Metric Differential Privacy at the User-Level via the Earth Mover's Distance

Jacob Imola[1], Amrita Roy Chowdhury[2], and Kamalika Chaudhuri[3]

[1]University of Copenhagen
[2]University of Michigan, Ann Arbor
[3]University of California, San Diego

October 10, 2024

Abstract

Metric differential privacy (DP) provides heterogeneous privacy guarantees based on a distance between the pair of inputs. It is a widely popular notion of privacy since it captures the natural privacy semantics for many applications (such as, for location data) and results in better utility than standard DP. However, prior work in metric DP has primarily focused on the *item-level* setting where every user only reports a single data item. A more realistic setting is that of user-level DP where each user contributes multiple items and privacy is then desired at the granularity of the user's *entire* contribution. In this paper, we initiate the study of one natural definition of metric DP at the user-level. Specifically, we use the earth-mover's distance (d_{EM}) as our metric to obtain a notion of privacy as it captures both the magnitude and spatial aspects of changes in a user's data.

We make three main technical contributions. First, we design two novel mechanisms under d_{EM}-DP to answer linear queries and item-wise queries. Specifically, our analysis for the latter involves a generalization of the privacy amplification by shuffling result which may be of independent interest. Second, we provide a black-box reduction from the general unbounded to bounded d_{EM}-DP (size of the dataset is fixed and public) with a novel sampling based mechanism. Third, we show that our proposed mechanisms can provably provide improved utility over user-level DP, for certain types of linear queries and frequency estimation.

1 Introduction

Differential privacy (DP) is the state-of-the art technique that enables useful data analysis while still providing a strong privacy guarantee at the granularity of individuals (Dwork, 2006). Over nearly two decades, DP has enjoyed significant academic attention and has proven its efficacy in practical applications as well. It has been successfully deployed in diverse settings, including the US census (Abowd, 2018), Apple's iOS platform (Cormode et al., 2018), and Google Chrome (Erlingsson et al., 2014).

Intuitively, DP guarantee makes a pair of input data to be indistinguishable from each other. The standard DP guarantee requires *all* pairs of inputs to be indistinguishable thereby providing a uniform privacy guarantee to all pairs. This implies that every pair of input is considered equally

sensitive. However, many practical applications call for a more tailored privacy semantics based on the heterogeneity of the data. In particular, input pairs that are closer or more similar to each other are considered to be more sensitive. For instance, for location data, revealing the exact city of residence is far more sensitive than revealing just the country. Metric DP ($d_{\mathcal{X}}$-DP; Chatzikokolakis et al. (2013)) is a notion of DP that formally captures this heterogeneity in privacy semantics. Specifically, similarity is measured via a distance metric $d_{\mathcal{X}}$ and the privacy guarantee degrades linearly with the $d_{\mathcal{X}}$ distance between the pair of inputs. In addition to offering a more nuanced privacy definition, metric DP also improves utility compared to standard DP. This improvement stems from metric DP requiring only similar pairs of input to be indistinguishable, which results in a significantly lower noise than standard DP.

Prior work in metric DP has primarily focused on the *item-level* setting where every user only reports a single data item (for e.g., a single record in a dataset). However, in many practical applications, a user contributes multiple items to a dataset. Privacy is then desired at the granularity of the user's *entire* contribution. This has spurred a large body of work known as *user-level* DP (Amin et al., 2019; Bassily and Sun, 2023; Cummings et al., 2022; Acharya et al., 2023). However, all of this work considers only standard DP and is thus susceptible to the same limitations in utility as noted earlier. To this end, we initiate the study of one natural definition of metric DP at the user-level. In particular, we look into a specific distance metric called *earth-mover's distance* (d_{EM}; Givens and Shortt (1984)). d_{EM} measures the similarity of two distributions and is quantified by the cost of transforming one distribution to another where the cost function can be defined by *any* metric over the underlying data. Thus, d_{EM} provides a general and naturally interpretable way of measuring similarity, suitable for capturing the privacy semantics of various real-world scenarios. While there have been some prior attempts at this, these works are limited to specific settings, such as text data Fernandes et al. (2019). To the best of our knowledge, this is the *first work to propose a definition of metric DP at the user-level for a general setting via* d_{EM}.

The immediate question when applying metric DP at the user level is how to define a metric on the entire collection of a user's data. We argue that d_{EM} is particularly well-suited for this task. Recall that metric DP caters to the privacy semantics that similar data is more sensitive. But the challenge here is that the similarity between two collections (sets) of data points has to be measured along two dimensions – (1) the distance between the individual data items, and (2) the fraction of the data items in the set that are different. In particular, note that in addition to small changes in the item-wise distances, changes in a *smaller amount* of the data also indicate more similarity and hence, correspond to more sensitive information (see below for concrete examples). This necessitates a measure that can express both of these quantities as a single metric, for which d_{EM} is a natural choice. Informally, the d_{EM} between two distributions is the minimum cost of transporting one distribution to another, where the cost is determined by the quantity of data items moved multiplied by the distance (measured via $d_{\mathcal{X}}$) over which they are moved. Our resulting privacy definition, denoted as d_{EM}-DP, yields the following privacy semantics. Under d_{EM}-DP, the strength of the privacy guarantee (indistinguishability) between two pairs of inputs K, K' (sets of data items) grows inversely with τq if K' can be obtained by changing τ fraction of K by an average distance of q (Def. 3.1). d_{EM} therefore takes into account both the structure of the distributions as well as the raw difference in their values. Consequently, the parameters τ and q provide flexibility in interpretation and offer a nuanced privacy definition suitable for many practical applications. We illustrate this as follows:

Location Data. We will use our location dataset as a canonical example throughout the paper.

Suppose that the location dataset consists of daily locations of users collected over a period of time. Here, the parameter τ can be interpreted in terms of the length of the time window the change in K' pertains to, and q corresponds to the extent of change in the location. Then, d_{EM}-DP makes it harder to distinguish between locations that are (1) close to each other, and (2) collected over a smaller time window. This is natural, since locations gathered over an extended period, such as a month, may reveal routine patterns that are less sensitive than locations recorded on a single day (for instance, a single-day location might reveal a non-routine visit to a friend or hospital).

Textual Data. Consider a natural language dataset of user conversations where each user's data is represented as a set of words. Typically, word embeddings ϕ map each word into a high-dimensional space, and word similarity is measured using a distance, such as the Euclidean distance, between $\phi(x_1)$ and $\phi(x_2)$. Now, the parameter τ corresponds to what fraction of the user's conversation has changed in K' from K, while q corresponds to the extent of the changes in the textual content. Thus, two conversations are harder to distinguish if (1) there is only a fine-grained difference in their textual semantics[1], and (2) if it pertains to just a small fraction of the conversation (indicating a user rarely discussed the topic, which typically implies more sensitive information).

Graph Data. Consider a graph $G = (V, E)$ in which connections in E are private. Suppose there is additional public information in the form of a covariate $\phi : V \to \mathbb{R}^d$, which captures some auxiliary information about a user—for instance, the *interests* of a user. Here similarity between users is measured via covariate distance. The parameter τ corresponds to the fraction of a user's connections which has changed in K' from K, and the parameter q corresponds to the extent of the change in their interests. Thus, two graphs are harder to distinguish between if (1) it is a fine-grained change to the interest[2], and (2) if it pertains to only a few of the user's connections (say a small, private group of friends). This again captures natural privacy semantics as users are more likely to share common interests with their close friends than with a larger group, such as all workplace colleagues.

In a nutshell, d_{EM}-DP offers a more fine-grained privacy definition compared to standard DP, that captures real-world privacy semantics more effectively while providing better utility. For example, consider the following two instances of K' in the aforementioned example of location data – one where the data for an entire month is different and another where only a single day's data differs. Standard DP treats both cases as equally sensitive (i.e., offers the same privacy guarantee for both cases), necessitating a larger noise addition even in the latter case, which results in reduced utility. In contrast, d_{EM}-DP offers a stronger privacy guarantee for the latter, thereby resulting in a better privacy-utility trade-off. d_{EM}-DP is in fact a natural relaxation of standard DP. We provide more details on the interpretation of d_{EM}-DP relative to standard DP in Sec. 8. A full version of this paper appears in Imola et al. (2024).

1.1 Details of Our Contributions

We consider n users who hold datasets $\{K_i\}_{i=1}^n$, each containing elements from a *data domain* $\mathcal{X}$ of size $k = |\mathcal{X}|$. Let $d_{\mathcal{X}}$ denote a distance metric defined over $\mathcal{X}$. WLOG, we consider $d_{\mathcal{X}}$ to be a normalized distance metric, i.e., all measures of distance are normalized to be at most 1. Let $\tilde{K}_i$ denote the normalized version of the dataset K_i. d_{EM} between any pair of datasets $\{K_i, K_i'\}$ can be defined by first normalizing them to $\{\tilde{K}_i, \tilde{K}_i'\}$, and then using $d_{\mathcal{X}}$ to measure the minimum cost

[1]Such as transitioning from text about algebra to trigonometry versus changing it from "math" to "classical music".

[2]for instance, shifting from movies featuring Dwayne Johnson to Vin Diesel instead of from "action" to "rom-com"

of transporting $\tilde{K}_i$ to $\tilde{K}_i'$. The global dataset is given by $K_G = K_1 \cup \cdots \cup K_n$, and there is an aggregator who wants to privately compute a query $V(K)$. In the **central model**, the aggregator already holds K_i from each user, and applies a private mechanism $\mathcal{M}(K_G)$ to obtain a private estimate for V. In the **local model**, the users do not trust the aggregator, and communicate private messages $\{m_i = \mathcal{M}_i(K_i)\}$ to the aggregator. The aggregator then post-processes these messages $\mathcal{V}(m_1, \ldots, m_n)$ to output a private estimate of V. For simplicity, in this work we assume the mechanisms $\mathcal{M}_i$ to be non-interactive.

We also make a distinction between *bounded* and *unbounded* data. Note that boundedness here refers to the size of each user's dataset and *not* the number of the users – throughout the paper, we assume that the number of users, n, is fixed and publicly known. In our specific context, bounded data corresponds to the case where the size of each user's dataset is publicly known, and the mechanism $\mathcal{M}$ only needs to preserve privacy between datasets of the same size. Furthermore, in the central model, each user's dataset has the *same* public size. The benefit of this simplification is that algorithm analysis is easier. Such a bounded data setting has been considered in many previous works (see Li et al. (2016)). We also consider the general unbounded data setting where each user can have datasets of varying sizes, with the size being *private* as well.

For each model and type of boundedness, we summarize how one would apply d_{EM}-DP, along with the resulting semantics, in Table 1. We also include a corresponding notion of the standard user-level DP Liu et al. (2023) (provides a uniform privacy guarantee to all pairs of datasets) which serves as our baseline. In what follows, we elaborate on our main contributions.

1.1.1 Mechanism Design

We provide novel mechanisms for answering two types of queries for d_{EM}-DP. In this section, we let K denote a general dataset of interest, which in the local model would be set to K_i and in the global model would be set to K_G.

Linear Query

First, we study how to release linear queries $F\tilde{K}$, where $F \in \mathbb{R}^{d \times |\mathcal{X}|}$ is a real-valued matrix with bounded entries. While computing the global sensitivity of such a query is easy under user-level DP, proving a sensitivity bound under d_{EM}-DP is considerably more complex. It involves making a stronger Lipschitz assumption about the points in F, and then leveraging this property to transport $\tilde{K}$ onto $\tilde{K}'$ to compute an upper bound on how much $F\tilde{K}$ can change by. To this end, we first prove the following bound:

Theorem 1.1. *(Informal version of Theorem 4.1): The sensitivity of $F\tilde{K}$ is upper bounded by*

$$\max_{K,K'} \frac{\|F\tilde{K} - F\tilde{K}'\|}{d_{EM}(\tilde{K}, \tilde{K}')} \leq \max_{x,x' \in \mathcal{X}} \frac{\|F[x] - F[x']\|}{d_{\mathcal{X}}(x, x')},$$

where the notation $F[x]$ indicates the column of F indexed by x.

Using the above result, we show that the sensitivity of F, which is a maximum over the space of all datasets, can be reduced to the Lipschitzness of F, which is simpler to bound. Sensitivity analysis for standard DP typically only requires F to be bounded. However, in the case of d_{EM}-DP, we need a stronger assumption – F must also be Lipschitz. In Section 4.1, we demonstrate that several commonly used queries, such as average, region, and similarity queries, are indeed Lipschitz.

Model	Granularity	Data Boundedness	Privacy Guarantee	Semantics	Notes
Local (applies to each $\mathcal{M}_i$)	User	Unbounded	(ε, δ)-user-level DP (Def. 2.1)	Two input datasets $K, K' \in \mathcal{X}^*$ are indistinguishable with parameters (ε, δ)	Recently proposed in Acharya et al. (2023). Acts our baseline for the local model.
	User	Bounded	(α, δ)-bounded d_{EM}-DP (Def. 3.1)	Two input datasets $K, K' \in \mathcal{X}^m$ are indistinguishable with parameters $(\alpha d_{\mathsf{EM}}(\tilde{K}, \tilde{K}'), \delta)$.	The size of each dataset, m, is public. Proofs of privacy easier due to Lemma 2.1.
	User	Unbounded	(α, δ)-unbounded d_{EM}-DP (Def. 3.1)	Two input datasets $K, K' \in \mathcal{X}^*$ is indistinguishable with parameters $(\alpha d_{\mathsf{EM}}(K, K'), \delta)$.	Implies user-level DP when $\alpha \leq \varepsilon$ since $d_{\mathsf{EM}}(\cdot, \cdot) \leq 1$.
	User	Unbounded	(ε, δ, r)-discrete d_{EM}-DP (Def. 5.1)	Two input datasets $K, K' \in \mathcal{X}^*$ such that $d_{\mathsf{EM}}(\tilde{K}, \tilde{K}') \leq r)$ are indistinguishable with parameters (ε, δ)	Using group privacy, can show $(\varepsilon\lceil\frac{d}{r}\rceil, \delta \exp(\varepsilon\lceil\frac{d}{r}\rceil)$ for two K, K' s.t. $d_{\mathsf{EM}}(\tilde{K}, \tilde{K}') \leq d$.
	Item	N/A	(α, δ)-$d_{\mathcal{X}}$-DP (Def. 2.3)	Two input items $x, x' \in \mathcal{X}$ is protected with parameters $(\alpha d_{\mathcal{X}}(x, x'), \delta)$	Proposed in Chatzikokolakis et al. (2015).
Central (applies to $\mathcal{M}$)	User	Unbounded	(ε, δ)-user-level DP (Def. 2.2)	Let $K_G = K_1 \cup \cdots K_n$ where $K_i \in \mathcal{X}^*$. Two input global datasets K_G, K'_G s.t. they differ only on the dataset of a single user $\{K_i, K'_i\}, i \in [n]$ are indistinguishable with parameters (ε, δ)	Studied widely Bassily and Sun (2023); Liu et al. (2020, 2023). Acts our baseline for the central model.
	User	Bounded	(α, δ)-bounded d_{EM}-DP (Def. 3.2)	Two input global datasets K_G, K'_G s.t. they differ only on $\{K_i, K'_i\} \in \mathcal{X}^m \times \mathcal{X}^m$ are indistinguishable with parameters $(\alpha d_{\mathsf{EM}}(\tilde{K}_i, \tilde{K}'_i), \delta)$	Each K_i has size m which is public.
	User	Unbounded	(ε, δ, r)-discrete d_{EM}-DP (Def. 5.2)	Two input global datasets K_G, K'_G s.t. they differ only on $\{K_i, K'_i\}$ and $d_{\mathsf{EM}}(\tilde{K}_i, \tilde{K}'_i) \leq r$ are indistinguishable with parameters (ε, δ).	Using group privacy, can show parameters $(\varepsilon\lceil\frac{d}{r}\rceil, \delta \exp(\varepsilon\lceil\frac{d}{r}\rceil))$ for any K_i, K'_i s.t. $d_{\mathsf{EM}}(\tilde{K}_i, \tilde{K}'_i) \leq d$. Implies user-level DP when $r \geq 1$.

Table 1: Summary of privacy definitions for this paper. The number of users, n, is fixed and publicly known for all the definitions.

Unordered Release of Item-wise Queries

We design a mechanism for performing itemwise queries on the *entire* dataset K. For now we consider bounded DP, so $|K|$ is known in advance. Our approach is simple – apply a private mechanism $\mathcal{A}$ to each item $x_j \in K$ and then release the set of noisy outputs $\{\mathcal{A}(x_j)\}$ after shuffling them. Here $\mathcal{A}$ can be an arbitrary mechanism satisfying $(\alpha, 0)$-$d_{\mathcal{X}}$ DP which makes our mechanism completely general-purpose (see Section 4.2 for some concrete examples of $\mathcal{A}$). The main technical novelty lies in the privacy analysis of the above mechanism. While one can use composition to show this release satisfies $O(\alpha m)$-d_{EM} DP Fernandes et al. (2019), this is far from being tight. As discussed in Section 4.2, composition is not the right tool for tight privacy analysis since it does not account for the fact that the output of our mechanism is an unordered list, i.e., the $\mathcal{A}(x_j)$s are released in a random order. Instead, we leverage privacy amplification by shuffling (Feldman et al., 2022) to prove a general shuffling result and adapt it for d_{EM}-DP.

Theorem 1.2. *(Informal version of Theorem 4.3) Suppose that $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ is an α-$d_{\mathcal{X}}$ DP algorithm with respect to $d_{\mathcal{X}}$. Let $(x_1, \ldots, x_m) \in \mathcal{X}^m$ be a dataset. Then, releasing $\mathsf{Shuffle}(\mathcal{A}(x_1), \ldots, \mathcal{A}(x_m))$ satisfies $(O(\alpha\sqrt{me^{\alpha}\ln(m/\delta)}), \delta e^{\alpha})$-$d_{\mathsf{EM}}$ DP.*

This analysis reduces the cost of releasing m points in the multiset from $m\alpha$ to $\sqrt{m}\alpha$, allowing for better utility. We keep the analysis general – we consider releasing the shuffled multiset of any black-box mechanism $\mathcal{A}$, that satisfies metric DP in the data domain $\mathcal{X}$, applied to each data point. Consequently, this result has broader applications to the shuffle model of privacy, and may be of independent interest.

1.1.2 Extending d_{EM}-DP to the Unbounded Setting

We start our mechanism designs by considering the bounded data setting in both the local and central models of privacy (see Table 1) as this enables easier privacy analysis (Section 4). However, the bounded setting might be restrictive in practice as it cannot support usecases where users have different amounts of data, or the data sizes are also private. To this end, we extend d_{EM}-DP to the more general unbounded setting. We achieve this through a general reduction that converts any bounded d_{EM}-DP mechanism into an unbounded one while treating the mechanism as a black box. If the data from each user is relatively homogeneous, such as being i.i.d., then the utility of the mechanism will be preserved.

Our reduction uses a projection mechanism that projects any dataset onto one with a fixed size, without significantly increasing the d_{EM} distance. The projection we use is *sampling with replacement*. Intuitively, this is a smooth projection because we can view sampling from two datasets K, K' in terms of a coupling between them, and show that the sampled points can also be coupled with expected cost given by $d_{\mathsf{EM}}(K, K')$. Using Bernstein's inequality, we show convergence to $d_{\mathsf{EM}}(K, K')$, up to a small additive factor.

