$), without making any of the values negative.

We have $\max_{(x,y) \to \phi'} \mathcal{D}(x,y) \leq \max_{(x,y) \to \phi} \mathcal{D}(x,y)$. This follows because the only mass added in $\phi'$ starting from $\phi$ was added to $\phi'(x,x)$, for $x \in X$, so $\max_{(x,y) \to \phi'} \mathcal{D}(x,y)$ cannot be more than $\max_{(x,y) \to \phi} \mathcal{D}(x,y)$, since there is no $(x,y), y \neq x$ which has non-zero mass in $\phi'$ but zero mass in $\phi$.

Now it follows that $W_\infty(p', q) \geq W_\infty(p, q')$, as shown below.

\[
W_\infty(p', q) = \max_{(x,y) \to \phi} \mathcal{D}(x,y) \\
\geq \max_{(x,y) \to \phi'} \mathcal{D}(x,y) \\
\geq \inf_{\phi' \in \Phi^0(p,q')} \max_{(x,y) \to \phi'} \mathcal{D}(x,y) \quad \text{(Since $\phi' \in \Phi^0(p,q')$)} \\
= W_\infty(p, q')
\]

This proves Claim 3. \hfill \Box

Observe that $W_\gamma(p, q) = \inf_{p', q'} W_\infty(p', q')$, which follows from the following set of equalities:

\[
W_\gamma(p, q) \stackrel{(a)}{=} \inf_{\phi \in \Phi^\gamma(p,q)} \max_{(x,y) \to \phi} \mathcal{D}(x,y) \\
\stackrel{(b)}{=} \inf_{p', q'} \left( \inf_{\phi \in \Phi^0(p',q')} \max_{(x,y) \to \phi} \mathcal{D}(x,y) \right) \quad \text{subject to $\Delta(p,p') + \Delta(q,q') \leq \gamma$} \\
\stackrel{(c)}{=} \inf_{p', q'} W_\infty(p', q') \quad \text{subject to $\Delta(p,p') + \Delta(q,q') \leq \gamma$} \tag{23}
\]

where (a) follows from the definition of $\gamma$-Lossy $\infty$-Wasserstein distance (Definition 5); (b) trivially holds by viewing the infimum set differently; and in (c) we substituted the definition of $W_\infty$.

Now, using Claim 3, in the following we show that

\[
\inf_{p', q'} W_\infty(p', q') = \inf_{p''} W_\infty(p'', q) \quad \text{subject to $\Delta(p,p'') + \Delta(q,q') \leq \gamma$} \tag{24}
\]

Note that $\leq$ trivially holds. To show the other inequality, it can be verified that the set $\{(p', q') : \Delta(p, p') + \Delta(q, q') \leq \gamma\}$ is a closed set and, therefore, infimum is achieved. Let $\bar{p}', \bar{q}'$ be a pair of distributions that achieves the infimum in the LHS of (24), which implies that $\Delta(p, \bar{p}') + \Delta(q, \bar{q}') \leq \gamma$. Consider $W_\infty(\bar{p}', \bar{q}')$. Note that $\Delta(q, \bar{q}') \leq \gamma - \Delta(p, \bar{p}')$. Now it follows from Claim 3 that there exists a distribution $p''$ such that $\Delta(p', p'') \leq \gamma - \Delta(p, \bar{p}')$ and that $W_\infty(\bar{p}', \bar{q}') \geq W_\infty(p'', q)$ hold. Note that we have $\Delta(p, p'') \leq \Delta(p, \bar{p}') + \Delta(\bar{p}', p'') \leq \gamma$.

Now taking infimum over all $p''$ such that $\Delta(p, p'') \leq \gamma$ yields $W_\infty(\bar{p}', \bar{q}') \geq \inf_{p''} W_\infty(p'', q)$, which proves the other inequality $\geq$ in (24). Thus, we have shown (24). Combining (23) and
(24) gives
\[ W_\gamma^\infty(p, q) = \inf_{\Delta(p, p') \leq \gamma} W^\infty(p', q). \] (25)

Now we are ready to show (22).
\[
W_{\gamma_1 + \gamma_2}^\infty(p, r) = \\
\inf_{\Delta(p, p') + \Delta(r, r') \leq \gamma_1 + \gamma_2} W^\infty(p', r') \\
\leq \\
\inf_{\Delta(p, p') \leq \gamma_1} W^\infty(p', r') \\
\leq \\
\inf_{\Delta(p, p') \leq \gamma_1} W^\infty(p', q) + W^\infty(q, r') \\
= \\
\inf_{\Delta(p, p') \leq \gamma_1} W^\infty(p', q) + \inf_{\Delta(p, p') \leq \gamma_1} W^\infty(q, r') \\
\leq \\
W_{\gamma_1}^\infty(p, q) + W_{\gamma_2}^\infty(q, r).
\]

Here (d) follows from (23); (e) follows from the fact that we are taking infimum in the RHS of (e) over a subset; (f) follows from the fact that the \(\infty\)-Wasserstein distance is a metric; and (g) follows from (25).

This completes the proof of Lemma 2. \(\square\)

**Lemma (Restating Lemma 3).** Let \(X, Y\) and \(Z\) be three random variables over a metric space \((\mathbb{R}^n, d)\) with \(Z\) being independent of \(X\) and \(Y\), and let \(p_0\) denote the distribution with all its mass at 0. Then \(\forall \gamma, \gamma_1\) such that \(\gamma, \gamma_1 \geq 0\) and \(\gamma_1 \leq \gamma/2\), we have
\[
W_\gamma(p_{X+Z}, p_{Y+Z}) \leq W_\gamma(p_X, p_Y) \leq W_{\gamma-2\gamma_1}(p_{X+Z}, p_{Y+Z}) + 2W_{\gamma_1}(p_0, p_Z), \quad (26)
\]
\[
W_\gamma^\infty(p_{X+Z}, p_{Y+Z}) \leq W_\gamma^\infty(p_X, p_Y) \leq W_{\gamma-2\gamma_1}^\infty(p_{X+Z}, p_{Y+Z}) + 2W_{\gamma_1}^\infty(p_0, p_Z). \quad (27)
\]

**Proof:** We only prove (26); and the proof can be extended to show (27). For (26), first we prove (a) and then we prove (b).