One caveat is that the introduced additive factor necessitates a slight adjustment to the privacy semantics of d_{EM}-DP. Whereas d_{EM}-DP protects a change of d_{EM}-distance d with an effective privacy parameter αd, for all $d > 0$, our modified definition includes r as a fixed parameter, and states that all changes of $d_{\mathsf{EM}} \leq r$ are protected uniformly with parameter αr. We may view this as a discretization of the d_{EM} by rounding it up to the nearest multiple of r. We refer to this notion as $(\alpha r, \delta, r)$-*discrete* user-level d_{EM}-DP (Def. 5.1). This privacy guarantee is weaker than d_{EM}-DP only for changes of $d_{\mathsf{EM}} < r$—for bigger changes, the two definitions are equivalent up to factors of

2 using group privacy. Typically, r is small and the practical difference between the definitions is negligible. Our reduction satisfies the following:

Theorem 1.3. *(Informal version of Theorem 5.3) Suppose that for n users, $\mathcal{M}$ is a mechanism which satisfies (α, δ)-bounded d_{EM}-DP. The algorithm which, given arbitrary user datasets $K_1, \ldots, K_n$, takes s i.i.d. samples from each K_i and then applies $\mathcal{M}$ on each of the sampled data items, satisfies $(\alpha r, \delta, r)$-discrete d_{EM}-DP (in the central model) for all $r \geq \frac{2\ln(1/\delta)}{s}$.*

The two notions of privacy are nearly equivalent for small r, showing that unbounded d_{EM}-DP can be reduced to bounded d_{EM}-DP with an almost exact translation of the privacy guarantee.

1.1.3 Demonstrating Improvements Over User-level DP

Finally, we evaluate the benefit of using the more nuanced privacy semantics of d_{EM}-DP over standard user-level DP by comparing the privacy and utility of our proposed mechanisms with baselines. Specifically, in Sec. 6.1, we study a special type of linear query called *linear embedding queries* and in Sec. 6.2, we study private frequency estimation. For simplicity, we consider the bounded data setting.

Let's start by understanding the relationship between (α, δ)-d_{EM}-DP and (ε, δ)-user-level DP. The following observations hold in both the central and local models:

- $\alpha = \varepsilon$: Since we assume $d_{\mathcal{X}}$ is normalized, we always have $d_{\mathsf{EM}} \leq 1$. Thus, in this case (ε, δ)-d_{EM}-DP implies (ε, δ)-user-level DP. However, any pair of input K, K' such that $d_{\mathsf{EM}}(\tilde{K}, \tilde{K'}) < 1$ the privacy protection of d_{EM}-DP is actually stronger. (Note that if $\alpha \leq \varepsilon$, then user-level DP is strictly weaker than d_{EM}-DP; the more appropriate baseline is to use $\alpha = \varepsilon$.)

- $\alpha > \varepsilon$: In this case, some pairs of inputs (with a large d_{EM} distance between them) are protected less strongly than they are under user-level DP. However, as indicated in our aforementioned real-life examples, input pairs with high d_{EM} (i.e., dissimilar input pairs) are typically less sensitive.

Now, we interpret the theoretical error bounds for linear embedding queries. From Table 2a, the error for releasing a d-dimensional linear embedding query under user-level DP is $O(\frac{d}{\varepsilon n})$, while it is $O(\frac{d}{\alpha n})$ for d_{EM}-DP. When $\alpha = \varepsilon$, these utilities are identical, but d_{EM}-DP offers stronger privacy. When $\alpha > \varepsilon$, then the utility of d_{EM}-DP is higher than that of user-level DP, with the the two guarantees offering differing privacy semantics. Thus, in both cases, there is a clear benefit of using d_{EM}-DP. These observations are the same in the local model.

Finally, for frequency estimation in the local model, Table 2b shows that the error of user-level DP is $O(\sqrt{\frac{k^2 \ln(m/\delta)}{n\varepsilon^2}})$, while it is $O(\sqrt{\frac{k^3}{n\alpha^2} \max\{\ln(\frac{m}{\delta}), \alpha\}}$ for d_{EM}-DP. For constant ε and $\alpha \geq \varepsilon^2 \frac{k}{\ln(m/\delta)}$, the utility is improved. In the central model, the error of the user-level DP algorithm is $O(\frac{k}{n\varepsilon})$ while it is $O(\frac{k^{3/2}}{n\alpha}\sqrt{\max\{\ln(\frac{m}{\delta}), \alpha\}})$ for d_{EM}-DP. The algorithm under d_{EM}-DP has the added benefit that it can be implemented in the shuffle model of privacy, which requires less trust and parallels prior work in the shuffle model Feldman et al. (2022). There is a utility improvement for $\alpha \geq \varepsilon^2 k$. When $\varepsilon \leq \alpha \leq \varepsilon^2 k$, we leave it as an interesting open problem whether d_{EM}-DP can offer utility improvements over user-level DP.

Linear Embedding Queries

Algorithm	Privacy Guarantee	Privacy Model	ℓ_2 Error		Notes
K-norm Mechanism	$(\varepsilon, 0)$-user level DP[♭]	Central, Bounded	$O(\frac{d}{\varepsilon n})$	(Lemma 6.3)	d_{EM}-DP gives same utility but stronger privacy for $\alpha = \varepsilon$; d_{EM}-DP gives better utility but different privacy for $\alpha > \varepsilon$.
PrivEMDLinear	(α, δ)-d_{EM}-DP	Central, Bounded	$O(\frac{d}{\alpha n}\sqrt{\ln \frac{1}{\delta}})$	(Lemma 6.2)	

[♭]For random linear queries, this algorithm also has best-known error among all (ε, δ)-user-level DP algorithms.

(a) Comparison of d_{EM}-DP to user-level DP in the central model for releasing a d-dimensional linear embedding query. The errors in the local model are a factor $\sqrt{n}$ higher.

Frequency Estimation

Algorithm	Privacy Guarantee	Privacy Model	d_{EM} Error		Notes
Hadamard Response	$(\varepsilon, 0)$-user-level DP[♭]	Local, Bounded	$O\left(\sqrt{\frac{k^2 \ln(m/\delta)}{n\varepsilon^2}}\right)$	(Lemma 6.4)	Assuming $k, \varepsilon, \alpha \leq \sqrt{m}$; d_{EM}-DP gives better utility for $\alpha \geq \varepsilon^2 \frac{k}{\ln(m/\delta)}$.
PrivEMDItemWise	(α, δ)-d_{EM}-DP	Local, Bounded	$O\left(\sqrt{\frac{k^3}{n\alpha^2} \max\left\{\ln(\frac{m}{\delta}), \alpha\right\}}\right)$	(Thm. 6.6)	
Laplace Mechanism	$(\varepsilon, 0)$-user-level DP[♭]	Central, Bounded	$O\left(\frac{k}{n\varepsilon}\right)$	(Lemma 6.7)	Assuming $n < \frac{m}{\alpha}$; d_{EM}-DP gives better utility when $\alpha > \varepsilon^2 k$.
PrivEMDItemWise	(α, δ)-d_{EM}-DP	Central, Bounded	$O\left(\frac{k^{3/2}}{n\alpha}\sqrt{\max\left\{\ln(\frac{m}{\delta}), \alpha\right\}}\right)$	(Corollary 6.8)	

[♯]This algorithm works in the shuffle model, which requires less trust than the central model.

[♭]This algorithm also has best-known error among all (ε, δ)-user-level DP algorithms.

(b) Comparison of d_{EM}-DP to user-level DP for frequency estimation in the setting defined in Section 6.2. k is the domain size $|\mathcal{X}|$.

Table 2: Summary of theoretical utility guarantees, assuming there are n users who hold datasets of size m.

2 Background

2.1 Differential Privacy

Intuitively, DP is a property of a mechanism which ensures that its output distribution remains insensitive to changes in the data of a single individual. The standard DP guarantee, which is also know as *item-level* DP, considers each user U_i to contribute only a single item $x_i \in \mathcal{X}$ to a global dataset, i.e., $K_i = x_i$. Instead we consider user-level DP, where each K_i is itself a multiset of elements from $\mathcal{X}$. We denote the set of all multisets of elements from $\mathcal{X}$ as $\mathcal{X}^*$. In the local model, our privacy definition is then:

Definition 2.1 (Unbounded User-level Local DP Acharya et al. (2023))**.** *We say a mechanism $\mathcal{M}$ acting on a dataset K satisfies (ε, δ)-unbounded user-level local DP if, for all $K, K' \in \mathcal{X}^*$ and all outputs O*

$$\Pr[\mathcal{M}(K) = O] \leq e^{\varepsilon} \Pr[\mathcal{M}(K') = O] + \delta. \tag{1}$$

Note that here we consider the more general unbounded data setting where the two datasets $\{K, K'\}$ can have arbitrary sizes.

Next, we present the definition for the central model.

Definition 2.2 (Unbounded User-level Central DP Liu et al. (2023))**.** *Let $K_G = K_1 \cup \cdots \cup K_n$ denote a global dataset from n users where $\forall i \in [n], K_i \in \mathcal{X}^*$. We say $K_G \sim K'_G$, if K'_G can be obtained from K_G by changing the dataset of a single user U_i from K_i to K'_i. We say a mechanism $\mathcal{M}$ acting on a dataset K satisfies (ε, δ)-unbounded user-level central DP if, for all K_G, K'_G such*

that $K_G \sim K'_G$, and all outputs O

$$\Pr[\mathcal{M}(K_G) = O] \leq e^{\varepsilon} \Pr[\mathcal{M}(K'_G) = O] + \delta. \tag{2}$$

Note that there is no restriction on the sizes of the datasets $\{K_i\}, i \in [n]$ in the above definition.

Next, we define metric DP that enables the privacy guarantee to depend on a metric $d_{\mathcal{X}}$ between the pair of inputs. We start by introducing it at the item-level (i.e., we consider changing only one item $x \in \mathcal{X}$ to another item $x' \in \mathcal{X}$). For simplicity, we consider the local model, so the mechanism acts on just a single item:

Definition 2.3 (Local $d_{\mathcal{X}}$-DP Alvim et al. (2018))**.** *We say $\mathcal{M}$ satisfies (α, δ)-local $d_{\mathcal{X}}$-DP if for all data elements $x, x' \in \mathcal{X}$, and all outputs O*

$$\Pr[\mathcal{M}(x) = O] \leq e^{\alpha d_{\mathcal{X}}(x,x')} \Pr[\mathcal{M}(x') = O] + \delta.$$

We replace the traditional privacy parameter ε with α in the above definition, because ε in Definitions. 2.1 and 2.2 is a unitless parameter while α has the inverse unit of $d_{\mathcal{X}}$.

2.2 Earth-Mover's Distance

Notations. We view datasets as multisets of elements from $\mathcal{X}$. We will also view a dataset $K \in \mathcal{X}^*$ as a *probability distribution* defined by its normalized histogram $\tilde{K}$. To do so, let $\Delta^{\mathcal{X}} \subseteq \mathbb{R}^{\mathcal{X}}$ denote the probability simplex indexed by $\mathcal{X}$—i.e. the set of all vectors $\langle v_x \rangle_{x \in \mathcal{X}}$ such that $v_x \geq 0$ and $\sum_{x \in \mathcal{X}} v_x = 1$. For a dataset K, $\tilde{K} \in \Delta^{\mathcal{X}}$ then denotes the probability distribution defined by K, meaning $\tilde{K}[x] = \frac{\text{Num. occurrences of } x \text{ in } K}{|K|}$. The earth-mover's (or 1-Wasserstein) distance Givens and Shortt (1984) is defined as follows. For a joint distribution $C(x_1, x_2) \in \Delta^{\mathcal{X} \times \mathcal{X}}$, let $C_{x_1}(x_2)$ denote the distribution conditioned on observing x_1, and let $C_1(x_1)$ denote the marginal distribution of x_1. We define $C_{x_2}(x_1)$ and $C_2(x_2)$ similarly.

Definition 2.4. *For distributions $P, Q \in \Delta^{\mathcal{X}}$, a joint distribution C on $\mathcal{X} \times \mathcal{X}$ is a coupling between P and Q if $C_1 = P$ and $C_2 = Q$. We let $\mathcal{C}(P, Q)$ denote the set of couplings between P and Q.*

A coupling C can be viewed as a "transportation plan" between P and Q, in the sense that if C places m probability mass at a point (x_1, x_2), then m probability mass from P at x_1 is transported to Q at x_2 (or vice-versa). We define the *cost* of a coupling as the expected transportation distance given by $\mathbb{E}_{(x,x') \sim C} d_{\mathcal{X}}(x, x')$. The *earth-mover's distance* (d_{EM}) between P, Q is equal to the minimum possible cost of a coupling between P and Q:

$$d_{\mathsf{EM}}(P, Q) = \inf_{C \in \mathcal{C}(P,Q)} \mathbb{E}_{(x,x') \sim C} \, d_{\mathcal{X}}(x, x').$$

Since we assume that $d_{\mathcal{X}}$ is bounded by 1, we have $d_{\mathsf{EM}}(\cdot, \cdot) \leq 1$.

Next, we present the Birkhoff-Von Neumann theorem which is useful in our privacy analysis in Section 4.2. The theorem states that if both P and Q are empirical distributions with the same number of points, then the d_{EM} between them is the cost of the coupling that moves the entire mass in each point to the same destination:

Lemma 2.1. *[Birkhoff-Von Neumann Theorem Konig (2001), Lemma A.1 in Fernandes et al. (2019)): For two datasets $K = \{x_1, \ldots, x_m\}$ and $K' = \{y_1, \ldots, y_m\}$, there is a permutation $\pi : [m] \to [m]$ such that*

$$d_{EM}(\tilde{K}, \tilde{K'}) = \frac{1}{m} \sum_{i=1}^{m} d_{\mathcal{X}}(x_i, y_{\pi(i)}). \tag{3}$$

3 Definition of d_{EM}-DP

In this section, we introduce our generalization of metric DP to the user-level. We start with the local model. We use the d_{EM} metric to measure the distance between two datasets K, K' since it captures the intuition that the changes which move smaller amounts of data by smaller distances are more sensitive (as discussed in Section 1).

Definition 3.1 ((Un)Bounded Local d_{EM}-DP)**.** *Let $\mathcal{M}$ be a mechanism which acts on a dataset K. We say $\mathcal{M}$ satisfies (α, δ)-**bounded** local d_{EM}-DP if for any two datasets K, K' such that $|K| = |K'|$, and for any output O, we have*

$$\Pr[\mathcal{M}(K) = O] \leq e^{\alpha d_{EM}(\tilde{K}, \tilde{K'})} \Pr[\mathcal{M}(K') = O] + \delta. \tag{4}$$

*If the above equation holds for all datasets K, K', regardless of whether $|K| = |K'|$, we say that $\mathcal{M}$ satisfies (α, δ)-**unbounded** local d_{EM}-DP.*

For bounded d_{EM}-DP, the size of the dataset is *not* protected, which is acceptable for applications where the amount of data is not sensitive. We explicitly differentiate between bounded and unbounded data since privacy analysis is easier under bounded d_{EM}-DP by leveraging Lemma 2.1 (see Section 4).
In the central model, our goal is to protect changes in a single user's dataset, transitioning from K_i to K_i', with a privacy guarantee that depends on $d_{\mathsf{EM}}(\tilde{K}_i, \tilde{K}_i')$. We consider the bounded data setting where each dataset K_i has a publicly known fixed size m.

Definition 3.2 (Bounded Central d_{EM}-DP)**.** *Let $K_G = K_1 \cup \cdots \cup K_n$ denote a global dataset from n users where $\forall i \in [n], K_i \in \mathcal{X}^m$. We say $K_G \sim K_G'$ if K_G' can be obtained from K_G by changing the dataset at a single index i from K_i to K_i'. We say a mechanism $\mathcal{M}$ satisfies (α, δ)-**bounded** central d_{EM}-DP if, for all K_G, K_G' such that $K_G \sim K_G'$, and all outputs O, we have*

$$\Pr[\mathcal{M}(K_G) = O] \leq e^{\alpha d_{EM}(\tilde{K}_i, \tilde{K}_i')} \Pr[\mathcal{M}(K_G') = O] + \delta.$$

In the above definition, the two global datasets K_G, K_G' are indistinguishable with a privacy parameter $\alpha d_{\mathsf{EM}}(K_i, K_i')$. Since we consider the bounded data setting, neither the number of total users, n, nor the size of the individual datasets, m, are protected.

It is important to note that the above definition cannot be directly translated to the unbounded data setting. This limitation arises from the fact that if each K_i is allowed to have an arbitrary size, then changing a single K_i could potentially change the entirety of K_G in the worst-case (where user U_i contributes the entire global dataset). This essentially reduces the central model (Def. 3.2) to the local model (Def. 3.1). We circumvent this challenge and provide a privacy definition for the unbounded data setting in Section 5, by controlling the amount of data from each user.

Setting the Privacy Parameters. There are some semantic differences between the parameter α in Defns. 3.1 and 3.2, and ε in Defns. 2.1 and 2.2. The privacy parameter ε is unitless. On the other hand, α is not unitless – it has a unit inversely proportional to d_{EM}. While $\varepsilon \gg 1$ is usually not considered acceptable for standard DP, it is not unreasonable to set $\alpha \gg 1$ in our case. This is acceptable if a strong privacy guarantee is needed only for input pairs that are close to each other since $d_{\mathsf{EM}}(\cdot, \cdot) < 1$. For all $q, \tau \in [0, 1]$, let $\mathcal{E}(q, \tau)$ refer to the minimum privacy parameter that is acceptable over all data changes of the form

A τ-fraction of K is changed by average distance q.

Then, α may be set as $\alpha = \inf_{q,\tau \in [0,1]} \frac{\mathcal{E}(q,\tau)}{q\tau}$, and we can verify that Defn. 3.1 will protect an input pair with the corresponding budget $\mathcal{E}(q,\tau)$. The parameter δ has the same interpretation as in standard DP, and should be set $\delta \ll \frac{1}{poly(n)}$.

Concrete Example. Throughout this paper, we consider a dataset of $n = 10^5$ users, each of whom contributes $m = 10^3$ location data points over the period of a month. We use the length of the shortest path on earth's surface as our metric $d_{\mathcal{X}}$. Suppose we want to protect a user's location over any particular day within a radius of 1000 miles, and the user's location over the entire time period within a distance of 100 miles. In the normalized metric space, these distances are $q_1 = 0.08$ and $q_2 = 0.008$, respectively[3]. They correspond to a fraction $\tau_1 = \frac{1}{30}$ and $\tau_2 = 1$ of the metric space changing, respectively. Suppose we want to protect both of these inputs with privacy parameter $\varepsilon = 0.2$. Hence, we set $\alpha = \min\left\{\frac{\varepsilon}{\tau_1 q_1}, \frac{\varepsilon}{\tau_2 q_2}\right\} = 25$. This value is much higher than typical privacy parameters used in DP, and yet it is able to adequately protect the desired inputs. Finally, we will set $\delta = 10^{-12}$ in our examples.

4 Mechanisms for d_{EM}-DP

Now, we describe our mechanisms for releasing queries under d_{EM}-DP. Throughout this section, we focus on the bounded data setting, and consider both the local and central models by considering a general dataset K which will be set to be K_i in the local model and K_G in the central one. In Section 4.1, we show how to bound the sensitivity of *linear queries*, which can then be released with the addition of calibrated noise. Then, in Section 4.2, we show that we can release a noisy representation of $\tilde{K}$ under d_{EM}-DP by applying *any* $d_{\mathcal{X}}$-DP mechanism to each item in K, and shuffling the outputs. Full proofs for this section appear in Appendix B.

4.1 Linear Queries

A non-adaptive linear query on a dataset K computes the value of $F\tilde{K}$, where $F \in \mathbb{R}^{d \times |\mathcal{X}|}$ is a matrix with d rows. The *linearity* comes from the linear transformation F; our linear queries are normalized since they operate on $\tilde{K}$ rather than K. Such normalized queries can be used for answering the fraction of users satisfying a predicate Blum et al. (2013). Nevertheless, one can estimate the non-normalized query by multiplying by an estimate of $|K|$.

Let us represent F by a function $f : \mathcal{X} \to \mathbb{R}^d$ where $f(x) = F[x]$, the xth column of F. The linear query can then be re-written as

$$q_f(K) = \mathbb{E}_{x \sim \tilde{K}}[f(x)]. \tag{5}$$

Thus, we may interpret a linear query on $\tilde{K}$ as *expected* value of f over a random item from K. Linear queries are simple but capable of expressing many indispensible tools in data analysis, and they are well-studied in differential privacy (Blum et al., 2013; Hardt and Talwar, 2010; Dwork et al., 2014). We will design a simple mechanism satisfying d_{EM}-DP for releasing a linear query, based on bounding the sensitivity of q_f under the d_{EM}. The sensitivity measures the maximum change

[3]The maximum surface distance between two points on Earth is ≈ 12930 miles.

output q_f, measured according to some norm $\|\cdot\|$ on $\mathbb{R}^d$, relative to a change in the inputs by a certain d_{EM}. This is given by:

$$\Delta_{\mathsf{EM}}(q_f) = \max_{K,K' \in \mathcal{X}^*} \frac{\|q_f(K) - q_f(K')\|}{d_{\mathsf{EM}}(\tilde{K}, \tilde{K}')}.$$

Naively, it is intractible to compute this sensitivity since there are exponentially many datasets of a given size. Additionally, this sensitivity might not always be bounded. For instance, consider two points x, x' that are close in $\mathcal{X}$, but $f(x)$ is very far from $f(x')$. In this case, we cannot bound Δ_{EM}, since the K and K' which put all their mass on x and x', respectively, will have $\frac{\|q_f(K)-q_f(K')\|}{d_{\mathsf{EM}}(\tilde{K},\tilde{K}')} = \|f(x) - f(x')\|$. We may exclude this case by assuming that f is ℓ-Lipschitz, meaning

$$\max_{x,x' \in \mathcal{X}} \frac{\|f(x) - f(x')\|}{d_{\mathcal{X}}(x, x')} \leq \ell.$$

It turns out that Lipschitzness is precisely the property needed in order to bound $\|q_f(\tilde{K}) - q_f(\tilde{K}')\|$ while effectively accommodating the underlying coupling between $\tilde{K}, \tilde{K}'$. This is a novel aspect of our analysis that has not been explored by previous sensitivity analysis.

Theorem 4.1. *Let $q_f(K)$ be a linear query of the form in (5), where $f : \mathcal{X} \to \mathbb{R}^d$ is ℓ-Lipschitz. Then, we have $\Delta_{\mathsf{EM}}(q_f) \leq \ell$.*

Remarks. The above result stands in contrast with traditional sensitivity arguments, which typically assume the weaker condition that f is merely bounded by ℓ. Although our Lipschitz assumption is stronger, it is satisfied by many queries of interest, such as

- *Average.* If $\mathcal{X} \subseteq \mathbb{R}^d$, then taking f to be the identity function will cause q_f to simply be the average of the elements in K.
- *Kernel Smoothed Region Queries.* If $R \subseteq \mathcal{X}$ is a region of interest, then $f(x) = \mathbf{1}[x \in R]$ will yield a q_f computing the fraction of elements that lie in R. Using a smooth interpolation of $\mathbf{1}[x \in R]$, such as kernel smoothers Hastie et al. (2009), will give a smooth approximation to this value.
- *Similarity Queries.* If $\mathcal{X} \subseteq \mathbb{R}^d$ and $c \in \mathbb{R}^d$ is a query vector, then letting $f(x) = \langle x, c \rangle$ will return the average similarity of each vector in K with c, with similarity measured by the dot product. This is an instance of a *linear embedding query* which we will study in detail in Section 6.1.

Additionally, the aforementioned example illustrates that this sensitivity analysis is tight. This means that $d_{\mathcal{X}}$, in addition to defining the privacy semantics, also influences the types of queries that can be answered with good utility.

Proof Sketch. At a high level, consider moving K onto K' one point at a time. The Lipschitz assumption allows us to bound the change in $q_f(K)$ in terms of how far the point is moving, which translates to the d_{EM} distance. More formally, let C be a minimum cost coupling between $\tilde{K}, \tilde{K}'$. We may write

$$\|q_f(\tilde{K}) - q_f(\tilde{K}')\| = \left\| \mathbb{E}_{x \sim \tilde{K}}[f(x)] - \mathbb{E}_{y \sim \tilde{K}'}[f(y)] \right\|.$$

We can view x and y as jointly generated by C, and the above expression becomes $\|\mathbb{E}_{x,y\sim C}[f(x) - f(y)]\|$. By the triangle inequality, we know this is at most

$$\|q_f(\tilde{K}) - q_f(\tilde{K}')\| \leq \mathbb{E}_{x,y\sim C}[\|f(x) - f(y)\|],$$

and by the Lipschitz assumption, we may upper bound this by

$$\ell \cdot \mathbb{E}_{x,y\sim C}[d_{\mathcal{X}}(x,y)] = \ell \times d_{\mathsf{EM}}(\tilde{K}, \tilde{K}'),$$

which completes the proof. □

Using the upper bound on $\Delta_{\mathsf{EM}}(q_f)$, we follow a well-known approach for privately releasing a point with known sensitivity under a norm: sample a point U uniformly from the ball $\{x \in \mathbb{R}^d : \|x\| = 1\}$, and release $q_f + \ell g U$, where $g \sim \Gamma(d, \frac{\omega}{\alpha})$ is the Gamma distribution with shape d and scale $\frac{\omega}{\alpha}$ (Hardt and Talwar, 2010). Here, ω is a scale parameter that may be different in the central or local model, since the sensitivity of f is less in the bounded central model. This mechanism, PrivEMDLinear, is outlined in Algorithm 1. Combining Theorem 4.1 with a standard privacy analysis, we can show that PrivEMDLinear satisfies $(\alpha, 0)$-d_{EM} DP.

Lemma 4.2. *PrivEMDLinear (Algorithm 1) with scale $\omega = \frac{1}{\alpha}$ satisfies $(\alpha, 0)$-unbounded local d_{EM}-DP and with scale $\omega = \frac{1}{\alpha n}$ satisfies $(\alpha, 0)$-bounded central d_{EM}-DP.*

Remarks. When using the 1-norm, PrivEMDLinear becomes the multidimensional Laplace mechanism. We may instantiate PrivEMDLinear with any noise mechanism that preserves (α, δ)-local $\|\cdot\|_p$-DP in the space $\mathbb{R}^d$. In Section 6.1, we will instantiate PrivEMDLinear using Gaussian noise of width $\frac{\omega\sqrt{1.25\ln(1/\delta)}}{\alpha}$ Dwork et al. (2014), which will give the proper error dependence on d under the 2-norm.

Concrete Example. In our location example, consider releasing the average distance of each point in K from a particular city in the local model. This can be expressed with $f(x) = d_{\mathcal{X}}(x, c)$, where c is the city; by the triangle inequality this is 1-Lipschitz. PrivEMDLinear could then be applied to release $q_f(K_i)$ plus noise of expected magnitude $\frac{\ell}{\alpha} = 0.04$ per user; the total noise will be $\frac{0.04}{\sqrt{n}} = 1.26 \times 10^{-4}$, corresponding to an error of just 1.6 miles.

Algorithm 1: PrivEMDLinear, an algorithm for releasing linear queries under bounded d_{EM}-DP.

Data: q_f – A d-dimensional linear query; ℓ – Upper bound of the Lipschitz constant of f; K – Input dataset; ω – scale parameter
Result: An estimate of $q_f(K)$
Sample U uniformly from $\{x \in \mathbb{R}^d : \|x\| = 1\}$;
Sample $g \in \mathbb{R}$ from $\Gamma(d, \omega)$;
return $\hat{q} = q_f(\tilde{K}) + \ell g U$;

4.2 Unordered Release of Item-wise Queries

We now consider the problem of directly releasing a private query applied to each item in K. This can provide a more fine-grained result than the aforementioned linear queries, which outputs the

average over all the items. We release the query results as an unordered list to take advantage of the fact that subsequent computation (such as, aggregation) often does not depend on the ordering of the data Feldman et al. (2022). Specifically, our second mechanism PrivEMDItemWise applies a mechanism $\mathcal{A}$, which satisfies $(\alpha_0, 0)$-$d_{\mathcal{X}}$-DP, *to each item* individually. We use $\mathcal{A}$ as a black-box making PrivEMDItemWise completely general-purpose. For example, one could let $\mathcal{A}$ be a private item-release mechanism (see Section 7 for some examples) and use PrivEMDItemWise to form a histogram of the dataset. $\mathcal{A}$ could also be a classifer, and PrivEMDItemWise can then release a simplified representation of the dataset. Once PrivEMDItemWise applies $\mathcal{A}$ to each element in the dataset, it shuffles the results (to remove any ordering of the data) and outputs the shuffled list. This appears in Algorithm 2, and a precursor appeared in Fernandes et al. (2019).

As PrivEMDItemWise does not hide the size of K, we show it satisfies bounded d_{EM}-DP. We use the following argument: for a neighboring dataset $K' = \{x'_1, \ldots, x'_m\}$, by Lemma 2.1 there exists a permutation $\pi : [m] \to [m]$ satisfying Eq. (3). Observe that we release the query responses in an unordered fashion by explicitly shuffling them. This allows us to pair up the element x_i with $x_{\pi(i)}$ and analyze the privacy guarantee of releasing $\mathcal{A}(x_1), \ldots, \mathcal{A}(x_m)$ versus $\mathcal{A}(x'_{\pi(i)}), \cdots, \mathcal{A}(x'_{\pi(m)})$. Prior work does this with composition Fernandes et al. (2019). However, composition is not the

Algorithm 2: PrivEMDItemWise, a general mechanism for releasing a item-wise queries from K as an unordered list under bounded d_{EM}-DP

Data: Dataset $K \in \mathcal{X}^m$, Mechanism $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ satisfying $(\alpha_0, 0)$-$d_{\mathcal{X}}$ DP
Result: $L \in \mathcal{Y}^m$, unordered list (multiset) of item-wise queries from K
$L = \emptyset$;
for $i = 1, \ldots, m$ **do**
  Add $\mathcal{A}(x_i)$ to L;
end
Shuffle(L);
return L

right tool for obtaining a tight privacy analysis. The reason is that composition assumes that each $\mathcal{A}(x_i)$ is output sequentially, and in particular it is possible to identify which point came from $\mathcal{A}(x_i)$ and which came from $\mathcal{A}(x_{\pi(i)})$. In our case, we output an unordered list, and it is not possible to link which point came from an index i. Based on this observation, our key idea is to leverage privacy amplification by shuffling (Feldman et al., 2022) instead, which can yield a much smaller privacy parameter when the output is order invariant.

In particular, our core technical contribution is to analyze a general formulation of the privacy amplification by shuffling problem, where the vector $x_1, \ldots, x_m$ is changed to an arbitrary vector $x'_1, \ldots, x'_m$. The key quantities we have control over are $\|v\|_1$ and $\|v\|_0$, where $v = (d_{\mathcal{X}}(x_i, x'_i))_{i=1}^m$ (the d_{EM} distance allows us to bound $\|v\|_1$ and the maximum contribution from a user allows us to bound $\|v\|_0$); thus, our privacy bound depends on them. We believe our general result is of independent interest in the field of metric DP. Formally,

Theorem 4.3. *Suppose that $(\mathcal{X}, d_{\mathcal{X}})$ is a metric space such that $d_{\mathcal{X}}(\cdot, \cdot) \leq 1$, and that $\mathcal{A}$ is an $(\alpha_0, 0)$ $d_{\mathcal{X}}$-DP algorithm. Let $(x_1, \ldots, x_m)$ and $(x'_1, \ldots, x'_m)$ be two vectors, and we define $v = (d_{\mathcal{X}}(x_i, x'_i))_{i=1}^m$. Let $0 < \delta < 1$ be a constant, and suppose it holds that $\alpha_0 < \ln(\frac{m}{16 \ln(4m/\delta)})$.*

Then, for all outputs O, we have that

$$\Pr[\mathsf{Shuffle}(\mathcal{A}(x_1),\ldots,\mathcal{A}(x_m)) \;=\; O] \;\le\; e^{\alpha}\Pr[\mathsf{Shuffle}(\mathcal{A}(x'_1),\ldots,\mathcal{A}(x'_m)) \;=\; O] \;+\; \delta e^{\alpha},$$

where

$$\alpha \le \|v\|_0 \ln\left(1+\frac{\exp(\alpha_0\|v\|_1/\|v\|_0)-1}{\exp(\alpha_0\|v\|_1/\|v\|_0)+1}\left(\frac{8\sqrt{e^{\alpha_0}\ln(4\|v\|_0/\delta)}}{\sqrt{m}}+\frac{8e^{\alpha_0}}{m}\right)\right).$$

Remarks. In particular, if $\alpha_0 \le \frac{\|v\|_0}{\|v\|_1}$, the above bound is $\approx \frac{\alpha_0\|v\|_1}{\sqrt{m}}$, which grows with just $\sqrt{m}$ (as $\|v\|_1 \le m$). The standard shuffling result only assumes that $\mathcal{A}$ satisfies α-*local* DP, and that just x_1 is changed to x'_1 (since each user owns a single item). Theorem 4.3 can be specialized to recover the state-of-the-art result for this special case Feldman et al. (2022), but it is significantly more general in its current form.

Proof Sketch. Using group privacy, we can analyze the privacy guarantee between $(x_1,\ldots,x_m)$ and $(x'_1,\ldots,x'_m)$, where up to $\|v\|_0$ points change, instead of just changing one point as a time. We then analyze a change of one point by generalizing and simplifying the state-of-the-art technique in Feldman et al. (2022). We show that the resulting privacy parameter for changing the point x_i to x'_i is $g(w_i)$ where

$$g(w_i) = \ln\left(1+\frac{e^{\alpha_0 w_i}-1}{e^{\alpha_0 w_i}+1}\left(\frac{8\sqrt{e^{\alpha_0}\ln(4/\delta)}}{\sqrt{m}}+\frac{8e^{\alpha_0}}{m}\right)\right).$$

By group privacy $\|v\|_0$ times, the overall privacy parameter is $\sum_{i=1}^{\|v\|_0} g(w_i)$. The final step is proving that g is concave so the worst-case amplification is simply $\|v\|_0 g(\frac{\|v\|_1}{\|v\|_0})$. □

Comparison with Composition. Analyzing Theorem 4.3 using the state-of-the-art composition results (Kairouz et al., 2015) and $\alpha_0 \le 1$ gives us

$$\alpha \le O\left(\alpha_0\|v\|_2\sqrt{\ln\tfrac{1}{\delta}}\right).$$

However, we cannot form a satisfying bound on the 2-norm of v—it is only possible to say $\|v\|_2 \le \|v\|_1$ which is tight when e.g. $\|v\|_0 = 1$. The bound is thus missing the factor of $\frac{1}{\sqrt{m}}$—composition here does not leverage the fact that all m items are released in a random order.

Combining (3) and Theorem 4.3, we obtain an improved privacy guarantee for PrivEMDItemWise. The guarantee can be stated in both the bounded local and central models. In the local model, recall that each user is applying PrivEMDItemWise to their data. In the central model, the aggregator applies PrivEMDItemWise to the entire dataset, and releases the frequencies of mn itemwise queries.

Theorem 4.4. *For any $\delta \in (0,1)$, PrivEMDItemWise shown in Algorithm 2 satisfies bounded local (α,δ')-d_{EM} DP, where*

$$\alpha = \sup_{w\in[0,1]} \frac{h(m;m,mw)}{w} \qquad \text{and } \delta' = \delta e^{h(m;m,m)},$$

and

$$h(m;x_0,x_1) = x_0 \ln\left(1+\frac{\exp(\alpha_0 x_1/x_0)-1}{\exp(\alpha_0 x_1/x_0)+1}\left(\frac{8\sqrt{e^{\alpha_0}\ln(4x_0/\delta)}}{\sqrt{m}}+\frac{8e^{\alpha_0}}{m}\right)\right).$$

Similarly, PrivEMDItemWise *satisfies bounded central* (α, δ')*-*d_{EM} *DP, where*

$$\alpha = \sup_{w \in [0,1]} \frac{h(mn; m, mw)}{w} \qquad \text{and } \delta' = \delta e^{h(mn;m,m)}.$$

Remarks. Theorem 4.4 gives the tightest possible privacy parameters, but we may also give an asymptotic formula as follows. For desired privacy parameters (α, δ), one should set

$$\alpha_0 = \begin{cases} \frac{\alpha}{32\sqrt{m \ln(4me^\alpha/\delta)}} & \text{if } \alpha \leq 32\sqrt{m \ln(4me^\alpha/\delta)} \\ 2 \ln\left(\frac{\alpha}{16\sqrt{m \ln(4me^\alpha/\delta)}}\right) & 32\sqrt{m \ln(4me^\alpha/\delta)} < \alpha < m \end{cases} \tag{6}$$

and

$$\alpha_0 = \begin{cases} \frac{\alpha\sqrt{n}}{32\sqrt{m \ln(4me^\alpha/\delta)}} & \text{if } \alpha\sqrt{n} \leq 32\sqrt{m \ln(4me^\alpha/\delta)} \\ 2 \ln\left(\frac{\alpha\sqrt{n}}{16\sqrt{m \ln(4me^\alpha/\delta)}}\right) & 32\sqrt{m \ln(4me^\alpha/\delta)} < \alpha\sqrt{n} < m\sqrt{n} \end{cases} \tag{7}$$

in order to achieve d_{EM}-DP in the bounded local and central model, respectively. Assuming $\alpha \leq O(\ln(m))$, this means that the privacy parameter will be roughly $\frac{\alpha}{\sqrt{m}}$ (resp. $\ln(\frac{\alpha\sqrt{n}}{\sqrt{m}})$) for releasing the m samples; this is asymptotically better than the analysis with composition which would require setting $\alpha_0 = \frac{\alpha}{m}$ (resp. $\frac{\alpha}{m}$)). Even with higher $\alpha = m^c$ for $c < 1$, the budget is still $\frac{\alpha}{\sqrt{m^{1+c}}}$ (resp. $\ln(\frac{\alpha\sqrt{n}}{\sqrt{m^{1+c}}})$), which are both significant asymptotic improvements.