**Proof of (a).** Let \(W_\gamma(p_X, p_Y) = \beta\) and let \(\phi \in \Phi_\gamma(p_X, p_Y)\) be any joint distribution with \(E_{(x,y)\sim\phi} [d(x, y)] = \beta\). Let \(p_\phi, q_\phi\) denote its marginals, i.e., \(p_\phi(x) = \int_{-\infty}^{\infty} \phi(x, y) \cdot dy\) and \(q_\phi(y) = \int_{-\infty}^{\infty} \phi(x, y) \cdot dx\). Consider the following joint distribution \(\phi'\):
\[
\phi'(x, y) = \int_{-\infty}^{\infty} p_z(z) \phi(x - z, y - z) \cdot dz.
\]

Now we show that \(\phi' \in \Phi_\gamma(p_{X+Z}, p_{Y+Z})\). For this, let \(p_{\phi'}, q_{\phi'}\) denote the marginals of \(\phi'\). We can easily show that \(p_{\phi'}(x) = \int_{x=-\infty}^{x=\infty} r(z)p_{\phi}(x - z) \cdot dz\) and \(q_{\phi'}(y) = \int_{y=-\infty}^{y=\infty} r(z)q_{\phi}(y - z) \cdot dz\). Now, we compute the statistical difference between \(p_{X+Z}\) and \(p_{\phi'}\):
\[
\Delta(p_{X+Z}, p_{\phi'}) = \frac{1}{2} \int_{x=-\infty}^{x=\infty} |p_{X+Z}(x) - p_{\phi'}(x)| \cdot dx
\]
Similarly, we can show \( \Delta(p_{YZ}, q_{\phi'}) = \Delta(p_Y, q_\phi) \). This, together with \( \Delta(p_X, p_{\phi'}) + \Delta(p_Y, q_\phi) \leq \gamma \) (which follows because \( \phi \in \Phi_\gamma(p_X, p_Y) \)) implies \( \Delta(p_{XZ}, p_{\phi'}) + \Delta(p_{YZ}, q_{\phi'}) \leq \gamma \). Thus, we have shown that \( \phi' \in \Phi_\gamma(p_{XZ}, p_{YZ}) \). Now, we show below that \( E_{(x,y)\sim\phi}[\Phi(x, y)] = E_{(x,y)\sim\phi'}[\Phi(x, y)] \):

\[
E_{(x,y)\sim\phi'}[\Phi(x, y)] = \int_{x=-\infty}^{x=\infty} \int_{y=-\infty}^{y=\infty} \Phi(x, y)\phi'(x, y) \cdot dx \cdot dy
\]

\[
= \int_{x=-\infty}^{x=\infty} \int_{y=-\infty}^{y=\infty} \Phi(x, y) \left[ \int_{z=-\infty}^{z=\infty} p_x(z)\phi(x-z, y-z) \cdot dz \right] \cdot dx \cdot dy
\]

\[
= \int_{z=-\infty}^{z=\infty} p_x(z) \left[ \int_{x=-\infty}^{x=\infty} \Phi(x, y)\phi(x-z, y-z) \cdot dx \right] \cdot dz
\]

\[
= \int_{z=-\infty}^{z=\infty} p_x(z) E_{(x,y)\sim\phi}[\Phi(x, y)] dz
\]

\[
= E_{(x,y)\sim\phi}[\Phi(x, y)]. \tag{28}
\]

Note that \( E_{(x,y)\sim\phi}[\Phi(x, y)] = \beta \) (by assumption) and \( W_\gamma(p_{XZ}, p_{YZ}) = \inf_{\phi' \in \Phi_\gamma(p_{XZ}, p_{YZ})} E_{(x,y)\sim\phi'}[\Phi(x, y)] \) (by definition). These, together with (28) yield \( W_\gamma(p_{XZ}, p_{YZ}) \leq \beta = W_\gamma(p_X, p_Y) \). This proves the inequality (a) of (26).

**Proof of (b).** By Lemma 2, we can show that

\[
W_\gamma(p_X, p_Y) \leq W_\gamma-2\gamma_1(p_{XZ}, p_{YZ}) + W_\gamma_1(p_X, p_{XZ}) + W_\gamma_1(p_Y, p_{YZ}).
\]

It follows from (a) that

\[
W_\gamma_1(p_X, p_{XZ}) \leq W_\gamma_1(p_0, p_Z)
\]

\[
W_\gamma_1(p_Y, p_{YZ}) \leq W_\gamma_1(p_0, p_Z).
\]

Combining the above three inequalities gives the required inequality (b) (26). This completes the proof of Lemma 3. \( \square \)

### B Omitted Details from Section 3.1

**Lemma** (Restating Lemma 5). On inputs of size at least \( n \), \( M_{trLap}^\tau \) (defined in Construction 1) is an \((\epsilon, \delta)\)-differentially private mechanism, where \( \epsilon = \frac{1}{\sqrt{\tau n}} \) and \( \delta = \frac{4}{\sqrt{\tau n}} e^{-\frac{\tau n}{2}} \).

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

Let \( q = n\tau \). Let \( x \) and \( x' \) be two neighbouring databases and let \( y = \text{hist}_p(x) \) and \( y' = \text{hist}_p(x') \). For simplicity, for every \( i \in [t] \), define \( n_i := y(i) \) and \( n'_i := y'(i) \). Since \( x \sim x' \), there
exists $g^* \in \mathcal{G}$ such that $\forall g \neq g^*, x(g) = x'(g^*)$, and $|x(g^*) - x'(g^*)| = 1$, which implies that there exists $i^* \in [t]$ such that $|n_{i^*} - n'_{i^*}| = 1$ and that $n_i = n'_i$ for every $i \neq i^* \setminus \{i^*\}$. Without loss of generality, assume that $n_{i^*} = n'_{i^*} + 1$. In order to prove the lemma, for every subset $S \subseteq \mathbb{R}^t$, we need to show that

$$
\begin{align*}
\Pr[M_{trLap}(x) \in S] &\leq e^\epsilon \Pr[M_{trLap}(x') \in S] + \delta, \\
\Pr[M_{trLap}(x') \in S] &\leq e^\epsilon \Pr[M_{trLap}(x) \in S] + \delta,
\end{align*}
$$

where $\epsilon = \frac{1}{\sqrt{q}}$ and $\delta = \frac{4e^{-\sqrt{q}}}{q}$. We only prove (29); (30) can be shown similarly.