Concrete Example. Our improved analysis makes the most significant improvements in the central model. Here, we would have to apply PrivEMDItemWise with $\alpha_0 = \frac{\alpha}{m} = 0.025$ for each of the $m = 10^3$ location data points per user. Using the guarantee of Theorem 4.4, it is possible to set $\alpha_0 \approx 3.0$ – a several orders of magnitude improvement.

Open Questions. The primary open question in the design of our mechanism is whether one can obtain a tighter privacy analysis of Theorem 4.4 that does not rely on group privacy, but rather analyzes the amplification potential of all points at once.

New Proof Techniques. First, in the case of linear queries, we bound the sensitivity under d_{EM}-DP using a Lipschitz assumption. This stands in contrast to sensitivity arguments in standard DP, which typically assume the weaker condition that the query is merely bounded. This is a novel aspect of our analysis that has not been explored in prior sensitivity analysis. Second, in the context of unordered release of item-wise queries, we prove a new result for privacy amplification by shuffling. This is a more generalized version of the standard shuffling result Feldman et al. (2022) in two ways – (1) the standard result assumes that the private mechanism satisfies *local*-DP whereas we work with mechanisms that satisfy $d_{\mathcal{X}}$-DP; (2) the standard result only considers changes to a single point, whereas we allow changes to all m input points. Using this result, we provide a tighter privacy analysis than what composition theorems would allow. Our new privacy amplification by shuffling result can be of independent interest in the field of metric DP.

5 Generalization to Unbounded DP

The mechanisms presented so far face two challenges when applied to the unbounded data setting. First, a direct privacy analysis of the unbounded data setting is difficult since we cannot leverage

Lemma 2.1, which significantly simplifies the analysis (for the bounded data setting). Second, and more importantly, the unbounded central model offers no utility improvement over the local model. In the worst-case scenario, a single user may contribute nearly all the data in the dataset, effectively reducing any algorithm to satisfying only local d_{EM}-DP. This issue has been noted in previous work in user-level DP Liu et al. (2023).

In this section, we tackle these challenges by showing a *blackbox* reduction from unbounded d_{EM}-DP to bounded d_{EM}-DP. Our reduction works in both the local and central models. The key idea of the reduction is to *smoothly* project a dataset K of any size to a dataset L of a given fixed size, such that the d_{EM} distance between any two input datasets and the d_{EM} distance between their projections are roughly the same. Then, it is easy to show that applying any bounded d_{EM}-DP algorithm to the smooth projections is sufficient to guarantee unbounded d_{EM}-DP for the entire scheme. Full proofs for this section appear in Appendix C.

Our proposed projection mechanism is smooth in a near-multiplicative sense, albeit with a small additive penalty when the d_{EM} between the two datasets is small. We account for this subtlety by slightly modifying the privacy semantics of d_{EM}-DP in the unbounded setting to not grow arbitrarily strong as $d_{\mathsf{EM}}(\tilde{K}, \tilde{K}') \to 0$. Instead, we introduce a distance threshold r such that all $d_{\mathsf{EM}}(\tilde{K}, \tilde{K}') \leq r$ enjoys a *uniform* privacy guarantee of ε. This refined privacy definition, termed discrete d_{EM}-DP, is formalized (in the local model) as:

Definition 5.1. *[Discrete Local d_{EM}-DP] Let $\mathcal{M}$ be a mechanism which acts on a dataset K. We say $\mathcal{M}$ satisfies (ε, δ, r)-discrete local d_{EM}-DP if, for any two datasets $K, K' \in \mathcal{X}^*$ such that $d_{\mathsf{EM}}(\tilde{K}, \tilde{K}') \leq r$,*

$$\Pr[\mathcal{M}(K) = O] \leq e^{\varepsilon} \Pr[\mathcal{M}(K') = O] + \delta.$$

Like in standard DP, the above definition uses the parameter ε because it is a *unitless* privacy parameter—the unit of the metric is expressed in the parameter r.

Fact 5.1. *For any K, K' such that $d = d_{\mathsf{EM}}(\tilde{K}, \tilde{K}')$, $\mathcal{M}$ satisfies*

$$\Pr[\mathcal{M}(K) = O] \leq e^{\varepsilon \lceil \frac{d}{r} \rceil} \Pr[\mathcal{M}(K') = O] + \delta \exp(\lceil \tfrac{d}{r} \rceil).$$

Fact 5.1 is implied from Definition 5.1 followed by a direct application of group privacy Dwork (2006). This guarantee can be interpreted as providing d_{EM}-DP at the granularity of units of d_{EM} distance r. Note that for all $d \geq r$, we have $\varepsilon \lceil \frac{d}{r} \rceil \leq \frac{2\varepsilon}{r} d$. Thus, (ε, δ, r)-discrete local d_{EM} DP is roughly equivalent to $(\frac{2\varepsilon}{r}, \delta)$-unbounded local d_{EM}-DP, except if $d_{\mathsf{EM}}(\tilde{K}, \tilde{K}') \leq r$. In this case, the privacy parameter will not go below ε. This adjustment does not significantly alter the overall privacy semantics of d_{EM}-DP; one may simply set α as described in Section 4.

In the central model, we make a similar definition:

Definition 5.2. *[Discrete Central d_{EM}-DP] Let $K_G = K_1 \cup \cdots \cup K_n$ denote a global dataset from n users (of any size). We say $K_G \sim_r K'_G$ if K'_G can be obtained from K_G by changing K_i to K'_i for just one user i, such that $d_{\mathsf{EM}}(\tilde{K}_i, \tilde{K}'_i) \leq r$. We say a mechanism $\mathcal{M}(K_G)$ satisfies (ε, δ, r)-discrete central d_{EM}-DP if, for all K_G, K'_G such that $K_G \sim_r K'_G$, we have*

$$\Pr[\mathcal{M}(K_G) = O] \leq e^{\varepsilon} \Pr[\mathcal{M}(K'_G) = O] + \delta.$$

As before, (ε, δ, r)-discrete central d_{EM}-DP is roughly equivalent to $(\frac{2\varepsilon}{r}, \delta)$-bounded central d_{EM}-DP when all user datasets have size m. We will see that Definition 5.2 is the appropriate generalization to unbounded user datasets under our projection mechanism which is described below.

Because our projection mechanism must preserve the d_{EM} between $\tilde{K}, \tilde{K}'$, it is not acceptable to select an arbitrary set of points from $\tilde{K}, \tilde{K}'$, as this could dramatically inflate the d_{EM} distance. However, couplings are intimately tied with *sampling* – observe that two random samples from $\tilde{K}, \tilde{K}'$ can actually be coupled so that their expected distance is $d_{\mathsf{EM}}(\tilde{K}, \tilde{K}')$. If this process is repeated multiple times, then the cost of coupling the samples will converge to $d_{\mathsf{EM}}(\tilde{K}, \tilde{K}')$. Thus, sampling with replacement will result in a projection that does not increase the d_{EM} distance by too much.

Algorithm 3: BoundedEMDReduction, a reduction from unbounded d_{EM}-DP to bounded d_{EM}-DP.

Data: K_G - Global datasets of n users; $\mathcal{A}$ - A mechanism satisfying bounded d_{EM}-DP; s - Number of samples.

$L = \emptyset$;
for $i = 1$ *to* n **do**
  Add s uniform samples with replacement from K_i to L
end
$O = \mathcal{A}(L)$;
return O

Lemma 5.2. *Let $\tilde{K}, \tilde{K}' \in \Delta^{\mathcal{X}}$ be probability distributions, and let C^* be the minimum cost coupling between $\tilde{K}, \tilde{K}'$. Let $\{(x_i, y_i)\}_{i=1}^s$ be s i.i.d. samples from C^*, $L = (x_1, \ldots, x_s)$ and $L' = (y_1, \ldots, y_s)$. Then,*

$$\Pr[d_{EM}(\tilde{L}, \tilde{L}') \geq (1 + \sqrt{2}) d_{EM}(\tilde{K}, \tilde{K}') + \tfrac{3}{s} \ln(\tfrac{1}{\delta})] \leq \delta.$$

Remarks. The multiplicative factor of $1 + \sqrt{2}$ shows that the projection is smooth when $d_{\mathsf{EM}}(\tilde{K}, \tilde{K}')$ dominates the additive factor of $\frac{3}{s} \ln(\frac{1}{\delta})$. This is achieved when the number of samples s is much larger than the inverse of $d_{\mathsf{EM}}(\tilde{K}, \tilde{K}')$. If s is not sufficiently large, then not enough samples are being taken to ensure convergence.

Proof Sketch. Let C be a minimum-cost coupling between $\tilde{K}, \tilde{K}'$. To show that $\tilde{L}, \tilde{L}'$ are not far from each other, we can view $\tilde{L}'$ as being generated from $\tilde{L}$, where for each point $x \in L$, a point $y \sim C_x(\cdot)$ is added to L'. This view of $\tilde{L}, \tilde{L}'$ shows there is a transportation plan from $L = \{x_1, \ldots, x_s\}$ and $L' = \{y_1, \ldots, y_s\}$ of expected cost $\mathbb{E}_{x \sim C_1, y \sim C_x} d_{\mathcal{X}}(x, y) = d_{\mathsf{EM}}(\tilde{K}_i, \tilde{K}_i')$. Using Bernstein's inequality, we can show with probability at least $1 - \delta$, $d_{\mathsf{EM}}(\tilde{L}, \tilde{L}')$ is upper bounded by $2 d_{\mathsf{EM}}(\tilde{K}_i, \tilde{K}_i') + \frac{6}{s} \log(\frac{1}{\delta})$. □

Our full reduction to bounded d_{EM}-DP first projects K onto a dataset of size m by taking samples with replacement. Next, it applies a *blackbox* bounded d_{EM}-DP mechanism, $\mathcal{A}$, to the projected dataset L. By blackbox application we mean that $\mathcal{A}$ can be any arbitrary mechanism as long as it satisfies bounded d_{EM}-DP. We call this mechanism BoundedEMDReduction, and it is illustrated in the central model in Algorithm 3 (in the local model, each user samples from their own K_i, so we would simply have $n = 1$). Using the projection guarantee of Lemma 5.2, BoundedEMDReduction enjoys the following privacy guarantee:

Theorem 5.3. *Let $\varepsilon > 0$ and $\delta, r \in [0,1]$ be arbitrary constants. Suppose $\mathcal{A}$ is a mechanism which satisfies (α, δ)-bounded local d_{EM}-DP (Definition 3.1), where*

$$\alpha = \frac{\varepsilon}{(1+\sqrt{2})r + \frac{3}{s}\ln(\frac{1}{\delta})}.$$

Then, BoundedEMDReduction satisfies $(\varepsilon, 2\delta, r)$-discrete local d_{EM}-DP. Similarly, if $\mathcal{A}$ is (α, δ)-bounded central d_{EM}-DP (Definition 3.1), then BoundedEMDReduction is $(\varepsilon, 2\delta, r)$-discrete central d_{EM}-DP.

Remarks. If the number of samples s is at least $\frac{\ln(1/\delta)}{r}$, then Theorem 5.3 shows there is only a small multiplicative cost to considering just bounded d_{EM}-DP (in the respective local or central model). In this case, $\mathcal{A}$ will need to roughly satisfy $(\frac{\varepsilon}{r}, \delta)$-bounded d_{EM}-DP, and this is roughly the same as the resulting (ε, δ, r)-discrete d_{EM} DP algorithm. There is no privacy disadvantage to taking a large number of samples, and the utility may also increase due to more information about the dataset being captured (recall that the projection does not providing privacy; it is being provided by $\mathcal{A}$). Thus, the number of samples may be set to be large with computational costs being the only constraint.

Proof Sketch. The proof of privacy is almost immediate from Lemma 5.2. The variables $\tilde{L}, \tilde{L}'$ in two executions of BoundedEMDReduction are random variables, but with probability δ, the d_{EM} distance between them is $c = (1+\sqrt{2})r + \frac{3}{s}\ln(\frac{1}{\delta})$. By the convexity of differential privacy, we can analyze the privacy parameter of every fixed choice of $\tilde{L}, \tilde{L}'$. With probability $1-\delta$, the privacy parameter will be $\frac{\alpha}{c}$, by the privacy guarantee of $\mathcal{A}$. □

BoundedEMDReduction can be used to bound the contribution of each user in the central setting, allowing us to apply the simpler Definition 3.2. In addition, it can be used to adapt PrivEMDItemWise to the unbounded data setting. One caveat is that utility may not be preserved if the number of user samples is too small or, in the central setting, if the users data distributions are heterogeneous. In particular, if users have varying numbers of samples, each from different distributions, applying BoundedEMDReduction equalizes the frequency of all user data. Nonetheless, it is often reasonable to assume the users have homogeneous data distributions Liu et al. (2020); Acharya et al. (2023).

Open Questions. Many sampling procedures are likely to be compatible with d_{EM}. This leads to the question of whether different procedures, such as sampling without replacement, are also smooth projections.

New Proof Techniques. Our blackbox reduction from unbounded to bounded d_{EM}-DP involves projecting a dataset onto one with a fixed size. The projection needs to be smooth, i.e., the d_{EM} distance between any two input datasets and the d_{EM} distance between their projections are roughly the same. We prove that sampling with replacement is indeed a smooth projection. The novelty of the analysis comes from viewing sampling from two datasets in terms of a coupling between them.

6 Applications of Proposed Mechanisms

In this section, we compare the utilities of PrivEMDLinear and PrivEMDItemWise to existing mechanisms satisfying user-level DP. For simplicity, we assume the bounded data setting. Full proofs for this section appear in Appendix D.

Notations. We define the following quantities of a real matrix $M \in \mathbb{R}^{d\times k}$. First, the (p,q) operator norm of M, denoted by $\|M\|_{p\to q}$, is given by $\|M\|_{p\to q} = \sup_{x\in\mathbb{R}^k, \|x\|_p\le 1} \|Mx\|_q$. We can show that $\|M\|_{1\to 2}$ is equal to the maximum ℓ_2 norm of a column of M. Furthermore, $\|M\|_{2\to 2}$, more commonly written as $\|M\|_2$, is the spectral norm and is equal to the maximum singular value of M. Matrix norms satisfy the important submultiplicative property, which states that $\|MN\|_{p\to r} \le \|M\|_{q\to r}\|N\|_{p\to q}$ for any matrices M, N and $p, q, r \ge 1$. Next, let I_d denote the $d \times d$ identity matrix, and again suppose that $M \in R^{d\times k}$ with $d \le k$. If M has full row rank, then there exists a matrix $N \in \mathbb{R}^{k\times d}$ such that $MN = I_d$. We call such a matrix N a *right inverse* of M. Finally, for $M \in \mathbb{R}^{s_1\times t_1}$ and $N \in \mathbb{R}^{s_2\times t_2}$, let $M \otimes N \in \mathbb{R}^{s_1 s_2 \times t_1 t_2}$ denote the Kronecker product of two real matrices, whose entry in $((i_1,i_2),(j_1,j_2))$ is $M_{i_1 j_1} N_{i_2 j_2}$.

6.1 Linear Embedding Queries

Many applications of metric DP assume there is an embedding function $\phi : \mathcal{X} \to \mathbb{R}^t$, which maps an item to its semantic representation in $\mathbb{R}^t$ (each of the examples in Section 1 have an embedding representation). The metric $d_{\mathcal{X}}$ is then the distance between $\phi(x)$ and $\phi(x')$; in this section, we consider the l_2 distance.

Since $\phi(x)$ also communicates information about the item x, we define *linear embedding queries* as linear queries applied to an item's embedding $\phi(x)$. Formally,

$$q_{f\circ\phi}(K) = \mathbb{E}_{x\sim\tilde{K}}[f \circ \phi(x)],$$

where $f(y) = Fy$ for a matrix $F \in \mathbb{R}^{d\times t}$ (meaning that f is a *linear* function). Assume each row F_i of F is normalized so that $\|F_i\|_2 \le 1$. Each coordinate of $f \circ \phi$ is equal to $\mathbb{E}_{x\sim\tilde{K}}[\langle F_i, \phi(x)\rangle]$. Thus, we may view each coordinate of a linear embedding query as a similarity query in the embedding space with query point F_i. Our analysis will assume that $d < |\mathcal{X}|$ and $d \ll n$, which is usually the case in practice. Note that we may write $q_{f\circ\phi}$ as $F\Phi\tilde{K}$, where $\Phi \in \mathbb{R}^{t\times\mathcal{X}}$ is the collection of embedding vectors in $\mathcal{X}$.

6.1.1 Local Model

Existing user-level DP mechanisms ask user at index i to privately release the query $\hat{q}_i = q_{f\circ\phi}(\tilde{K}_i)$. The aggregator computes the average $\hat{q} = \frac{1}{n}\sum_{i=1}^n \hat{q}_i$. The current best solutions have the following error guarantee (Duchi et al., 2013; Bassily, 2019)[4]:

Lemma 6.1. *(From Proposition 3 in Duchi et al. (2013)) There exists an $(\varepsilon, 0)$-bounded user-level DP in the local model algorithm which produces an estimate $\hat{q}$ such that, for all $\tilde{K}$,*

$$\mathbb{E}[\|\hat{q} - q_{f\circ\phi}(\tilde{K})\|_2] \le O\left(\|F\Phi\|_{1\to 2}\frac{\sqrt{d}}{\varepsilon\sqrt{n}}\right).$$

To interpret the term $\|F\Phi\|_{1,2}$, we can use the inequality $\|F\Phi\|_{1\to 2} \le \|F\|_2\|\Phi\|_{1\to 2}$, which is tight for certain choices of F and Φ. By assumption, we know $\|F\|_2 \le \sqrt{d}$ and $\|\Phi\|_{1,2} \le 1$, both of which can also be tight. The bound is thus $O(\frac{d}{\varepsilon\sqrt{n}})$.

On the other hand, for d_{EM}-DP, by Theorem 4.1, we know that $\Delta_{\mathsf{EM}}(q_{f\circ\phi})$ is at most the Lipschitz constant of $f \circ \phi$ given by:

$$\max_{x,x'\in\mathcal{X}} \frac{\|F(\phi(x)) - F(\phi(x'))\|}{\|\phi(x) - \phi(x')\|} \le \max_{x,x'\in\mathcal{X}} \frac{\|F(\phi(x)) - \phi(x'))\|}{\|\phi(x) - \phi(x')\|} \le \|F\|_2.$$

[4]No algorithms are known which satisfy (ε, δ)-DP and have better error than the $(\varepsilon, 0)$-DP algorithm shown.

Hence, each user can apply PrivEMDLinear with the Gaussian mechanism with $\ell = \|F\|_2$, which gives the following utility guarantee:

Lemma 6.2. *There exists an (α, δ)-bounded d_{EM}-DP algorithm in the local model which produces an estimate $\hat{q}$ such that, for all $\tilde{K}$,*

$$\mathbb{E}[\|\hat{q} - q_{f\circ\phi}(\tilde{K})\|_2] \leq \|F\|_2 \frac{\sqrt{1.25d\ln(1/\delta))}}{\alpha\sqrt{n}}.$$

Remarks. We use the Gaussian mechanism because it performs better under the ℓ_2 error than the pure $(\alpha, 0)$-bounded local d_{EM} DP illustrated in Algorithm 1. However, this forces us to use $\delta > 0$. Compared to Lemma 6.1, the above bound differs by a factor of $\frac{\varepsilon}{\alpha}$ (and small $\ln\frac{1}{\delta}$ terms)—when $\alpha = \varepsilon$, we know that d_{EM}-DP provides better privacy. When $\varepsilon \ll \alpha$, that PrivEMDItemWise offers lower error than Lemma 6.1.

6.1.2 Central Model

In the central model, linear query release has been extensively studied, and optimal algorithms under item-level DP are known (Hardt and Talwar, 2010; Bhaskara et al., 2012; Nikolov et al., 2013). These algorithms can be easily adapted to user-level DP, which will provide the following guarantee[5]:

Lemma 6.3. *(From Theorem 1.3 in Hardt and Talwar (2010)) There exists an $(\varepsilon, 0)$-bounded user-level DP algorithm in the central model which produces an estimate $\hat{q}$ such that, for all $\tilde{K}$,*

$$\mathbb{E}[\|\hat{q} - q_{f\circ\phi}(\tilde{K})\|_2] \leq O\left(\|F\Phi\|_{1\to 2}\frac{\sqrt{d}}{\varepsilon n}\ln\left(\frac{k}{d}\right)\right).$$

To provide (α, δ)-bounded d_{EM}-DP in the central model, we can use PrivEMDLinear with the Gaussian mechanism with scale $\omega = \frac{1}{n\alpha}$. Following the same approach as in Lemma 6.2, this results in $O(\|F\|_2\frac{\sqrt{d\ln\frac{1}{\delta}}}{\alpha n})$ error. Again, this is worse than Lemma 6.3 by a factor of $\frac{\varepsilon}{\alpha}$, and similar observations apply.

6.2 Frequency Estimation

Here, we evaluate the error of PrivEMDItemWise for private frequency estimation, where the goal is to obtain a private estimate $\tilde{H}$ of the (normalized) histogram $\tilde{K}$. This problem has been extensively studied in privacy Hay et al. (2009); Xu et al. (2013); Suresh (2019); Kairouz et al. (2016); Acharya et al. (2018); Chen et al. (2020); Acharya et al. (2023); the high-level goal is to minimize the ℓ_p distance between $\tilde{H}$ and $\tilde{K}$. However when the data domain is a general metric space $\mathcal{X}$, not all ℓ_p perturbations to $\tilde{K}$ are the same. Therefore, we measure the similarity between $\tilde{K}, \tilde{H}$ via $d_{\mathsf{EM}}(\tilde{K}, \tilde{H})$, as we do in our privacy definition.

To simplify the analysis while still demonstrating the effectiveness of our mechanisms, we fix $\mathcal{X}$ to be the following "clustered" metric space. Let $\mathcal{X} = \mathcal{B}\times\mathcal{C}$, where $\mathcal{B} = \{b_1, \ldots, b_s\}$, $\mathcal{C} = \{c_1, \ldots, c_t\}$

[5]For simplicity, we do not state or compare to the exact instance-optimal upper bounds known for linear queries. Instead, the upper bound in Lemma 6.3 is the optimal one for random linear queries Hardt and Talwar (2010). Like in the local case, there is no separation between (ε, δ)-DP and $(\varepsilon, 0)$-DP for this problem.

and $s \cdot t = k$. For some $r < \frac{1}{2}$, the distance is given by the following:

$$d_{\mathcal{B}\times\mathcal{C}}((b,c),(b',c')) = \begin{cases} 0 & \text{if } b = b' \text{ and } c = c' \\ r & \text{if } b = b' \\ 1 & \text{otherwise.} \end{cases}$$

We can think of this metric space as a collection of s clusters consisting of the t items $\{(b,c_1),\ldots,(b,c_t)\}$ for each $b \in \mathcal{B}$. Points in a cluster are more related, being at distance r apart, than items in two different clusters, which are distance 1 apart. We will assume that privacy is only needed between two items in the same cluster, so we will set $\alpha = \frac{\varepsilon}{r}$.

6.2.1 Algorithms in the Local Model

At a high level, in the local model each user applies a private mechanism $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ (with $\mathcal{Y}$ discrete and $|\mathcal{Y}| \geq |\mathcal{X}|$) to each sample and releases it. The central server forms an aggregate vector $v \in \mathbb{R}^{\mathcal{Y}}$. Let $A \in \mathbb{R}^{\mathcal{X}\times\mathcal{Y}}$ denote the transition probability matrix of $\mathcal{A}$; we have by linearity of expectation that $\mathbb{E}[v] = \tilde{K}A$. Assuming that A has a right inverse B, the central server returns the estimate $\tilde{H} = vB$, which is unbiased. All previous work in distribution estimation under local DP can be expressed in this way (Kairouz et al., 2016; Acharya et al., 2018; Chen et al., 2020; Acharya et al., 2023). We summarize this in Algorithm 4.

Algorithm 4: FreqEstLocal, a general framework for histogram estimation under local DP

Data: K, a family of datasets from n users each with size m; $\mathcal{A}$, a mechanism from $\mathcal{X}$ to $\mathcal{Y}$; $B \in \mathbb{R}^{\mathcal{Y}\times\mathcal{X}}$, a right inverse of $\mathcal{A}$.

for *each* *user* i *from* 1 *to* n **do**
- $L_i = \emptyset$;
- **for** $l_j \in K_i$ **do**
  - $r_j = \mathcal{A}(l_j)$;
  - Add r_j to L_i;
- **end**
- Release $\tilde{L}_i$;

end

$v = \frac{1}{n}\sum_{i=1}^{n} \tilde{L}_i$;

$\tilde{H} = vB$;

return $\tilde{H}$

The state-of-the-art approach for frequency estimation is the Hadamard response Acharya et al. (2018); Chen et al. (2020), which is based off of the Hadamard matrices (which form a robust encoding of $\mathcal{X}$). Specifically, the matrix A is given by $q_1\mathbf{1} + q_2 H$, where H is a Hadamard matrix and q_1, q_2 are constants chosen so that A is normalized and that each element is proportional to either e^{ε} or 1. This mechanism has the following utility:

Lemma 6.4. *(From Theorem 3.1 in Chen et al. (2020)) There exists a mechanism $\mathcal{A}$ such that* FreqEstLocal *satisfies (ε,δ)-bounded user-level DP and returns an estimator $\tilde{H}$ such that*

$$\max_K \mathbb{E}[d_{EM}(\tilde{K},\tilde{H})] \leq O\left(\sqrt{\frac{k}{mn}} + \sqrt{\frac{k^2\ln(m/\delta))}{n\varepsilon^2}}\right).$$

Remarks. In order to adapt the Hadamard response to the user-level setting, we suppose each user applies $\mathcal{A}$ to each sample with privacy budget $\frac{\varepsilon}{\sqrt{m\ln(m/\delta)}}$, and (ε,δ)-user level DP follows from composition Kairouz et al. (2015). The term $\sqrt{\frac{k}{mn}}$ is the sampling error which does not depend on ε, and the second $\sqrt{\frac{k^2\ln(m/\delta)}{n\varepsilon^2}}$ term is the cost of privacy. The cost of privacy usually dominates, and furthermore its dependence on m is not significant. This is because m reduces both the effect of each sample on the final estimate, and the privacy budget per sample, countervailing itself.

With d_{EM}-DP under our chosen metric, we can use a transition probability matrix A that is *less noisy*. This comes from the fact that only items in the same cluster need a strong privacy guarantee, whereas traditional user-level DP would require all points to have this level of privacy guarantee. To provide formal guarantees, we first derive an error bound on FreqEstLocal in terms of A (specifically its right inverse), which we will then optimize later.

Theorem 6.5. *For the metric space $\mathcal{X}=\mathcal{B}\times\mathcal{C}$ and any mechanism $\mathcal{A}$ satisfying $(\alpha_0,0)$ $d_{\mathcal{X}}$-DP where $\alpha_0=O(\frac{\alpha}{\sqrt{m\ln(me^{\alpha}/\delta)}})$ (α_0 is specifically defined in Theorem 4.4), FreqEstLocal is (α,δ)-bounded d_{EM}-DP in the local model and returns an estimator $\tilde{H}$ such that*

$$\max_{K}\mathbb{E}[d_{EM}(\tilde{H},\tilde{K})]\le r\sqrt{\frac{st(\|B^T\|_{1\to2}^2-1)}{mn}}+\sqrt{\frac{s(\|P^TB^T\|_{1\to2}^2-1)}{mn}},\tag{8}$$

where B is a right inverse of $\mathcal{A}$, $P=I_{\mathcal{B}}\otimes 1_{\mathcal{C}}^{+}$, and $1_{\mathcal{C}}^{+}$ is a column vector of 1s indexed by $\mathcal{C}$.

Remarks. The first term in the RHS of Eq. (8) is the cost of equalizing the mass between clusters, and the second term is the cost of equalizing the mass across clusters (since the matrix P essentially projects $\mathcal{A}$ to act between clusters). For small r, the first term approaches 0, and the latter term may also approach 0 because $\mathcal{A}$ will not often map a point outside its cluster under $d_{\mathcal{X}}$-DP (and thus, $\|P^TB^T\|_{1\to2}^2-1\to0$).

Proof Sketch. Observe that we may upper bound $d_{\mathsf{EM}}(\tilde{H},\tilde{K})$ with any transportation plan between $\tilde{H},\tilde{K}$. We will use the following one: first map the probability masses in each cluster so that they match, putting extra mass in an arbitrary point. This incurs at most $r\|\tilde{H}-\tilde{K}\|_1$ cost, since the intra-cluster distance is at most r. Next, equalizing the mass between clusters, which incurs at most $\|P(\tilde{H}-\tilde{K})\|_1$ cost, where P is the given matrix which can be viewed as the linear operator that adds the mass within each cluster together. Both of the error terms can then be bounded by viewing $\tilde{H}-\tilde{K}$ as the sum of mn independent variables drawn from a Dirichlet distribution with mean 0, and applying a variance analysis. □

Now, the task is to pick a mechanism $\mathcal{A}$ satisfying d_{EM}-DP, which minimizes the error in (8). The constraint of d_{EM}-DP is quite different from standard DP, and permits novel mechanism design. We use a natural generalization of k-randomized response Kairouz et al. (2016), adapted to $d_{\mathcal{X}}$-DP. Specifically, GKRR_{α_0} has probabilities given by, for each $(b,c)\in\mathcal{X}$,

$$\begin{aligned}\Pr[\mathsf{GKRR}(b,c)=(b,c))]&\propto e^{\alpha_0},\\ \Pr[\mathsf{GKRR}(b,c))=(b,c')]&\propto e^{(1-r)\alpha_0}\qquad\forall c'\neq c,\\ \Pr[\mathsf{GKRR}(b,c))=(b',c')]&\propto 1\qquad\forall b'\neq b,c'.\end{aligned}$$

Using this mechanism, the higher-order terms of Eq. (8) will approach 0 with r, as follows:

Theorem 6.6. *For the metric space $\mathcal{X} = \mathcal{B} \times \mathcal{C}$, FreqEstLocal with the mechanism $\mathcal{A} = \mathsf{GKRR}_{\alpha_0}$ satisfies (α, δ)-d_{EM} DP in the local model and returns an estimator $\tilde{H}$ such that*

$$\max_K \mathbb{E}[d_{\mathsf{EM}}(\tilde{H}, \tilde{K})] \quad \leq \quad r\sqrt{\frac{st^3}{mn}}\left(\frac{e^{\alpha_0}+s}{e^{\alpha_0}-e^{(1-r)\alpha_0}}\right) \; + \; \sqrt{\frac{s^2t^2}{mn}}\left(\frac{\sqrt{s+2(e^{\alpha_0}-1)}}{e^{\alpha_0}+(t-1)e^{(1-r)\alpha_0}-t}\right), \quad (9)$$

where α_0 is defined in Eq. (6)*.*

Remarks. Specifically, for our choice of $\alpha = \frac{\varepsilon}{r}$, we have

$$\max_K \mathbb{E}[d_{\mathsf{EM}}(\tilde{K}, \tilde{H})] \leq 4\sqrt{\tfrac{k^3}{mn}} + 64\tfrac{\sqrt{k^3}}{\alpha\sqrt{n}}\sqrt{\ln(4m\exp(\alpha)/\delta)}.$$

Similar to Lemma 6.4, the $\sqrt{\frac{k^3}{mn}}$ term is the cost of sampling. The $r\frac{\sqrt{k^3}}{\varepsilon\sqrt{n}}$ term is the cost of privacy, and it dominates when $\alpha \leq \sqrt{m}$. We will compare Theorem 6.6 with Lemma 6.4 when $k, \varepsilon, \alpha < \sqrt{m}$—then the cost of privacy dominates. Specifically, the cost of Lemma 6.4 is $O(\sqrt{\frac{k^2\ln(m/\delta)}{n\varepsilon^2}})$, and the cost of Theorem 6.6 is $O(\sqrt{\frac{k^3}{\alpha^2 n}\max\{\ln(\frac{m}{\delta}), \alpha\}})$. Given ε, the error will be smaller if

$$\alpha > \begin{cases} \varepsilon\sqrt{k} & \varepsilon < \frac{1}{\sqrt{k}}\ln(\frac{m}{\delta}) \\ \varepsilon^2\frac{k}{\ln(m/\delta)} & \text{otherwise} \end{cases}$$

i.e. if there is a *gap* between α, ε of size at least $\sqrt{k}$. This is possible if $k \ll \frac{1}{r}$, and for these instances d_{EM} DP offers better utility than user-level DP. In Theorem 6.6, the super-linear factor of $k^{3/2}$ comes from the fact that the k-RR is suboptimal in terms of k (Acharya et al., 2018).

6.2.2 Algorithms in the Central Model

The Laplace mechanism has been shown to be optimal for many instances of frequency estimation (Dwork et al., 2014). To attain user-level privacy, the baseline Laplace mechanism releases, for each $x \in \mathcal{X}$, the values $F_x = \tilde{K}_G(x) + Y$, where $Y \sim Lap(\frac{1}{n\varepsilon})$. The distribution function $\tilde{H}$ is then the normalization of $\langle F_x : x \in \mathcal{X}\rangle$. This gives us the following guarantee.

Lemma 6.7. *For the metric space $\mathcal{X} = \mathcal{B} \times \mathcal{C}$, the Laplace mechanism described above satisfies $(\varepsilon, 0)$-user level DP, and produces an estimate $\tilde{H}$ such that*

$$\max_K \mathbb{E}[d_{\mathsf{EM}}(\tilde{K}, \tilde{H})] \leq O\left(\tfrac{k}{n\varepsilon}\right).$$

Again, this utility does not depend on m, since each user contributes $\frac{1}{n}$ fraction of the whole dataset which is independent of m. Consistent with central DP, the error decreases with $\frac{1}{n}$, which is much faster than the $\frac{1}{\sqrt{n}}$ in the local model.

It is possible to adapt FreqEstLocal to bounded central d_{EM}-DP by simply pretending to be one user who holds the global dataset K_G. The privacy analysis of Theorem 4.4, and the utility analysis of Theorem 6.6 may be combined for the following corollary:

Corollary 6.8. *For the metric space $\mathcal{X} = \mathcal{B} \times \mathcal{C}$, FreqEstLocal with $\mathcal{A} = \mathsf{GKRR}_{\alpha_0}$ with α_0 given in Eq.* (7) *satisfies (α, δ)-d_{EM} DP in the central model and returns an estimator $\tilde{H}$ with error given in Eq.* (9)*.*

Remarks. In particular

$$\max_K \mathbb{E}[d_{\mathsf{EM}}(\tilde{H}, \tilde{K})] \leq 4\frac{\sqrt{k^3}}{\sqrt{mn}} + 64\frac{\sqrt{k^3}}{\alpha n}\sqrt{\ln(4m\exp(\alpha)/\delta)}.$$

The same sampling error is present, but the cost of privacy is reduced from a $\frac{1}{\sqrt{n}}$ dependence in Theorem 6.6 to just $\frac{1}{n}$. To compare just the cost of privacy in Corollary 6.6 to Lemma 6.7, we will assume we are in the regime $n \leq \frac{m}{\alpha}$. Then, the cost in Corollary 6.8 is $O(\frac{\sqrt{k^3}}{\alpha n})\sqrt{\max\{\ln(\frac{m}{\delta}), \alpha\}}$. The error of Corollary 6.8 is less when

$$\alpha \geq \begin{cases} \varepsilon\sqrt{k\ln(\frac{m}{\delta})} & \varepsilon \leq \sqrt{\frac{\ln(m/\delta)}{k}} \\ \varepsilon^2 k & \text{otherwise} \end{cases}$$

Thus, the utility is improved when α is bigger than ε by a factor of at least $\sqrt{k}$, which is achieved when $k \ll \frac{1}{r}$. One final advantage of Corollary 6.8 is that it may be implemented in the shuffle model of DP, which requires less trust than the central model. This parallels prior results of the shuffle model of DP (Feldman et al., 2022).

Open Questions. For linear queries, one open question is whether it is possible to obtain error competitive with user-level DP under $(\alpha, 0)$-d_{EM}-DP. For frequency estimation, an immediate open question is whether it is possible to reduce the super-linear dependence k to one that matches that of Lemma 6.4, and whether an improvement in error can be made when $\alpha \leq \varepsilon^2 k$.

New Proof Techniques. Frequency estimation is a classic problem that has been thoroughly studied in the standard DP literature. Despite the well-studied nature of the problem, the use of d_{EM}-DP enables a novel mechanism design. In particular, we can work with a transition matrix (corresponding to the private mechanism $\mathcal{A}$) that is less noisy. Consequently, we use a natural generalization of k-Randomized Response, which allows for better utility analysis than standard DP.

7 Related Work

Item-level DP. DP was originally considered at the item-level (Dwork, 2006). Relevant to our setting are results in distribution estimation (Hay et al., 2009; Xu et al., 2013; Suresh, 2019); these results study more complex estimation problems than frequency. We also consider linear query release (Hardt and Talwar, 2010; Bhaskara et al., 2012; Nikolov et al., 2013; Blum et al., 2013; Li et al., 2015). The mechanism in Hardt and Talwar (2010) is often optimal and easy to adapt to our setting; we compare our algorithms with it.

User-level DP. User-level privacy is gaining increasing interest (Amin et al., 2019; Narayanan et al., 2022; Bassily and Sun, 2023; Levy et al., 2021; Liu et al., 2020; Cummings et al., 2022). The most relevant work to ours involves user-level private mean estimation (Cummings et al., 2022) and histogram estimation (Liu et al., 2023; Acharya et al., 2023), though these problems are more complex than those we study. Another related area is deciding the amount of data to collect from each user when users have varying amounts of data (Amin et al., 2019; Liu et al., 2023; Cummings et al., 2022), which relates to our unbounded DP setup. These techniques apply to more specialized settings than our general blackbox reduction and are not immediately comparable.