Fix an arbitrary subset $S \subseteq \mathbb{R}^t$. Since histogram has $t$ bars, and $M_{trLap}^*$ adds independent noise in each bar according to $\pi_q(z)$, we have that for every $s = (s_1, \ldots, s_t) \in \mathbb{R}^t$, we have $p_{M_{trLap}^*(s)}(s) = \prod_{i=1}^t \pi_q(s_i - n_i)$. Thus we have

$$
\begin{align*}
\Pr[M_{trLap}(x) \in S] &= \int_{s \in S} \left[ \prod_{i=1}^t \pi_q(s_i - n_i) \right] \cdot ds, \\
\Pr[M_{trLap}(x') \in S] &= \int_{s \in S} \left[ \prod_{i=1}^t \pi_q(s_i - n_i') \right] \cdot ds.
\end{align*}
$$

Now, using the fact that $\forall k \neq i^*, n_k = n_k'$ and $n_{i^*} = n'_{i^*} + 1$, we partition $S$ into 6 disjoint sets:
1. $S_0 := \{s \in \mathbb{R}^t : -\frac{q}{2} \leq s_{i^*} - n_{i^*}, s_i - n_i' < 0\}$.
2. $S_1 := \{s \in \mathbb{R}^t : -q < s_{i^*} - n_{i^*}, s_i - n_i' \leq -\frac{q}{2}\}$.
3. $S_2 := \{s \in \mathbb{R}^t : -q < s_{i^*} - n_{i^*} < -\frac{q}{2} < s_{i^*} - n_{i^*}' \leq 0\}$.
4. $S_3 := \{s \in \mathbb{R}^t : s_{i^*} - n_{i^*} \leq -q < s_{i^*} - n_{i^*}' \leq -\frac{q}{2}\}$.
5. $S_4 := \{s \in \mathbb{R}^t : -\frac{q}{2} < s_{i^*} - n_{i^*} \leq 0 \leq s_{i^*} - n_{i^*}'\}$.
6. $S_5 := \{s \in \mathbb{R}^t : 0 \leq s_{i^*} - n_{i^*}, s_i - n_i' \text{ or } s_{i^*} - n_{i^*}, s_i - n_i' \leq -q\}$.

It can be verified that, whenever $s \in (s_1, \ldots, s_t) \in S_k, k \in [5] \setminus \{4\}$, we have, $\pi_q(s_{i^*} - n_{i^*}) \leq e^{\epsilon} \pi_q(s_{i^*} - n_{i^*}')$, where $\epsilon = \frac{1}{\sqrt{q}}$. This follows from (i) for $s \in S_0$, we have $\pi_q(s_{i^*} - n_{i^*}) = e^{\epsilon} \pi_q(s_{i^*} - n_{i^*}')$, (ii) for $s \in S_1$, we have $\pi_q(s_{i^*} - n_{i^*}) = e^{-\epsilon} \pi_q(s_{i^*} - n_{i^*}')$, (iii) for $s \in S_2$, we have $\pi_q(s_{i^*} - n_{i^*}) \leq e^{\epsilon} \pi_q(s_{i^*} - n_{i^*}')$, (iv) for $s \in S_3$, we have $\pi_q(s_{i^*} - n_{i^*}) = 0$, and (v) for $s \in S_5$, we have $\pi_q(s_{i^*} - n_{i^*}) = 0$ and $\pi_q(s_{i^*} - n_{i^*}') > 0$, which does not satisfy the desired inequality.

Define $S_{-4} := S \setminus S_4$. In order to show (29), first we show $\Pr[M_{trLap}^*(x) \in S_{-4}] \leq e^\epsilon \Pr[M_{trLap}(x') \in S_{-4}]$ and then we show $\Pr[M_{trLap}(x) \in S_4] \leq \delta$. These together imply (29).

$$
\begin{align*}
\Pr[M_{trLap}^*(x) \in S_{-4}] &= \int_{s \in S_{-4}} \left[ \prod_{i=1}^t \pi_q(s_i - n_i) \right] \cdot ds \\
&= \int_{s \in S_{-4}} \left[ \prod_{\substack{i=1 \atop i \neq i^*}}^t \pi_q(s_i - n_i) \pi_q(s_{i^*} - n_{i^*}') \right] \cdot ds \\
&\leq \int_{s \in S_{-4}} \left[ \prod_{\substack{i=1 \atop i \neq i^*}}^t \pi_q(s_i - n_i) e^{\epsilon} \pi_q(s_{i^*} - n_{i^*}') \right] \cdot ds \\
&= e^\epsilon \int_{s \in S_{-4}} \left[ \prod_{i=1}^t \pi_q(s_i - n_i) \right] \cdot ds
\end{align*}
$$
\[ e^\delta \Pr[\mathcal{M}_{\text{triLap}}^\tau(x') \in S_{-4}] \] (33)

Observe that, for every \( s \in S_4 \), we have \( -\frac{q}{2} < s_i - n_i < 0 \leq s_i - n_i' \), where \( n_i' = n_i + 1 \). This implies that \((s_i - n_i') \in [-1, 0)\). For \( i \in [t] \), define \( S_4(i) := \{ s_i : \exists s \in S_4 \text{ s.t. } s_i = s \} \).

\[
\Pr[\mathcal{M}_{\text{triLap}}^\tau(x) \in S_4] = \int_{s \in S_4} \left[ \prod_{i=1}^t \pi_q(s_i - n_i) \right] \cdot ds
\]

\[
= \int_{s_1 \in S_4(1)} \cdots \int_{s_i \in S_4(i^*)} \cdots \int_{s_t \in S_4(t)} \left[ \prod_{i=1}^t \pi_q(s_i - n_i) \right] \cdot ds_1 \cdots ds_i \cdots ds_t
\]

\[
= \int_{s_i \in S_4(i^*)} \pi_q(s_i - n_i \ast) \left( \int_{s_1 \in S_4(1)} \cdots \int_{s_t \in S_4(t)} \left[ \prod_{i=1, i \neq i^*}^t \pi_q(s_i - n_i) \right] \cdot ds_1 \cdots ds_t \right) ds_{i^*}
\]

\[
\leq \int_{s_i \in S_4(i^*)} \pi_q(s_i - n_i \ast) ds_{i^*}
\]

\[
\leq \int_{n_i \ast -1}^0 \pi_q(z) dz
\]

\[
= \frac{e^{1/\sqrt{\eta}} - 1}{2(1 - e^{-\sqrt{\eta}/2})} e^{-\sqrt{\eta}/2}
\]

\[
\leq 2(e^{1/\sqrt{\eta}} - 1)e^{-\sqrt{\eta}/2}
\]

\[
\leq \frac{4}{\sqrt{\eta}} e^{-\sqrt{\eta}/2}
\]

\[
= \delta
\]

Now, by combining (33) and (34), we show (29):

\[
\Pr[\mathcal{M}_{\text{triLap}}^\tau(x) \in S] = \Pr[\mathcal{M}_{\text{triLap}}^\tau(x) \in S_{-4}] + \Pr[\mathcal{M}_{\text{triLap}}^\tau(x) \in S_4]
\]

\[
\leq e^\delta \Pr[\mathcal{M}_{\text{triLap}}^\tau(x') \in S_{-4}] + \delta
\]

\[
\leq e^\delta \Pr[\mathcal{M}_{\text{triLap}}^\tau(x') \in S] + \delta
\]

This completes the proof of Lemma 5.

\[ \square \]

C Omitted Details from Section 4.1

Lemma (Lemma 6 Restated). If \( \hat{\partial} \) is a measure of distortion over \( A \), then \( \hat{\partial} \) is quasi-metric.

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

1. The first property follows from the definition of \( \hat{\partial} \) (see Definition 8): If \( \hat{\partial}(p, q) = 0 \), 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 is only one coupling \( \phi \) in \( \Phi(p, q) \), which is a diagonal distribution.
2. For the second property, let $\phi_2 \in \Phi(p, r)$ and $\phi_3 \in \Phi(r, q)$ denote the optimal couplings for $\hat{\partial}(p, r)$ and $\hat{\partial}(r, q)$, respectively, i.e., $\partial(p, r) = \sup_{(x,y)} \partial(x, y)$ and $\partial(r, q) = \sup_{(y,z)} \partial(y, z)$.

It follows from the Gluing Lemma [Vil08] that we can find a coupling $\phi'$ over $A \times A \times A$ such that the projection of $\phi'$ onto its first two coordinates is equal to $\phi_2$ and its last two coordinates is equal to $\phi_3$. Let $\hat{\phi}_1$ denote the projection of $\phi'$ onto its first and the third coordinates. Note that $\hat{\phi}_1 \in \Phi(p, q)$, but it may not be an optimal coupling for $\hat{\partial}(p, q)$. Now the triangle inequality follows from the following set of inequalities:

$$\hat{\partial}(p, q) = \inf_{\phi \in \Phi(p, q)} \sup_{(x,z)} \partial(x, z) \leq \sup_{(x,z)} \partial(x, z) \leq \sup_{(x,y,z)} \partial(x, y) + \sup_{(y,z)} \partial(y, z) \leq \sup_{(x,y,z)} \partial(x, y) + \sup_{(y,z)} \partial(y, z) = \hat{\partial}(p, r) + \hat{\partial}(r, q),$$

where (a) follows from the fact that $\partial$ is a measure of distortion, which is a quasi-metric.

This completes the proof of Lemma 6.

**Lemma (Lemma 7 Restated).** If $\mathcal{M} : A \rightarrow B$ is an $(\alpha, \beta, \gamma)$-accurate mechanism for $f$ w.r.t. $\partial$, then for any random variable $X$ over $A$, there is a random variable $X^*$ such that

$$\hat{\partial}(p_X, p_{X^*}) \leq \alpha, \quad W_\gamma(f(X^*), \mathcal{M}(X)) \leq \beta.$$

**Proof:** From the accuracy guarantee of $\mathcal{M}$, we have for each $x$, there is a random variable $X_x'$ such that for each $x'$ in the support of $X_x'$, $\partial(x, x') \leq \alpha$ and $W_\gamma(f(X_x'), \mathcal{M}(x)) \leq \beta$.

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

To see the second part, let $\phi_x$ denote the distribution that achieves $W_\gamma(f(X'_x), \mathcal{M}(x))$. That is, for each $x$, $\phi_x \in \Phi^i(f(X'_x), \mathcal{M}(x))$ and $W_\gamma(f(X'_x), \mathcal{M}(x)) = E_{(a,b) \sim \phi_x} [\partial(a, b)]$. Let $\phi$ be
defined by $\phi(a, b) = p_X(x)\phi_x(a, b)$. It is easy to verify that $\phi \in \Phi^\gamma(f(X^*), \mathcal{M}(X))$. Further,

$$W_\gamma(f(X^*), \mathcal{M}(X)) \leq \mathbb{E}_{(a,b) \sim \phi} [d(a, b)]$$

$$= \mathbb{E}_{x \sim p_X} \left[ \mathbb{E}_{(a,b) \sim \phi_x} [d(a, b)] \right]$$

$$= \mathbb{E}_{x \sim p_X} [W_\gamma(f(X'_x), \mathcal{M}(x))]$$

$$\leq \beta.$$ 

This completes the proof of Lemma 7. 

The following lemma translates the definition of distortion sensitivity (Definition 9) to apply to distortion of input distributions. It is stated assuming that all infima in the definition of $\sigma_j^\phi$ are achieved. This is the case, e.g., when the measure of distortion is discrete.

Lemma (Lemma 8 Restated). Suppose $f : A \to B$ has distortion sensitivity $\sigma_j^\phi$, w.r.t. $(\partial_1, \partial_2)$. Then, for random variables $X_0$ over $A$, and $Y$ over $B$ such that $\hat{\partial}_2(f(X_0), p_Y) \leq \alpha$, there exists a distribution $X$ over $A$ such that $W_\theta(f(X), p_Y) = 0$ and $\hat{\partial}_1(p_{X_0}, p_X) \leq \sigma_j^\phi(\alpha)$.