Local DP. The results most relevant to our work in local DP are locally-private linear query release Duchi et al. (2013); Bassily (2019) and distribution estimation (Duchi et al., 2013; Kairouz et al., 2016; Acharya et al., 2018; Chen et al., 2020; Acharya et al., 2023). We directly compare our work to the optimal algorithms in Bassily (2019) and Chen et al. (2020) for our problems, which can be adapted to user-level DP easily. The other related line of work is privacy amplification via shuffling (Erlingsson et al., 2019; Girgis et al., 2021; Feldman et al., 2022). We extend the state-of-the-art analysis in Feldman et al. (2022) to general metric DP.

Metric DP. Metric DP was first proposed in Chatzikokolakis et al. (2013) in the central model. In the local model, this has led to work on releasing numeric data (Roy Chowdhury et al., 2022), location data (Andrés et al., 2013; Bordenabe et al., 2014; Chatzikokolakis et al., 2015; Weggenmann and Kerschbaum, 2021) and text (Feyisetan et al., 2019, 2020; Feyisetan and Kasiviswanathan, 2021; Imola et al., 2022). Unlike these works, we consider privacy in a general metric space. The most related work is that of Fernandes et al. (2019), which proposes metric DP based on d_{EM} for releasing text embeddings. As explained in the introduction, we consider a more general setting than Fernandes et al. (2019).

8 Interpretation of d_{EM}-DP

In this section, we elaborate on how to interpret the d_{EM}-DP guarantee (and metric DP in general) relative to standard DP. We start by discussing the advantages offered by d_{EM}-DP. As discussed in Section 1, the primary benefit of d_{EM}-DP is that it offers a more fine-grained and nuanced privacy definition compared to standard DP. This results in a more flexible privacy-utility trade-off that is better suited than standard DP for many real-world applications. In addition, d_{EM}-DP unlocks new proof techniques that may also be applicable to standard DP. Specifically, the d_{EM} metric introduces couplings that need to be explicitly addressed in privacy analysis. For instance, in Section 5 we showed that sampling with replacement is a smooth projection by explicitly viewing sampling from two datasets in terms of a coupling between them. While standard DP privacy analysis often implicitly uses couplings, we believe that some of our proof techniques for explicitly handling general couplings could also be beneficial in the context of standard DP.

Next, let us understand the technical relation between metric DP and standard DP. Metric DP is essentially a relaxation of standard DP. Any mechanism that satisfies metric DP (user-level or item-level) also satisfies standard DP, albeit with a potentially higher privacy parameter. In particular, any mechanism satisfying (α, δ)-$d_{\mathcal{X}}$-DP also satisfies $(\alpha \cdot d_{max}, \delta)$-DP, where d_{max} is the maximum $d_{\mathcal{X}}$ distance between any two pairs of inputs. Conversely, any mechanism that satisfies (ε, δ)-DP also satisfies $(\frac{\varepsilon}{d_{min}}, \delta)$-$d_{\mathcal{X}}$-D,P where d_{min} is the minimum $d_{\mathcal{X}}$ distance between any two pairs of inputs. Hence, although in theory one can translate between these two privacy guarantees, the translation is very loose. Tightly analyzing the privacy parameter under metric DP (whether d_{EM} or otherwise) for an arbitrary mechanism that satisfies standard DP is non-trivial and there is no one-size-fits-all method to do so. For instance, Algorithm 1 can be instantiated via the Laplace and Gaussian mechanism – both classic standard DP mechanisms – under some conditions. However, as discussed in Section 4.1, a more complex analysis is required to evaluate privacy under d_{EM}-DP. Additionally, in terms of mechanism design, a mechanism optimized for d_{EM}-DP might not be ideal for standard DP and vice-versa. For instance, our proposed Algorithm 4 for performing frequency estimation may not work well under standard DP (i.e., have high privacy parameters).

In what follows, we outline three concrete scenarios, where a practitioner should prefer d_{EM}-DP over standard DP. First, if the practitioner has a prior on the sensitive data indicating that the distance between any two user's data is overwhelmingly likely to be < 1 (assuming all distances are normalized), then d_{EM}-DP is clearly the better choice. This is because such a prior makes the worst-case scenario of antipodal data pairs—where two users' data are completely dissimilar (the case that standard DP safeguards against)—highly unlikely in practice. For instance, this scenario may arise in the context of location data when the data corresponds to location information of employees of the same firm. In this case, weekday locations will be the same across all users, leading to small pairwise d_{EM} distances. Second, a practitioner should opt for d_{EM}-DP when it captures a more realistic privacy semantics of the underlying data. Although ideally, we would like to prevent any data leakage about an individual, this is unfortunately not feasible in practice due to the vast amount of auxiliary information already publicly available about every individual. For instance, most people's occupations are publicly available on social media profiles. Returning to our location data example, data collected over a month would reveal routine patterns, such as a person's workplace, which is already public and hence doesn't require protection. Rather, the more sensitive information is short-term location data gathered over say the course of a day (which might reveal non-routine visit to a friend or hospital). In such a scenario, d_{EM}-DP would offer a better privacy-utility trade-off with more realistic privacy guarantees than standard DP. Third, d_{EM}-DP may be preferable if the practitioner is restricted to work within a low privacy parameter regime (for instance, due to some government guideline). This is because for the same privacy parameter (i.e., $\alpha = \varepsilon$), d_{EM}-DP can offer a stronger privacy guarantee than standard DP while maintaining the same utility for certain queries, such as linear queries (Section 4.1).

Finally, we conclude with some caveats regarding the use of metric DP. Metric DP assigns varying levels of sensitivity to different neighboring pairs. Specifically, smaller changes between neighboring pairs are considered more sensitive and are therefore protected with a higher privacy guarantee. However, this approach may not be suitable for all contexts. For instance, in the case of medical records, where the data between individuals can be vastly different, standard DP may offer better privacy protection. When adopting metric DP, it is crucial for practitioners to clearly define what is considered sensitive and what is not, and to engage in discussions about whether these definitions align with acceptable privacy semantics. This transparency enables users to make informed decisions about whether the privacy guarantees provided meet their needs.

9 Conclusion

We have proposed metric DP at the user level using the earth-mover's distance, d_{EM}. This captures both the magnitude and structural aspects of changes in the data, resulting in a tailored privacy semantic. We have designed two novel privacy mechanisms under d_{EM}-DP which improves the utility over standard DP. Additionally, we have shown that general (unbounded) d_{EM}-DP can be reduced to the simpler case (bounded) where all users have the same amount of data. Finally, we have demonstrated that d_{EM}-DP, when tailored to the application, can offer improved utility over standard DP.

References

John M Abowd. 2018. The US Census Bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 2867–2867.

Jayadev Acharya, Yuhan Liu, and Ziteng Sun. 2023. Discrete distribution estimation under user-level local differential privacy. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 8561–8585.

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. 2018. Communication Efficient, Sample Optimal, Linear Time Locally Private Discrete Distribution Estimation. *CoRR* abs/1802.04705 (2018). arXiv:1802.04705 `http://arxiv.org/abs/1802.04705`

Mário Alvim, Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Anna Pazii. 2018. Invited Paper: Local Differential Privacy on Metric Spaces: Optimizing the Trade-Off with Utility. In *2018 IEEE 31st Computer Security Foundations Symposium (CSF)*. 262–267. `https://doi.org/10.1109/CSF.2018.00026`

Kareem Amin, Alex Kulesza, Andres Munoz, and Sergei Vassilvtiskii. 2019. Bounding user contributions: A bias-variance trade-off in differential privacy. In *International Conference on Machine Learning*. PMLR, 263–271.

Miguel E Andrés, Nicolás E Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. 2013. Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. 901–914.

Borja Balle, Gilles Barthe, and Marco Gaboardi. 2018. Privacy amplification by subsampling: Tight analyses via couplings and divergences. *Advances in neural information processing systems* 31 (2018).

Gilles Barthe and Federico Olmedo. 2013. Beyond differential privacy: Composition theorems and relational logic for f-divergences between probabilistic programs. In *International Colloquium on Automata, Languages, and Programming*. Springer, 49–60.

Raef Bassily. 2019. Linear queries estimation with local differential privacy. In *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 721–729.

Raef Bassily and Ziteng Sun. 2023. User-level private stochastic convex optimization with optimal rates. In *International Conference on Machine Learning*. PMLR, 1838–1851.

Aditya Bhaskara, Daniel Dadush, Ravishankar Krishnaswamy, and Kunal Talwar. 2012. Unconditional differentially private mechanisms for linear queries. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. 1269–1284.

Avrim Blum, Katrina Ligett, and Aaron Roth. 2013. A learning theory approach to noninteractive database privacy. *Journal of the ACM (JACM)* 60, 2 (2013), 1–25.

Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. 2014. Optimal Geo-Indistinguishable Mechanisms for Location Privacy. *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (Nov. 2014), 251–262. `https://doi.org/10.1145/2660267.2660345` arXiv: 1402.5029.

Konstantinos Chatzikokolakis, Miguel E Andrés, Nicolás Emilio Bordenabe, and Catuscia Palamidessi. 2013. Broadening the scope of differential privacy using metrics. In *PETS*.

Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Marco Stronati. 2015. Constructing elastic distinguishability metrics for location privacy. *arXiv preprint arXiv:1503.00756* (2015).

Wei-Ning Chen, Peter Kairouz, and Ayfer Ozgur. 2020. Breaking the communication-privacy-accuracy trilemma. *Advances in Neural Information Processing Systems* 33 (2020), 3312–3324.

Graham Cormode, Somesh Jha, Tejas Kulkarni, Ninghui Li, Divesh Srivastava, and Tianhao Wang. 2018. Privacy at scale: Local differential privacy in practice. In *Proceedings of the 2018 International Conference on Management of Data*. 1655–1658.

Imre Csiszár. 1975. I-divergence geometry of probability distributions and minimization problems. *The annals of probability* (1975), 146–158.

Rachel Cummings, Vitaly Feldman, Audra McMillan, and Kunal Talwar. 2022. Mean estimation with user-level privacy under data heterogeneity. *Advances in Neural Information Processing Systems* 35 (2022), 29139–29151.

John C Duchi, Michael I Jordan, and Martin J Wainwright. 2013. Local privacy, data processing inequalities, and minimax rates. *arXiv preprint arXiv:1302.3203* (2013).

Cynthia Dwork. 2006. Differential privacy. In *International colloquium on automata, languages, and programming*. Springer, 1–12.

Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (2014), 211–407.

Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. 2019. Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2468–2479.

Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. 1054–1067.

Vitaly Feldman, Audra McMillan, and Kunal Talwar. 2022. Hiding among the clones: A simple and nearly optimal analysis of privacy amplification by shuffling. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 954–964.

Natasha Fernandes, Mark Dras, and Annabelle McIver. 2019. Generalised differential privacy for text document processing. In *Principles of Security and Trust: 8th International Conference, POST 2019, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2019, Prague, Czech Republic, April 6–11, 2019, Proceedings 8*. Springer International Publishing, 123–148.

Oluwaseyi Feyisetan, Borja Balle, Thomas Drake, and Tom Diethe. 2020. Privacy-and utility-preserving textual analysis via calibrated multivariate perturbations. In *Proceedings of the 13th international conference on web search and data mining*. 178–186.

Oluwaseyi Feyisetan, Tom Diethe, and Thomas Drake. 2019. Leveraging hierarchical representations for preserving privacy and utility in text. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 210–219.

Oluwaseyi Feyisetan and Shiva Kasiviswanathan. 2021. Private release of text embedding vectors. In *Proceedings of the First Workshop on Trustworthy Natural Language Processing*. 15–27.

Antonious M Girgis, Deepesh Data, Suhas Diggavi, Ananda Theertha Suresh, and Peter Kairouz. 2021. On the renyi differential privacy of the shuffle model. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. 2321–2341.

Clark R Givens and Rae Michael Shortt. 1984. A class of Wasserstein metrics for probability distributions. *Michigan Mathematical Journal* 31, 2 (1984), 231–240.

Moritz Hardt and Kunal Talwar. 2010. On the geometry of differential privacy. In *Proceedings of the forty-second ACM symposium on Theory of computing*. 705–714.

Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer. 191–216 pages.

Michael Hay, Vibhor Rastogi, Gerome Miklau, and Dan Suciu. 2009. Boosting the accuracy of differentially-private histograms through consistency. *arXiv preprint arXiv:0904.0942* (2009).

Jacob Imola, Amrita Roy Chowdhury, and Kamalika Chaudhuri. 2024. Metric Differential Privacy at the User-Level. *arXiv preprint arXiv:2405.02665* (2024).

Jacob Imola, Shiva Kasiviswanathan, Stephen White, Abhinav Aggarwal, and Nathanael Teissier. 2022. Balancing utility and scalability in metric differential privacy. In *Uncertainty in Artificial Intelligence*. PMLR, 885–894.

Peter Kairouz, Keith Bonawitz, and Daniel Ramage. 2016. Discrete distribution estimation under local privacy. In *International Conference on Machine Learning*. PMLR, 2436–2444.

Peter Kairouz, Sewoong Oh, and Pramod Viswanath. 2015. The composition theorem for differential privacy. In *International conference on machine learning*. PMLR, 1376–1385.

Dénes Konig. 2001. *Theorie der endlichen und unendlichen Graphen*. Vol. 72. American Mathematical Soc.

Daniel Levy, Ziteng Sun, Kareem Amin, Satyen Kale, Alex Kulesza, Mehryar Mohri, and Ananda Theertha Suresh. 2021. Learning with User-Level Privacy. In *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (Eds.), Vol. 34. Curran Associates, Inc., 12466–12479. https://proceedings.neurips.cc/paper_files/paper/2021/file/67e235e7f2fa8800d8375409b566e6b6-

Chao Li, Gerome Miklau, Michael Hay, Andrew McGregor, and Vibhor Rastogi. 2015. The matrix mechanism: optimizing linear counting queries under differential privacy. *The VLDB journal* 24 (2015), 757–781.

N. Li, M. Lyu, D. Su, and W. Yang. 2016. *Differential Privacy: From Theory to Practice*. Morgan and Claypool. https://ieeexplore.ieee.org/document/7731575

Yuhan Liu, Ananda Theertha Suresh, Felix Xinnan X Yu, Sanjiv Kumar, and Michael Riley. 2020. Learning discrete distributions: user vs item-level privacy. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 20965–20976. https://proceedings.neurips.cc/paper_files/paper/2020/file/f06edc8ab534b2c7ecbd4c2051d9cb1e-

Yuhan Liu, Ananda Theertha Suresh, Wennan Zhu, Peter Kairouz, and Marco Gruteser. 2023. Algorithms for bounding contribution for histogram estimation under user-level privacy. In *International Conference on Machine Learning*. PMLR, 21969–21996.

Shyam Narayanan, Vahab Mirrokni, and Hossein Esfandiari. 2022. Tight and robust private mean estimation with few users. In *International Conference on Machine Learning*. PMLR, 16383–16412.

Aleksandar Nikolov, Kunal Talwar, and Li Zhang. 2013. The geometry of differential privacy: the sparse and approximate cases. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. 351–360.

Amrita Roy Chowdhury, Bolin Ding, Somesh Jha, Weiran Liu, and Jingren Zhou. 2022. Strengthening Order Preserving Encryption with Differential Privacy. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security* (Los Angeles, CA, USA) *(CCS '22)*. Association for Computing Machinery, New York, NY, USA, 2519–2533. https://doi.org/10.1145/3548606.3560610

Ananda Theertha Suresh. 2019. Differentially private anonymized histograms. *Advances in Neural Information Processing Systems* 32 (2019).

Benjamin Weggenmann and Florian Kerschbaum. 2021. Differential privacy for directional data. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. 1205–1222.

Jia Xu, Zhenjie Zhang, Xiaokui Xiao, Yin Yang, Ge Yu, and Marianne Winslett. 2013. Differentially private histogram publication. *The VLDB journal* 22 (2013), 797–822.

A Omitted Technical Details

An alternative characterization of differential privacy is through the hockey-stick divergence (Barthe and Olmedo, 2013). For probability distributions P, Q defined on a space $\mathcal{Y}$, this is given by the following:

Definition A.1. *Let $\varepsilon, \delta > 0$, and let P, Q be distributions on a space $\mathcal{Y}$. The Hockey Stick Divergence is given by*

$$D_{e^\varepsilon}(P\|Q) = \int_{\mathcal{Y}} \max\left\{\frac{P(y)}{Q(y)} - e^\varepsilon, 0\right\} Q(y)dy.$$

It is easy to show that $D_{e^\varepsilon}(M(K)\|M(K')) \leq \delta$ implies (1), so Definition A.1 provides an alternative way to prove privacy.

Definition A.1 satisfies a number of useful properties. First, because it is an f-divergence (Csiszár, 1975), it satisfies the *data-processing inequality*: for any function f, we have

$$D_{e^\varepsilon}(f(P)\|f(Q)) \le D_{e^\varepsilon}(f(P)\|f(Q)).$$

This property is used to show that DP is invariant to post-processing by any function f. The second property, again holding for all f-divergences, is *convexity*. This states that for two pairs of distributions $P_1, P_2, Q_1, Q_2 \in \Delta^{\mathcal{Y}}$ and a real number $\lambda \in [0,1]$ we have

$$D_{e^\varepsilon}(\lambda P_1 + (1-\lambda)P_2\|\lambda Q_1 + (1-\lambda)Q_2) \\ \le \lambda D_{e^\varepsilon}(P_1\|Q_1) + (1-\lambda)D_{e^\varepsilon}(P_2\|Q_2).$$

Stated in terms of couplings, we may generalize convexity as follows:

Lemma A.1. *Suppose $X, Y \in \mathcal{X}$ are random variables with probability distributions $P_X, P_Y \in \Delta^{\mathcal{X}}$. Suppose $\mathcal{M} : \mathcal{X} \to \mathcal{Y}$ is a randomized function. Then, for any coupling $C \in \mathcal{C}(P_X, P_Y)$, we have*

$$D_{e^\varepsilon}(\mathcal{M}(X)\|\mathcal{M}(Y)) \le \mathbb{E}_{(x,y)\sim C}[D_{e^\varepsilon}(\mathcal{M}(x)\|\mathcal{M}(y))].$$

Proof. We may write

$$\mathcal{M}(X) = \sum_{x\in\mathcal{X}} P_X(x)\mathcal{M}(x) = \sum_{x,y\in\mathcal{X}} C(x,y)\mathcal{M}(x)$$
$$\mathcal{M}(Y) = \sum_{x,y\in\mathcal{X}} C(x,y)\mathcal{M}(y).$$

Applying convexity, we have

$$D_{e^\varepsilon}(\mathcal{M}(X)\|\mathcal{M}(Y)) \le \sum_{x,y\in\mathcal{X}} C(x,y) D_{e^\varepsilon}(\mathcal{M}(x)\|\mathcal{M}(y)),$$

and the claim follows. □

Third, D_{e^ε} satisfies a "weak" triangle inequality (also known as group privacy):

Lemma A.2. *For distributions P, Q, R on $\mathcal{Y}$, we have $D_{e^{\alpha+\beta}}(P\|R) \le D_{e^\alpha}(P\|Q) + e^\alpha D_{e^\beta}(Q\|R)$.*

Proof. For any P, Q, ε, we may view $D_{e^\varepsilon}(P\|Q)$ through its dual form as

$$D_{e^\varepsilon}(P\|Q) = \sup_{Y\subseteq\mathcal{Y}} (P(Y) - e^\varepsilon Q(Y)).$$

Thus, let Y^* denote the maximal set such that

$$D_{e^{\alpha+\beta}}(P\|R) = (P(Y^*) - e^{\alpha+\beta}R(Y^*)).$$

We may rewrite this as

$$D_{e^{\alpha+\beta}}(P\|R) = (P(Y^*) - e^\alpha Q(Y^*)) + e^\alpha(Q(Y^*) - e^\beta R(Y^*)) \\ \le D_{e^\alpha}(P\|Q) + e^\alpha D_{e^\beta}(Q\|R),$$

showing the claim. □

B Omitted Proofs from Section 4

B.1 Proof of Theorem 4.1

Theorem 4.1. *Let $q_f(K)$ be a linear query of the form in (5), where $f : \mathcal{X} \to \mathbb{R}^d$ is ℓ-Lipschitz. Then, we have $\Delta_{\mathsf{EM}}(q_f) \leq \ell$.*

For any two distributions $\tilde{K}, \tilde{K}'$, we have

$$\begin{aligned} q_f(K) - q_f(K') &= \mathbb{E}_{x\sim\tilde{K}}[f(x)] - \mathbb{E}_{x\sim\tilde{K}'}[f(x)] \\ &= \sum_{x\in\mathcal{X}} f(x)\tilde{K}(x) - \sum_{x\in\mathcal{X}} f(y)\tilde{K}'(y). \end{aligned}$$

Let $C(x,y) = \{C_x(y)\}_{x\in\mathcal{X}}$ be the minimum-transport coupling between $\tilde{K}$ and $\tilde{K}'$. By Definition 2.4, we have $\tilde{K}'(y) = \sum_{x\in\mathcal{X}} C(x,y)$, and $d_{\mathsf{EM}}(\tilde{K}, \tilde{K}') = \sum_{x,y\in\mathcal{X}} d_{\mathcal{X}}(x,y)C(x,y)$. Now, we write

$$\begin{aligned} &\sum_{x\in\mathcal{X}} f(x)\tilde{K}(x) - \sum_{y\in\mathcal{X}} f(y)\tilde{K}'(y) \\ &\quad= \sum_{x\in\mathcal{X}} f(x)\tilde{K}(x) - \sum_{y\in\mathcal{X}} f(y) \sum_{x\in\mathcal{X}} C(x,y) \\ &\quad= \sum_{x\in\mathcal{X}} \left(f(x) - \sum_{y\in\mathcal{X}} f(y)C_x(y) \right) \tilde{K}(x) \\ &\quad= \sum_{x\in\mathcal{X}} \left(\sum_{y\in\mathcal{X}} f(x)C_x(y) - \sum_{y\in\mathcal{X}} f(y)C_x(y) \right) \tilde{K}(x) \\ &\quad= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{X}} \left(f(x) - f(y)\right) C_x(y)\tilde{K}(x) \\ &\quad= \sum_{x,y\in\mathcal{X}} \left(f(x) - f(y)\right) C(x,y). \end{aligned}$$

By the triangle inequality and the fact that f is ℓ-Lipschitz, we may write

$$\begin{aligned} \|q_f(K) - q_f(K')\| &\leq \sum_{x,y\in\mathcal{X}} \|f(x) - f(y)\| C(x,y) \\ &\leq \sum_{x,y\in\mathcal{X}} \ell d_{\mathcal{X}}(x,y)C(x,y) \\ &= \ell d_{\mathsf{EM}}(\tilde{K}, \tilde{K}'). \end{aligned}$$

The last equation tells us that $\Delta_{d_{\mathsf{EM}}}(q_f) \leq \ell$.

B.2 Proof of Lemma 4.2

Lemma 4.2. *PrivEMDLinear (Algorithm 1) with scale $\omega = \frac{1}{\alpha}$ satisfies $(\alpha, 0)$-unbounded local d_{EM}-DP and with scale $\omega = \frac{1}{\alpha n}$ satisfies $(\alpha, 0)$-bounded central d_{EM}-DP.*

In the local model, by Theorem 4.1, we have $\|q_f(K) - q_f(K')\| \leq \ell$. By adding noise drawn from $\Gamma(d, \frac{1}{\alpha})$, it is known this satisfies $(\alpha, 0)$-DP Hardt and Talwar (2010). In the bounded central setting, we have $\|q_f(K) - q_f(K')\| \leq \frac{\ell}{n}$, and thus we may add noise drawn from $\Gamma(d, \frac{1}{n\alpha})$.

B.3 Proof of Theorem 4.3

Theorem 4.3. *Suppose that $(\mathcal{X}, d_{\mathcal{X}})$ is a metric space such that $d_{\mathcal{X}}(\cdot, \cdot) \leq 1$, and that $\mathcal{A}$ is an $(\alpha_0, 0)$ $d_{\mathcal{X}}$-DP algorithm. Let $(x_1, \ldots, x_m)$ and $(x'_1, \ldots, x'_m)$ be two vectors, and we define $v = (d_{\mathcal{X}}(x_i, x'_i))_{i=1}^m$. Let $0 < \delta < 1$ be a constant, and suppose it holds that $\alpha_0 < \ln(\frac{m}{16 \ln(4m/\delta)})$. Then, for all outputs O, we have that*

$$\Pr[\mathsf{Shuffle}(\mathcal{A}(x_1), \ldots, \mathcal{A}(x_m)) = O] \leq e^{\alpha} \Pr[\mathsf{Shuffle}(\mathcal{A}(x'_1), \ldots, \mathcal{A}(x'_m)) = O] + \delta e^{\alpha},$$

where

$$\alpha \leq \|v\|_0 \ln\left(1 + \frac{\exp(\alpha_0 \|v\|_1 / \|v\|_0) - 1}{\exp(\alpha_0 \|v\|_1 / \|v\|_0) + 1}\left(\frac{8\sqrt{e^{\alpha_0} \ln(4\|v\|_0/\delta)}}{\sqrt{m}} + \frac{8e^{\alpha_0}}{m}\right)\right).$$

We will first assume the following lemma:

Lemma B.1. *Suppose that $\mathcal{A}$ is an $\alpha_0 d_{\mathcal{X}}$-metric DP algorithm, where $d_{\mathcal{X}} \leq 1$. Let $x_1^0, x_1^1, x_2, \ldots, x_m \in \mathcal{X}$ be a set of inputs such that $d_{\mathcal{X}}(x_1^0, x_1^1) \leq d$, and let $\delta > 0$ be a constant such that $\alpha_0 \leq \ln(\frac{m}{16 \ln(2/\delta)})$. Then, we have that*

$$D_{e^{\alpha}}(\mathsf{Shuffle}(\mathcal{A}(x_1^0), \ldots, \mathcal{A}(x_m)), \mathsf{Shuffle}(\mathcal{A}(x_1^1), \mathcal{A}(x_2), \ldots, \mathcal{A}(x_m))) \leq \delta,$$

where

$$\alpha \leq \ln\left(1 + \frac{e^{\alpha_0 d} - 1}{e^{\alpha_0 d} + 1}\left(\frac{8\sqrt{e^{\alpha_0} \ln(4/\delta)}}{\sqrt{m}} + \frac{8e^{\alpha_0}}{m}\right)\right).$$

To prove Theorem 4.3, let

$$S(\mathbf{x}_i, \mathbf{x}'_{m-i}) = \mathsf{Shuffle}(\mathcal{A}(x_1), \ldots, \mathcal{A}(x_i), \mathcal{A}(x'_{i+1}), \ldots, \mathcal{A}(x'_m)).$$

Let $m' = \|v\|_0$, and WLOG suppose that $x_i = x'_i$ for $i > m'$. Our goal is to show that

$$D_{e^{\alpha}}(S(\mathbf{x}_{m'}, \mathbf{x}'_0) \| S(\mathbf{x}_0, \mathbf{x}'_{m'})) \leq \delta.$$

By Lemma B.1, we have for each $1 \leq i \leq m'$ that

$$D_{\exp(\alpha(i))}(S(\mathbf{x}_{i-1}, \mathbf{x}'_{m'-i+1}) \| S(\mathbf{x}_i, \mathbf{x}'_{m'-i})) \leq \frac{\delta}{m'},$$

where

$$\alpha(i) = \ln\left(1 + \frac{e^{\alpha_0 d_{\mathcal{X}}(x_i, x'_i)} - 1}{e^{\alpha_0 d_{\mathcal{X}}(x_i, x'_i)} + 1}\left(\frac{8\sqrt{e^{\alpha_0} \ln(4m'/\delta)}}{\sqrt{m}} + \frac{8e^{\alpha_0}}{m}\right)\right).$$

Applying Lemma A.2 m' times, we see

$$\begin{aligned}
&D_{\exp(\alpha(1)+\cdots+\alpha(m'))}(S(\mathbf{x}_{m'}, \mathbf{x}'_0) \| S(\mathbf{x}_0, \mathbf{x}'_{m'})) \\
&\leq D_{\exp(\alpha(m'))}(S(\mathbf{x}_{m'}, \mathbf{x}'_0) \| S(\mathbf{x}_{m'-1}, \mathbf{x}'_1)) \\
&+ e^{\alpha(m')} D_{\exp(\alpha(m'-1))}(S(\mathbf{x}_{m'-1}, \mathbf{x}'_1) \| S(\mathbf{x}_{m'-2}, \mathbf{x}'_2)) \\
&+ \cdots \\
&+ e^{\alpha(2)+\cdots+\alpha(m')} D_{\exp(\alpha(1))}(S(\mathbf{x}_1, \mathbf{x}'_{m'-1}) \| S(\mathbf{x}_0, \mathbf{x}'_{m'})) \\
&\leq e^{\alpha(1)+\cdots+\alpha(m')} \sum_{i=1}^{m'} D_{\exp(\alpha(i))}(S(\mathbf{x}_{i-1}, \mathbf{x}'_{m'-i+1}) \| S(\mathbf{x}_i, \mathbf{x}'_{m'-i})) \\
&\leq e^{\alpha(1)+\cdots+\alpha(m')} \delta.
\end{aligned}$$

We now show that $\alpha(i)$ is a concave function of $d_{\mathcal{X}}(x_i, x'_i)$; to do this we write $\alpha(i) = f(d) = \ln(1 + g(d)K)$, where $g(d) = \frac{e^d - 1}{e^d + 1}$ and $K > 0$ is a suitable constant. We will show that $f''(d) \leq 0$. Taking derivatives, it is easy to show that $f''(d)$ has the same sign as $(1 + Kg(d))g''(d) - Kg'(d)^2$. Thus, we will show that $(1 + Kg(d))g''(d) \leq Kg'(d)^2$. We may write

$$\begin{aligned}
g(d) &= 1 - \frac{2}{e^d + 1} \\
g'(d) &= \frac{2e^d}{(e^d + 1)^2} \\
g''(d) &= 2\frac{(e^d + 1)^2 e^d - 2e^d(e^d + 1)e^d}{(e^d + 1)^4} = 2\frac{e^d - e^{2d}}{(e^d + 1)^3}.
\end{aligned}$$

Now, we have

$$\begin{aligned}
&(1 + Kg(d))g''(d) \leq Kg'(d)^2 \\
\iff &(1 + K - \frac{2K}{e^d + 1})2\frac{e^d - e^{2d}}{(e^d + 1)^3} \leq K\frac{4e^{2d}}{(e^d + 1)^4} \\
\iff &((e^d + 1)(K + 1) - 2K)(1 - e^d) \leq 2Ke^d \\
\iff &(Ke^d - K + e^d + 1)(1 - e^d) \leq 2Ke^d \\
\iff &Ke^d - K + 1 - Ke^{2d} + Ke^d - e^{2d} \leq 2Ke^d \\
\iff &-K + 1 - Ke^{2d} - e^{2d} \leq 0
\end{aligned}$$

We are done by observing that $1 - e^{2d} \leq 0$, and $-K - Ke^{2d} \leq 0$. Having shown convexity, we establish the maximum occurs when each $\alpha(i)$ is equal to $\frac{\|v\|_1}{\|v\|_0}$. This gives us a bound of

$$\begin{aligned}
&\alpha(1) + \cdots + \alpha(m') \\
&\qquad \leq \|v\|_0 \ln\left(1 + \frac{e^{\alpha_0 \|v\|_1/\|v\|_0} - 1}{e^{\alpha_0 \|v\|_1/\|v\|_0} + 1}\left(\frac{8\sqrt{e^{\alpha_0} \ln(4\|v\|_0/\delta)}}{\sqrt{m}} + \frac{8e^{\alpha_0}}{m}\right)\right).
\end{aligned}$$

B.4 Proof of Lemma B.1

This lemma can be viewed as a generalization of amplification by shuffling, which has the same setup but sets $d = 1$ and merely requires that $\mathcal{M}$ satisfy ε-local DP. We generalize the approach of Feldman et al. (2022), starting with the the following preliminary claims.

B.4.1 Preliminary Lemmas

Lemma B.2. *(Generalization of Lemma 3.3 in Feldman et al. (2022)). Let $X = \{x_1^0, x_1^1, x_2 \ldots, x_m\}$ be a set of indices, and for $x \in X$, let $R(x), Q(x)$ be two families of distributions and $\alpha \in [0,1], \beta \in [0, \frac{1}{2}]$ be coefficients such that*

$$\begin{aligned} R(x_1^0) &= (1-\alpha)Q(x_1^0) + \alpha Q(x_1^1) \\ R(x_1^1) &= \alpha Q(x_1^0) + (1-\alpha)Q(x_1^1) \\ R(x_j) &= \beta Q(x_1^0) + \beta Q(x_1^1) + (1-2\beta)Q(x_j) \quad \forall j \geq 2. \end{aligned}$$

Then, there exists a post-processing mechanism $\mathcal{S}$ such that

$$\begin{aligned} \mathsf{Shuffle}(R(x_1^0), R(x_2), \ldots R(x_m)) &= \mathcal{S}(A + 1 - \Delta, C - A + \Delta) \qquad \textit{and} \\ \mathsf{Shuffle}(R(x_1^1), R(x_2), \ldots, R(x_m)) &= \mathcal{S}(A + \Delta, C - A + 1 - \Delta), \end{aligned}$$

where $C \sim Bin(s-1, 2\beta)$, $A \sim Bin(C, \frac{1}{2})$, and $\Delta \sim Bernoulli(\alpha)$, and $\mathsf{Shuffle}$ is a uniformly random shuffle.

Proof. Let $Y_1^0, Y_1^1, Y_2, \ldots, Y_m$ be distributions where Y_1^b is defined over $\{0,1\}$ and satisfies $Y_1^0(0) = 1-\alpha$ and $Y_1^1(1) = \alpha$ (with reversed probabilities if $b = 1$), and Y_j for $j \geq 2$ is defined over $\{0,1,2\}$ and satisfies $Y_j(0) = Y_j(1) = \beta$ and $Y_j(2) = 1 - 2\beta$. Let F be a function returning a distribution satisfying

$$F_j(v) = \begin{cases} Q(x_1^0) & v = 0 \\ Q(x_1^1) & v = 1 \\ Q(x_j) & \text{otherwise} \end{cases}$$

Observe that by definition, the following probability distributions are equal for $b \in \{0,1\}$:

$$R(x_1^b), R(x_2), \ldots, R(x_m) = F_1(Y_1^b), F_2(Y_2), \ldots, F_m(Y_m).$$

Let $\mathbf{0}(Y_1, \ldots, Y_m)$ denote the number of indices j such that $Y_j = 0$, and define $\mathbf{1}(Y_1, \ldots, Y_m)$ similarly. We will show that there exists a post-processing function $\mathcal{S}$ such that, for both $b \in \{0,1\}$, we have

$$\begin{aligned} &\mathsf{Shuffle}(F_1(Y_1^b), F_2(Y_2), \ldots, F_m(Y_m)) \\ &\qquad = \mathcal{S}(\mathbf{0}(Y_1^b, \ldots, Y_m), \mathbf{1}(Y_1^b, \ldots, Y_m)). \quad (10) \end{aligned}$$

We will do this by conditioning on the event $E_{u,v}$ that

$$(\mathbf{0}(Y_1^b, Y_2, \ldots, Y_m), \mathbf{1}(Y_1^b, Y_2, \ldots, Y_m)) = (u, v),$$

where $u, v \in \mathbb{N}$ satisfy $1 \leq u+v \leq m$. Now, define the vector $r = \mathsf{Shuffle}(F(Y_1), F_2(Y_2), \ldots, F_m(Y_m))$. Conditioned on $E_{u,v}$, r is distributed according to the following process: First, select a random

partition $U \sqcup V \sqcup W = [m]$ such that $|U| = u$ and $|V| = v$, corresponding to the indices (after shuffling) where $Y_1^b, Y_2, \ldots, Y_m$ are equal to $0, 1$, or 2. Next, let π be a random injection from W to $[m] \setminus 1$. Then, r is distributed according to:

$$r(u) = Q(x_1^0) \quad \forall u \in U \tag{11}$$

$$r(v) = Q(x_1^1) \quad \forall v \in V \tag{12}$$

$$r(w) = Q(x_{\pi(w)}) \quad \forall w \in W. \tag{13}$$

The above process is independent of α, β given $E_{u,v}$. In particular, it does not care whether we replace α with $1-\alpha$, and thus it serves as our process $\mathcal{S}$ satisfying (10) for both values of b. Having established this, it is easy to show that $\mathbf{0}(Y_1^0, \ldots, Y_m) = A + 1 - \Delta$, $\mathbf{1}(Y_1^0, \ldots, Y_m) = C - A + \Delta$ for $b = 0$, and $\mathbf{0}(Y_1^1, \ldots, Y_m) = A + \Delta$, $\mathbf{1}(Y_1^1, \ldots, Y_m) = C - A + 1 - \Delta$ for $b = 1$. □

Having reduced the shuffling problem to a divergence between two fixed probability distributions, we follow the method of Feldman et al. (2022) to compute this divergence. We use the following two results:

Lemma B.3. *(Restatement of Lemma A.1 from Feldman et al. (2022)): Suppose $p \geq \frac{16\ln(2/\delta)}{m}$, $C \sim Bin(m-1, p)$ and $A \sim Bin(C, \frac{1}{2})$. Define $P = (A+1, C-A)$ and $Q = (A, C-A+1)$. Then, $D_{e^\varepsilon}(P\|Q) \leq \delta$, where*

$$\varepsilon = \ln\left(1 + \frac{8\sqrt{\ln(4/\delta))}}{\sqrt{pm}} + \frac{8}{pm}\right)$$

The next result, advanced joint convexity, originally appeared in the privacy amplification by sampling literature and can be used to improve the parameter ε when computing $D_\alpha(P\|Q)$ between two distributions which are nearly the same.