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

$$\hat{\partial}_2(f(X_0), p_Y) = \sup_{(u,y)\sim \phi} \partial_2(u, y). \quad (35)$$

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

$$\hat{\partial}_2(f(x_0), p_{Y_{x_0}}) = \inf_{\phi \in \Phi^{\phi}(f(x_0), p_{Y_{x_0}})} \sup_{(u,y)\sim \phi} \partial_2(u, y)$$

$$\leq \sup_{(u,y)\sim \phi} \partial_2(u, y)$$

$$\leq \sup_{(u,y)\sim \phi} \partial_2(u, y) \leq \alpha. \quad (a)$$

Here (a) follows from the fact that $\text{support}(\phi_{x_0}) \subseteq \text{support}(\phi)$ and (b) follows from (35). Thus for each $x_0 \in \text{support}(X_0)$, we have $\hat{\partial}_2(f(x_0), p_{Y_{x_0}}) \leq \alpha$. By the definition of $\sigma_j^\phi$, there exist $X_{x_0}$ such that,

$$W_\theta(f(X_{x_0}), p_{Y_{x_0}}) = 0, \quad \hat{\partial}_1(x_0, p_{X_{x_0}}) \leq \sigma_j^\phi(\alpha)$$

Now, define $X = \sum_{x_0 \in \text{support}(X_0)} p_{X_{x_0}} (x_0) X_{x_0}$. Now, using the similar argument as used to prove the second part of Lemma 7, in the following we show that $W_\theta(f(X), p_Y) = 0$ and $\hat{\partial}_1(p_{X_{x_0}}, p_X) \leq \sigma_j^\phi(\alpha)$.

• Showing $W_\theta(f(X), p_Y) = 0$: For each $x_0 \in \text{support}(X_0)$, let $\psi_{x_0}$ be the optimal coupling that achieves the infimum in the definition of $W_\theta(f(X_{x_0}), p_{Y_{x_0}})$. That is, for each $x_0$, $\psi_{x_0} \in$
\( \Phi^\theta(f(X_{x_0}), p_{y_{x_0}}) \) and \( W_\theta(f(X_{x_0}), p_{y_{x_0}}) = E_{(a,b) \leftarrow \psi_{x_0}} [d(a,b)] \). Let \( \psi \) be defined by \( \psi(a,b) = p_{x_0}(x_0)\psi_{x_0}(a,b) \). It is easy to verify that \( \psi \in \Phi^\theta(f(X), p_y) \). Further,

\[
W_\theta(f(X), p_Y) \leq E_{(a,b) \leftarrow \psi} [d(a,b)]
= E_{x_0 \leftarrow p_{X_0}} [E_{(a,b) \leftarrow \psi_{x_0}} [d(a,b)]]
= E_{x_0 \leftarrow p_{X_0}} [W_\theta(f(X_{x_0}), p_{y_{x_0}})]
= 0.
\]

• Showing \( \widehat{\partial}_1(p_{x_0}, p_X) \leq \sigma^\theta_f(\alpha) \): For each \( x_0 \in \text{support}(X_0) \), let \( \psi_{x_0} \) be the optimal coupling that achieves the infimum in the definition of \( \widehat{\partial}_1(x_0, p_{x_0}) \). That is, for each \( x_0, \psi_{x_0} \in \Phi^\theta(x_0, p_{x_0}) \) and \( \widehat{\partial}_1(x_0, p_{x_0}) = \sup_{(a,b) \leftarrow \psi_{x_0}} \partial_1(a,b) \). Let \( \psi \) be defined by \( \psi(a,b) = p_{x_0}(x_0)\psi_{x_0}(a,b) \). It is easy to verify that \( \psi \in \Phi^\theta(p_{x_0}, p_X) \). Further,

\[
\widehat{\partial}_1(p_{x_0}, p_X) \leq \sup_{(a,b) \leftarrow \psi} \partial_1(a,b)
= \sup_{x_0 \leftarrow p_{X_0}} \sup_{(a,b) \leftarrow \psi_{x_0}} \partial_1(a,b)
= \sup_{x_0 \leftarrow p_{X_0}} \widehat{\partial}_1(x_0, p_{x_0})
\leq \sigma^\theta_f(\alpha).
\]

This completes the proof of Lemma 8.

\[\square\]

D Omitted Details from Section 4.3

**Theorem** (Restating Theorem 3). *On an n-element database, \( M_{1D-hist}^{\alpha, \beta, B} \) (from Construction 3) is \( (\alpha, \beta, 0) \)-accurate w.r.t. the distortion measure \( \partial_{\text{drop}} \) and metric \( \delta_{\text{hist},(\cdot)} \), and \( (1/\sqrt{\alpha}, 4\sqrt{\alpha}e^{-\beta^2/2}) \)-differentially private, where \( q = 2n^{\alpha/\beta}/B \), for the histogram function.*

**Proof:** We have \( M_{1D-hist}^{\alpha, \beta, B} = M_{trLap}^{\tau, \alpha} \circ M_{buc}^{\beta, 1} \) with \( \tau = B/\sqrt{2} \), \( \alpha \), and \( \beta := (I_1, \ldots, I_t) \), where \( I_i = [B(i - 1)/t, B(i)/t] \). Let \( q = \tau n = 2n^{\alpha/\beta}/B \), and let \( f_1 \) denote the underlying function that \( M_{buc}^{\beta, 1} \) computes. Note that \( f_1 \) is the identity function.

We have from Lemma 9 that \( M_{buc}^{\beta, 1} \) is \( (0, \beta, 0) \)-accurate. Also, since \( f_1 \) is the identity function, we have from Remark 1 that \( \sigma^\theta_{f_1}(\alpha) = \alpha \) for all \( \alpha \geq 0 \), i.e., \( f_1 \) has identity \( \theta \)-distortion sensitivity for \( \theta = 0 \).

Since \( M_{trLap}^{\tau, 1} \) operates on \( t \) bars, where \( \tau \alpha = \alpha / \sqrt{2} \), we have from Lemma 4 and Lemma 5 that \( M_{trLap}^{\tau, 1} \) is \( (\alpha, 0, 0) \)-accurate w.r.t. the distortion measure \( \partial_{\text{drop}} \) and metric \( \delta_{\text{hist},(\cdot)} \), and \( (1/\sqrt{\alpha}, 4\sqrt{\alpha}e^{-\beta^2/2}) \)-differentially private, respectively. Note that \( M_{trLap}^{\tau, 1} \) can be thought of as adding multi-dimensional Laplacian noise \( Z \). Now, take an arbitrary \( \beta \geq 0 \). Then it follows from Lemma 3 that for any two input distributions \( p_X, p_Y \) over histograms, such that \( W(p_X, p_Y) \leq \beta \), we have \( W(p_{X + Z}, p_{Y + Z}) \leq W(p_X, p_Y) \leq \beta \), where the underlying metric is \( \delta_{\text{hist},(\cdot)} \). Hence \( \tau^\theta_{M_{trLap}^{\tau, 1}}(\beta) \leq \beta \), i.e., the \( \theta \)-error sensitivity is upper-bounded by the identity function for \( \theta = 0 \).  

\(^3\)Note that Lemma 4 is not stated w.r.t. the metric \( \delta_{\text{hist},(\cdot)} \), but it can be verified that it does hold for this particular metric.
Since $\mathcal{M}^{\alpha,\beta,\tau}_{1D\text{-hist}} = \mathcal{M}^{\tau,\rho}_{\text{trLap}} \circ \mathcal{M}^{\lambda}_{\text{buc}}$, we have from Theorem 1 that $\mathcal{M}^{\alpha,\beta,\tau}_{1D\text{-hist}}$ is $(\alpha, \beta, 0)$-accurate. For the privacy guarantee, first observe that $\mathcal{M}^{\lambda}_{\text{buc}}$ is a neighborhood preserving mechanism, i.e., for any two multisets $x, x'$ such that $x \sim_{\text{mset}} x'$, we have $\mathcal{M}^{\lambda}_{\text{buc}}(x) \sim_{\text{mset}} \mathcal{M}^{\lambda}_{\text{buc}}(x')$. Now, it follows from Theorem 2 that $\mathcal{M}^{\alpha,\beta,\tau}_{1D\text{-hist}}$ is $(\frac{1}{\sqrt{q}}, \frac{4}{\sqrt{q}} e^{-\frac{q^2}{2}})$-differentially private. This completes the proof of Theorem 3.

**Theorem** (Restating Theorem 4). On an $n$-element database, $\mathcal{M}^{\alpha,\beta,\lambda}_{1D\text{-supp}}$ (from Construction 4) is $(\frac{1}{\sqrt{q}}, \frac{4}{\sqrt{q}} e^{-\frac{q^2}{2}})$-differentially private, with $q = 2n\alpha\beta$, and $(\alpha, \beta, 0)$-accurate w.r.t. the distortion measure $\partial_{\text{drop}}$ and metric $\partial_{\text{hist}}(\cdot, \cdot)$ for the function $f_{\text{supp}}$.

**Proof:** We have $\mathcal{M}^{\alpha,\beta,\lambda}_{1D\text{-supp}} = f_{\text{supp}} \circ \mathcal{M}^{\alpha,\beta,\tau}_{1D\text{-hist}}$. Let $f_1$ denote the underlying function that $\mathcal{M}^{\alpha,\beta,\tau}_{1D\text{-hist}}$ computes. We have from Theorem 3 that $\mathcal{M}^{\alpha,\beta,\tau}_{1D\text{-hist}}$ is $(\frac{1}{\sqrt{q}}, \frac{4}{\sqrt{q}} e^{-\frac{q^2}{2}})$-differentially private, with $q = 2n\alpha\beta$, and $(\alpha, \beta, 0)$-accurate w.r.t. the distortion measure $\partial_{\text{drop}}$ and the metric $\partial_{\text{hist}}(\cdot, \cdot)$.

By definition $f_{\text{supp}}$ upon histograms is a deterministic and accurate mechanism for computing support of histograms. Hence, it is $(0, 0, 0)$-accurate w.r.t. distortion $\partial_{\text{drop}}$ and metric $\partial_{\text{hist}}(\cdot, \cdot)$.

Now we compute parameters required for composition. Since $\alpha_2 = 0$ for composition, we trivially have $\sigma_1^0(0) = 0$.

For any coupling $\phi$ of $X, Y$ over histograms, we define coupling $\phi'$ on $f_{\text{supp}}(X), f_{\text{supp}}(Y)$ as $Pr(\phi' = (a, b)) = \sum_{x: y: (f_{\text{supp}}(x), f_{\text{supp}}(y)) = (a, b)} Pr(\phi = (x, y))$. Since, for two histograms $y, y'$, by definition of metrics, $\partial_{\text{hist}}(f_{\text{supp}}(y), f_{\text{supp}}(y')) \leq \partial_{\text{hist}}(y, y')$. Therefore, $E(\sigma(a, b) \circ \partial_{\text{hist}}(a, b)) \leq E(\partial_{\text{hist}}(x, y))$. Therefore, we have, $W(p_{f_{\text{supp}}(x)}, p_{f_{\text{supp}}(y)}) \leq W(p_x, p_y)$, which implies that $\tau_0(\beta) \leq \beta$ for any $\beta \geq 0$.

Hence, by Theorem 1, the composed mechanism $\mathcal{M}^{\alpha,\beta,\lambda}_{1D\text{-supp}}$ is $(\alpha, \beta, 0)$-accurate w.r.t. the distortion measure $\partial_{\text{drop}}$ and metric $\partial_{\text{hist}}(\cdot, \cdot)$ for the function $f_{\text{supp}}$.

**Theorem** (Restating Theorem 5). On and $n$-element database, $\mathcal{M}^{\alpha,\beta,\lambda}_{\text{max}}$ (from Construction 5) is a $(\frac{1}{\sqrt{q}}, \frac{4}{\sqrt{q}} e^{-\frac{q^2}{2}})$-differentially private mechanism, with $q = 2n\alpha\beta$, that is $(\alpha, \beta, 0)$-accurate w.r.t. distortion measure $\partial_{\text{drop}}$ and the standard metric for $\mathbb{R}$, for the function $f_{\text{max}}$.

**Proof:** We have $\mathcal{M}^{\alpha,\beta,\lambda}_{\text{max}} = f_{\text{max}} \circ \mathcal{M}^{\alpha,\beta,\tau}_{1D\text{-hist}}$. Let $f_1$ denote the underlying function that $\mathcal{M}^{\alpha,\beta,\tau}_{1D\text{-hist}}$ computes. Again, we have from Theorem 3 that $\mathcal{M}^{\alpha,\beta,\tau}_{1D\text{-hist}}$ is $(\frac{1}{\sqrt{q}}, \frac{4}{\sqrt{q}} e^{-\frac{q^2}{2}})$-differentially private, with $q = 2n\alpha\beta$, and $(\alpha, \beta, 0)$-accurate w.r.t. distortion $\partial_{\text{drop}}$ and standard metric on $\mathbb{R}$.

We shall now compute the parameters required for composition. Since $\alpha_2 = 0$ for composition, we trivially have $\sigma_1^0(0) = 0$.

Also, as for any two histograms $y, y'$, we have by definition of the metrics that $|f_{\text{max}}(y) - f_{\text{max}}(y')| \leq \partial_{\text{hist}}(y, y')$. Now, using the same coupling as in the proof of Theorem 4, we can show that $\tau_0(f_{\text{max}}(\beta)) \leq \beta$ for any $\beta \geq 0$.

Therefore, by Theorem 1, the composed mechanism $\mathcal{M}^{\alpha,\beta,\lambda}_{\text{max}}$ is $(\alpha, \beta, 0)$-accurate w.r.t. distortion measure $\partial_{\text{drop}}$ and the standard metric for $\mathbb{R}$. Again, this composition can be viewed as
a post processing over the support mechanism, thus it is also \((\frac{1}{\sqrt{d}}, \frac{4}{\sqrt{q}} e^{-\frac{\sqrt{q}}{2}})\)-differentially private for the \(f_{\text{max}}\) function.

\[ \square \]

## E Omitted Details from Section 5

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

**Proof:** Let \(p(x) = \sum_{u \in U} q_u \text{Lap}(x|u, b)\), where \(\sum_{u \in U} q_u = 1\) and \(q_u \geq 0\). Then we have

\[
\frac{p(x)}{p(x+d)} = \frac{\sum_{u \in U} q_u \text{Lap}(x|u, b)}{\sum_{u \in U} q_u \text{Lap}((x+d)|u, b)}
\]

It follows from the definition of Laplace distribution (Definition 3) that

\[
\frac{\text{Lap}(x|\mu, b)}{\text{Lap}(x+d|\mu, b)} = e^{(|x-\mu|-|x+d-\mu|)/b} \leq e^{|d|/b},
\]

\[
\frac{\text{Lap}(x|\mu, b)}{\text{Lap}(x+d|\mu, b)} = e^{(|x-\mu|-|x+d-\mu|)/b} \geq e^{-|d|/b}.
\]

Using (37) in (36) yields \(\frac{p(x)}{p(x+d)} \leq e^{|d|/b}\) as follows:

\[
\frac{p(x)}{p(x+d)} = \frac{\sum_{u \in U} q_u \text{Lap}(x|u, b)}{\sum_{u \in U} q_u \text{Lap}((x+d)|u, b)} \leq \frac{\sum_{u \in U} q_u e^{|d|/b} \text{Lap}((x+d)|u, b)}{\sum_{u \in U} q_u \text{Lap}((x+d)|u, b)} = e^{|d|/b}
\]

Using (38) in (36) yields \(\frac{p(x)}{p(x+d)} \geq e^{-|d|/b}\) as follows:

\[
\frac{p(x)}{p(x+d)} \geq \frac{\sum_{u \in U} q_u e^{-|d|/b} \text{Lap}((x+d)|u, b)}{\sum_{u \in U} q_u \text{Lap}((x+d)|u, b)} = e^{-|d|/b}
\]

This proves Claim 1.

\[ \square \]

**Claim** (Claim 2 Restated). Suppose two distributions \(p, q\) (defined over the same alphabet) both satisfy the log-Lipschitz condition given in Claim 1, and \(W_\theta^\infty(p, q) \leq \rho\). Then \(p, q\) satisfy the \((\epsilon, \delta)\)-DP condition (i.e., for every \(S \subseteq \mathbb{R}\), we have \(\text{Pr}_{x+p}[x \in S] \leq \epsilon \text{Pr}_{x+q}[x \in S] + \delta\), where \(\epsilon = \frac{2\rho}{b} + \ln(2)\) and \(\delta = \frac{\rho^2}{b(1-e^{-\frac{\rho}{b}})}\).

**Proof:** We have from (25) that \(W_\theta^\infty(p, q) = \inf_{\Delta(p',q') \leq \theta} W_\theta^\infty(p', q)\). This, together with the assumption \(W_\theta^\infty(p, q) \leq \rho\), implies that \(\inf_{\Delta(p',q') \leq \theta} W_\theta^\infty(p', q) \leq \rho\). Let \(p'\) be the optimal distribution that achieves the infimum, which implies that \(\Delta(p, p') \leq \theta\) and \(W_\theta^\infty(p', q) \leq \rho\).
For $x \in \mathbb{R}$, define two functions $a(x)$ and $z(x)$ as follows:

$$a(x) = \max\{0, p(x) - p'(x)\} \quad (39)$$

$$z(x) = \int_{y=x+\rho}^{y=x-\rho} a(y) dy. \quad (40)$$

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 4.** $\int_{x \in \mathbb{R}} z(x) dx = 2p \int_{x \in \mathbb{R}} a(x) dx \leq 2p \theta$.

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

For the second inequality, note that $\int_{x \in \mathbb{R}} a(x) dx = \int_{x \in \mathbb{R}} \max\{0, p(x) - p'(x)\} dx = \frac{1}{2} \int_{x \in \mathbb{R}} |p(x) - p'(x)| dx = \Delta(p, p') \leq \theta$. Here (a) follows from the fact that $\int_{x : p'(x) \geq p(x)} (p'(x) - p(x)) dx = \int_{x : p'(x) < p(x)} (p(x) - p'(x)) dx$. \hfill \Box

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 $p(x') \geq p(x) - a(x)$ and by the definition of $z(x)$ from (40), we have $M_{p'} \geq M_p - z(x)$. We shall also prove the following inequality and use it in our proof.

**Claim 5.** $M_q \geq M_{p'}$.

**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 $W^\infty(p', q) \leq \theta$ 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 $M_{p'} \leq M_q$. Note that $p'$ may not satisfy the log-Lipschitz condition of Claim 1, but we do not need this in order to prove the claim. \hfill \Box

Since $p$ and $q$ satisfy the log-Lipschitz condition of Claim 1, we have that $\frac{p(y)}{p(x)} \geq e^{-\frac{|x-y|}{\theta}}$ and $\frac{q(y)}{q(x)} \leq e^{-\frac{|x-y|}{\theta}}$ hold for every $x, y \in \mathbb{R}$. We can lower-bound $M_p$ and upper-bound $M_q$ as follows:

$$M_p = \int_{x-\rho}^{x+\rho} p(y) dy \geq \int_{x-\rho}^{x+\rho} e^{-\frac{|x-y|}{\theta}} p(y) dy = p(x) \int_{x-\rho}^{x+\rho} e^{-\frac{|x-y|}{\theta}} dy = p(x) \cdot 2b(1 - e^{-\frac{x}{\theta}}) \quad (41)$$

$$M_q = \int_{x-2\rho}^{x+2\rho} q(y) dy \leq \int_{x-2\rho}^{x+2\rho} e^{-\frac{|x-y|}{\theta}} q(y) dy = q(x) \int_{x-2\rho}^{x+2\rho} e^{-\frac{|x-y|}{\theta}} dy = q(x) \cdot 2b(e^{2\frac{x}{\theta}} - 1) \quad (42)$$

By substituting the bounds on $M_p$ and $M_q$ from (41) and (42), respectively, we have that

$$p(x) \cdot 2b(1 - e^{-\frac{x}{\theta}}) \leq q(x) \cdot 2b(e^{2\frac{x}{\theta}} - 1) + z(x)$$

$$p(x) \leq q(x) \left( \frac{e^{2\frac{x}{\theta}} - 1}{1 - e^{-\frac{x}{\theta}}} \right) + \frac{z(x)}{2b(1 - e^{-\frac{x}{\theta}})}$$

$$p(x) \leq q(x) e^{\frac{x}{\theta}} \left( \frac{e^{2\frac{x}{\theta}} - 1}{e^{\frac{x}{\theta}} - 1} \right) + \frac{z(x)}{2b(1 - e^{-\frac{x}{\theta}})}$$

$$p(x) \leq q(x) e^{\frac{x}{\theta}} (e^{\frac{x}{\theta}} + 1) + \frac{z(x)}{2b(1 - e^{-\frac{x}{\theta}})}$$

$$p(x) \leq q(x) 2e^{\frac{2x}{\theta}} + \frac{z(x)}{2b(1 - e^{-\frac{x}{\theta}})}$$
which implies \( p(x) \leq e^{\epsilon}q(x) + \frac{z(x)}{2b(1-e^{-\frac{x}{b}})} \), where \( \epsilon = \frac{2\rho}{\theta} + \ln(2) \). Using this and integrating over an arbitrary subset \( S \subseteq \mathbb{R} \) and using \( \int_{x \in S} z(x) dx \leq \frac{2\rho}{\theta} \) (from Claim 4) gives

\[
\Pr_{x \sim p}[x \in S] \leq e^{\epsilon} \Pr_{x \sim q}[x \in S] + \frac{\rho\theta}{b(1-e^{-\frac{x}{b}})}.
\]

(43)

Take \( \delta = \frac{\rho\theta}{b(1-e^{-\frac{x}{b}})} \). Note that \( \delta \to 0 \) as \( \theta \to 0 \). This concludes the proof of Claim 2. \( \square \)

**Remaining proof of Lemma 11**: We show below that \( \mathbb{E}_{(y,t) \sim \phi^*}[|y-t|] \leq \frac{b}{(1-\gamma)(1 + \ln(\frac{1}{\gamma}))} \):

\[
\mathbb{E}_{(y,t) \sim \phi^*}[|y-t|] = \int_{w=-\infty}^{\infty} \int_{w=-\infty}^{w} |w| \cdot \Pr_{(y,t) \sim \phi^*}[|y-t| = w] \cdot dw
\]

\[
\begin{align*}
&= \int_{w=-\infty}^{w} w \cdot \Pr_{(y,t) \sim \phi^*}[|y-t| = w] \cdot dw \\
&= \int_{w=-\infty}^{w} w \cdot \left[ \Pr_{(y,t) \sim \phi^*}[y - t = w] + \Pr_{(y,t) \sim \phi^*}[y - t = -w] \right] \cdot dw \\
&= \int_{w=-\infty}^{\infty} w \cdot \left[ \int_{-\infty}^{\infty} \phi^*(i, i - w) \cdot di + \int_{-\infty}^{\infty} \phi^*(i, i + w) \cdot di \right] \cdot dw \\
&= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} w \cdot 2\phi^*(i, i + w) \cdot dw \right] \cdot di
\end{align*}
\]

Let \( d = b\ln(\frac{1}{\gamma}) \). By definition of \( \phi^* \), \( \phi^*(i, i + w) \) is non-zero only if \( i + w \in [i - d, i + d] \), which implies that \( w \in [-d, d] \). Using this in above gives:

\[
\mathbb{E}_{(y,t) \sim \phi^*}[|y-t|] = \int_{-\infty}^{\infty} \left[ \int_{0}^{d} w \cdot 2\phi^*(i, i + w) \cdot dw \right] \cdot di
\]

\[
= \int_{-\infty}^{\infty} x(i) \left[ \int_{0}^{d} w \cdot \frac{1}{1 - \gamma} \left( \frac{1}{b} e^{-\frac{i + w - i}{b}} \right) \cdot dw \right] \cdot di
\]

\[
= \frac{1}{b(1-\gamma)} \int_{-\infty}^{\infty} x(i) \left[ \int_{0}^{d} w \cdot \left( e^{-\frac{w}{b}} \right) \cdot dw \right] \cdot di
\]

\[
= \frac{1}{b(1-\gamma)} \int_{-\infty}^{\infty} x(i) \left[ \int_{0}^{d} w \cdot \left( e^{-\frac{w}{b}} \right) \cdot dw \right] \cdot di
\]

\[
= \frac{1}{b(1-\gamma)} \int_{-\infty}^{\infty} x(i) \left[ \int_{0}^{d/b} (bw) \cdot e^{-w} \cdot (b \cdot dw) \right] \cdot di
\]

\[
= \frac{b}{(1-\gamma)} \int_{-\infty}^{\infty} x(i) \left[ \int_{0}^{\ln(\frac{1}{\gamma})} w \cdot e^{-w} \cdot dw \right] \cdot di \quad \text{(since } d = b\ln(\frac{1}{\gamma})\text{)}
\]

\[
= \frac{b}{(1-\gamma)} \int_{-\infty}^{\infty} x(i) \left[ (-we^{-w} - e^{-w})|_{w=0}^{\ln(\frac{1}{\gamma})} \right] \cdot di
\]

\[
= \frac{b}{(1-\gamma)} \int_{-\infty}^{\infty} x(i) \left[ (1 - \gamma)(1 + \ln(\frac{1}{\gamma})) \right] \cdot di
\]

\[
= \frac{b}{(1-\gamma)} \left( 1 - \gamma(1 + \ln(\frac{1}{\gamma})) \right)
\]

32
\[ = b \left( 1 - \frac{\ln\left(\frac{1}{\gamma}\right)}{\frac{1}{\gamma} - 1} \right) \]

Note that at \( \gamma = 0 \), we have \( E_{(y,t)\epsilon-\phi^*}[|y - t|] = b \). \( \square \)