Lemma B.4. *(Restatement of Theorem 2 from Balle et al. (2018)) Let P, Q be probability distributions satisfying $P = \nu M + (1-\nu)N$ and $Q = \nu M' + (1-\nu)N$ for distributions M, M', N and $\nu \in [0, 1]$. Given $\alpha \geq 1$, define $\alpha' = 1 + \nu(\alpha - 1)$ and $\beta = \frac{\alpha'}{\alpha}$. Then,*

$$D_{\alpha'}(P\|Q) \leq \nu D_\alpha(M\|(1-\beta)N + \beta M').$$

Finally, we require a result from local DP:

Lemma B.5. *(Restatement of Theorem 2.5 from Kairouz et al. (2015)) Let P, Q be two distributions and $\alpha \geq 1$ be a parameter such that $D_\alpha(P\|Q) = 0$. Then, there exist distributions M, N such that*

$$P = \frac{\alpha}{\alpha+1}M + \frac{1}{\alpha+1}N$$
$$Q = \frac{1}{\alpha+1}M + \frac{1}{\alpha+1}N.$$

With these results in order, we are ready to complete the proof.

B.4.2 Completing the proof of Lemma B.1

Using the definition of $d_{\mathcal{X}}$-DP and the fact that $d_{\mathcal{X}} \leq 1$, we have

$$D_{\exp(\varepsilon_0 d)}(\mathcal{A}(x_1^0) \| \mathcal{A}(x_1^1)) = 0$$
$$D_{\exp(\varepsilon_0)}(\mathcal{A}(x_1^0) \| \mathcal{A}(x_j)) = 0 \quad \forall j \geq 2$$
$$D_{\exp(\varepsilon_0)}(\mathcal{A}(x_1^1) \| \mathcal{A}(x_j)) = 0 \quad \forall j \geq 2.$$

Applying Lemma B.5 to the first equation, we obtain

$$\mathcal{A}(x_1^0) = (1-\beta)Q(x_1^0) + \beta Q(x_1^1) \tag{14}$$
$$\mathcal{A}(x_1^1) = \beta Q(x_1^0) + (1-\beta)Q(x_1^1) \tag{15}$$

where $\beta = \frac{1}{1+\exp(\varepsilon_0 d)}$. Applying the lemma to the second and third sets of equations, we obtain

$$\mathcal{A}(x_1^0) = (1-\gamma)R(x_1^0, x_j) + \gamma R'(x_1^0, x_j) \quad \forall j \geq 2 \tag{16}$$
$$\mathcal{A}(x_j) = \gamma R(x_1^0, x_j) + (1-\gamma)R'(x_1^0, x_j) \quad \forall j \geq 2 \tag{17}$$
$$\mathcal{A}(x_1^1) = (1-\gamma)R(x_1^1, x_j) + \gamma R'(x_1^1, x_j) \quad \forall j \geq 2 \tag{18}$$
$$\mathcal{A}(x_j) = \gamma R(x_1^1, x_j) + (1-\gamma)R'(x_1^1, x_j) \quad \forall j \geq 2. \tag{19}$$

where $\gamma = \frac{1}{1+\exp(\varepsilon_0)}$. Subtracting 16 and 17, we obtain that

$$\mathcal{A}(x_j) = \frac{\gamma}{1-\gamma}\mathcal{A}(x_1^0) + \frac{1-2\gamma}{1-\gamma}R'(x_1^0, x_j) \quad \forall j \geq 2, \tag{20}$$

and likewise 18 and 19 imply

$$\mathcal{A}(x_j) = \frac{\gamma}{1-\gamma}\mathcal{A}(x_1^1) + \frac{1-2\gamma}{1-\gamma}R'(x_1^1, x_j) \quad \forall j \geq 2. \tag{21}$$

Taking the average of 20 and 21, we obtain

$$\mathcal{A}(x_j) = \frac{\gamma}{2(1-\gamma)}\mathcal{A}(x_1^0) + \frac{\gamma}{2(1-\gamma)}\mathcal{A}(x_1^1) + \frac{1-2\gamma}{1-\gamma}Q(x_j) \quad \forall j \geq 2, \tag{22}$$

where $Q(x_j) = \frac{1}{2}R'(x_1^0, x_j) + \frac{1}{2}R'(x_1^1, x_j)$. Now, equations 14 and 15 imply that

$$\mathcal{A}(x_1^0) + \mathcal{A}(x_1^1) = Q(x_1^0) + Q(x_1^1).$$

This implies

$$\mathcal{A}(x_j) = \frac{\gamma}{2(1-\gamma)}Q(x_1^0) + \frac{\gamma}{2(1-\gamma)}Q(x_1^1) + \frac{1-2\gamma}{1-\gamma}Q(x_j) \quad \forall j \geq 2. \tag{23}$$

Applying Lemma B.2, there exists a function S such that

$$\mathsf{Shuffle}(\mathcal{A}(x_1^0), \mathcal{A}(x_2), \ldots, \mathcal{A}(x_m)) = S(A + 1 - \Delta, C - A + \Delta)$$
$$\mathsf{Shuffle}(\mathcal{A}(x_1^1), \mathcal{A}(x_2), \ldots, \mathcal{A}(x_m)) = S(A + \Delta, C - A + 1 - \Delta),$$

where $C \sim Bin(m-1, \frac{\gamma}{1-\gamma}) = Bin(m-1, e^{-\varepsilon_0})$, $A \sim Bin(C, \frac{1}{2})$, and $\Delta \sim Bernoulli(\beta)$. By the post-processing inequality, we have for any $\alpha \geq 1$ that

$$D_\alpha(\mathsf{Shuffle}(\mathcal{A}(x_1^0), \mathcal{A}(x_2), \ldots, \mathcal{A}(x_s)) \| \mathsf{Shuffle}(\mathcal{A}(x_1^1), \mathcal{A}(x_2), \ldots, \mathcal{A}(x_s))) \leq D_\alpha((A+1-\Delta, C-A+\Delta) \| (A+\Delta, C-A+1-\Delta)).$$

Observe we can write

$$(A+1-\Delta, C-A+\Delta) = (1-\beta)(A+1, C-A) + \beta(A, C-A+1)$$
$$(A+\Delta, C-A+1-\Delta) = \beta(A+1, C-A) + (1-\beta)(A, C-A+1).$$

Define $X = (A+1, C-A)$ and $Y = (A, C-A+1)$. We can rewrite the above as

$$(A+1-\Delta, C-A+\Delta) = 2\beta\frac{X+Y}{2} + (1-2\beta)X$$
$$(A+\Delta, C-A+1-\Delta) = 2\beta\frac{X+Y}{2} + (1-2\beta)Y.$$

Applying Lemma B.4, we have

$$D_{\alpha'}((A+1-\Delta, C-A+\Delta) \| (A+\Delta, C-A+1-\Delta)) \leq (1-2\beta) D_\alpha(X \| (1-\eta)(\tfrac{X+Y}{2}) + \eta Y),$$

where $\alpha' = 1 + (1-2\beta)(\alpha-1)$ and $\eta = \frac{\alpha'}{\alpha}$. By convexity, the RHS above is at most

$$D_{\alpha'}((A+1-\Delta, C-A+\Delta) \| (A+\Delta, C-A+1-\Delta)) \leq (1-2\beta) D_\alpha(X \| Y).$$

Now, we finally set $\alpha = 1 + \frac{8\sqrt{\exp(-\varepsilon_0)\ln(4/\delta)}}{\sqrt{m}} + \frac{8\exp(-\varepsilon_0)}{m}$. Lemma B.3 (using the assumption that $\varepsilon_0 \leq \ln(\frac{m}{16\ln(2/\delta)})$) implies $D_\alpha(X \| Y) \leq \delta$. From this, we obtain our desired result that

$$D_{\alpha'}(\mathsf{Shuffle}(\mathcal{A}(x_1^0), \mathcal{A}(x_2), \ldots, \mathcal{A}(x_m)) \| \mathsf{Shuffle}(\mathcal{A}(x_1^1), \mathcal{A}(x_2), \ldots, \mathcal{A}(x_m))) \leq (1-2\beta) D_\alpha(X \| Y) \leq \delta,$$

where

$$\alpha' = 1 + \frac{e^{\varepsilon_0 d} - 1}{e^{\varepsilon_0 d} + 1}\left(\frac{8\sqrt{e^{\varepsilon_0}\ln(4/\delta)}}{\sqrt{m}} + \frac{8e^{\varepsilon_0}}{m}\right).$$

B.5 Proof of Theorem 4.4

Theorem 4.4. *For any $\delta \in (0,1)$, PrivEMDItemWise shown in Algorithm 2 satisfies bounded local (α, δ')-d_{EM} DP, where*

$$\alpha = \sup_{w \in [0,1]} \frac{h(m; m, mw)}{w} \qquad \text{and } \delta' = \delta e^{h(m;m,m)},$$

and

$$h(m; x_0, x_1) = x_0 \ln\left(1 + \frac{\exp(\alpha_0 x_1/x_0) - 1}{\exp(\alpha_0 x_1/x_0) + 1}\left(\frac{8\sqrt{e^{\alpha_0}\ln(4x_0/\delta)}}{\sqrt{m}} + \frac{8e^{\alpha_0}}{m}\right)\right).$$

Similarly, PrivEMDItemWise satisfies bounded central (α, δ')-d_{EM} DP, where

$$\alpha = \sup_{w \in [0,1]} \frac{h(mn; m, mw)}{w} \qquad \text{and } \delta' = \delta e^{h(mn;m,m)}.$$

First, consider the local model. Fix any two itemsets $K = \{x_1, \ldots, x_m\}$ and $K' = \{x_1, \ldots, x'_m\}$ such that $d_{\mathsf{EM}}(\tilde{K}, \tilde{K}') \le w$. By Lemma 2.1, there exists a permutation $\pi : [m] \to [m]$ such that

$$\sum_{i=1}^{m} d_{\mathcal{X}}(x_i, x'_{\pi(i)}) = mw.$$

Let

$$\tilde{L} = \text{Shuffle}(\mathcal{A}(x_1), \ldots, \mathcal{A}(x_m)) \tag{24}$$

$$\tilde{L}' = \text{Shuffle}(\mathcal{A}(x'_{\pi(i)}), \ldots, \mathcal{A}(x'_{\pi(m)})). \tag{25}$$

By Theorem 4.3, we know that $D_{\exp(\alpha(w))}(\tilde{L} \| \tilde{L}') \le \delta e^{\alpha(w)}$, where $\alpha(w) = h(m; m, mw)$. The final privacy parameters for a fixed w will be $\frac{\alpha(w)}{w}$ and $\delta e^{\alpha(w)}$; the worst-case privacy parameters are thus $\sup_{w \in [0,1]} \frac{\alpha(w)}{w}$ and $\sup_{w \in [0,1]} \delta e^{\alpha(w)}$. Since $\alpha(w)$ is an increasing function, the latter term reduces to $\delta e^{\alpha(w)}$.

In the bounded central model, the same logic applies, except that $\tilde{L}, \tilde{L}'$ have size mn, differ in only m coordinates, and

$$\sum_{i=1}^{mn} d_{\mathcal{X}}(x_i, x'_{\pi(i)}) = mw.$$

We apply Theorem 4.3 to obtain $D_{\exp(\alpha(w))}(\tilde{L} \| \tilde{L}') \le \delta e^{\alpha(w)}$, where $\alpha(w) = h(mn; m, mw)$, and we complete the proof similarly.

C Omitted Proofs from Section 5

C.1 Proof of Lemma 5.2

Lemma 5.2. *Let $\tilde{K}, \tilde{K}' \in \Delta^{\mathcal{X}}$ be probability distributions, and let C^* be the minimum cost coupling between $\tilde{K}, \tilde{K}'$. Let $\{(x_i, y_i)\}_{i=1}^{s}$ be s i.i.d. samples from C^*, $L = (x_1, \ldots, x_s)$ and $L' = (y_1, \ldots, y_s)$. Then,*

$$\Pr[d_{\mathsf{EM}}(\tilde{L}, \tilde{L}') \ge (1 + \sqrt{2}) d_{\mathsf{EM}}(\tilde{K}, \tilde{K}') + \tfrac{3}{s} \ln(\tfrac{1}{\delta})] \le \delta.$$

For $i = 1, \ldots, s$, define $X_i = d_{\mathcal{X}}(x_i, y_i)$, and observe that $d_{\mathsf{EM}}(\tilde{L}, \tilde{L}') \le \frac{1}{s}(X_1 + \cdots + X_s)$. Now, let μ denote $d_{\mathsf{EM}}(\tilde{K}, \tilde{K}')$. Observe each X_i is i.i.d. and satisfies $\mathbb{E}[X_i] = \mu$ and $0 \le X_i \le 1$. Due to the last two facts, we have $\mathbb{E}[X_i^2] \le \mu$. By Bernstein's inequality, we have, for all $t \ge 0$,

$$\Pr\left[X_1 + \cdots + X_s - s\mu \ge t\right] \le e^{-t^2/2(v + bt/3)},$$

where $v = \sum_{i=1}^{s} \mathbb{E}[X_i^2] \le s\mu$ and $b = 1$. By setting

$$t = \max\{\sqrt{4s\mu \ln(1/\delta)}, \tfrac{4}{3} \ln(1/\delta)\},$$

we ensure that the probability is at most δ. We have

$$s\mu + t \le s\mu + 2\sqrt{s\mu \ln(1/\delta)} + \tfrac{4}{3} \ln(1/\delta) \le (1 + \sqrt{2}) s\mu + (\tfrac{4}{3} + \sqrt{2}) \ln \tfrac{1}{\delta}.$$

Finally,

$$\Pr[d_{\mathsf{EM}}(\tilde{L}, \tilde{L}') \ge (1 + \sqrt{2})\mu + \tfrac{3}{s} \ln \tfrac{1}{\delta}] \le$$
$$\Pr[X_1 + \cdots + X_s \ge (1 + \sqrt{2}) s\mu + 3 \ln \tfrac{1}{\delta}] \le \delta.$$

C.2 Proof of Theorem 5.3

Theorem 5.3. *Let $\varepsilon > 0$ and $\delta, r \in [0,1]$ be arbitrary constants. Suppose $\mathcal{A}$ is a mechanism which satisfies (α, δ)-bounded local d_{EM}-DP (Definition 3.1), where*

$$\alpha = \frac{\varepsilon}{(1+\sqrt{2})r + \frac{3}{s}\ln(\frac{1}{\delta})}.$$

Then, BoundedEMDReduction satisfies $(\varepsilon, 2\delta, r)$-discrete local d_{EM}-DP. Similarly, if $\mathcal{A}$ is (α, δ)-bounded central d_{EM}-DP (Definition 3.1), then BoundedEMDReduction is $(\varepsilon, 2\delta, r)$-discrete central d_{EM}-DP.

First, we will consider the local model. Let K, K' denote two datasets such that $d_{\mathsf{EM}}(\tilde{K}, \tilde{K'}) \le r$. Let L, L' denote the set of s samples when K (resp. K') is used. Our goal is to show that $D_{\exp(\varepsilon)}(\mathcal{M}(L) \| \mathcal{M}(L')) \le \delta$. Observe we may define the objects $\mathbf{L}, \mathbf{L}' \in \Delta^{\mathcal{X}^s}$ to be the probability distributions of L, L' (which lie in $\mathcal{X}^s$). By Lemma A.1, for any coupling $C \in \mathcal{C}(\mathbf{L}, \mathbf{L}')$, we have

$$D_{\exp(\varepsilon)}(\mathcal{M}(L) \| \mathcal{M}(L')) \le \mathbb{E}_{(L,L') \sim C}[D_{\exp(\varepsilon)}(\mathcal{M}(L) \| \mathcal{M}(L'))].$$

Let A denote the event that we have $d_{\mathsf{EM}}(\tilde{L}, \tilde{L'}) \le (1+\sqrt{2})r + \frac{3}{s}\ln\frac{1}{\delta}$. When A holds, then $D_{\exp(\varepsilon)}(\mathcal{M}(L) \| \mathcal{M}(L')) \le \delta$ by assumption. When this does not hold, then trivially $D_{\exp(\varepsilon)}(\mathcal{M}(L) \| \mathcal{M}(L')) \le 1$. Conditioning on the above expectation, we have

$$\begin{aligned}\mathbb{E}_{(L,L') \sim C}[D_{\exp(\varepsilon)}(\mathcal{M}(L) \| \mathcal{M}(L'))] &\le \delta \Pr[A] + \Pr[\overline{A}] \\ &\le \delta + \Pr[\overline{A}].\end{aligned}$$

Now, let $C^* \in \Delta^{\mathcal{X} \times \mathcal{X}}$ denote the optimal coupling between $\tilde{K}, \tilde{K'}$. We will take $C = (C^*)^s \in \Delta^{\mathcal{X}^s \times \mathcal{X}^s}$, the s-fold Kronecker product of C^*. Observe this is indeed a coupling between $\mathbf{L}, \mathbf{L}'$, and each coordinate of $(L, L') \sim C$ is simply a sample from C^*. Thus, the event A above is equivalent to

$$\Pr[A] = \Pr_{(L,L') \sim (C^*)^s}[d_{\mathsf{EM}}(\tilde{L}, \tilde{L'}) \le (1+\sqrt{2})r + \frac{3}{s}\ln\frac{1}{\delta}],$$

where the notation $(L, L') \sim (C^*)^s$ indicates that $L = \{x_1, \ldots, x_s\}$ and $L = \{y_1, \ldots, y_s\}$, and each $(x_i, y_i) \sim C^*$. By Lemma 5.2, we know that $\Pr[A] \ge 1 - \delta$, and thus the above expectation is at most 2δ. This proof may be generalized easily to the central model.

D Omitted Proofs from Section 6

D.1 Proof of Lemma 6.2

Lemma 6.2. *There exists an (α, δ)-bounded d_{EM}-DP algorithm in the local model which produces an estimate $\hat{q}$ such that, for all $\tilde{K}$,*

$$\mathbb{E}[\|\hat{q} - q_{f \circ \phi}(\tilde{K})\|_2] \le \|F\|_2 \frac{\sqrt{1.25d\ln(1/\delta))}}{\alpha\sqrt{n}}.$$

As the sensitivity of the query is bounded by $\|F\|_2$, is easy to show (e.g. Dwork et al. (2014)) that adding d-dimensional Gaussian noise with width $\|F\|_2 \frac{r\sqrt{1.25\ln\frac{1}{\delta}}}{\alpha}$ in each coordinate will satisfy $(\frac{\alpha}{r}, \delta)$ local d_{EM}-DP. The standard deviation in each coordinate of $\hat{q}$ is thus $\|F\|_2 \frac{r\sqrt{1.25\ln\frac{1}{\delta}}}{\alpha\sqrt{n}}$, and this gives the desired expected error.

D.2 Proof of Theorem 6.5

Theorem 6.5. *For the metric space $\mathcal{X} = \mathcal{B} \times \mathcal{C}$ and any mechanism $\mathcal{A}$ satisfying $(\alpha_0, 0)$ $d_{\mathcal{X}}$-DP where $\alpha_0 = O(\frac{\alpha}{\sqrt{m \ln(me^\alpha/\delta)}})$ (α_0 is specifically defined in Theorem 4.4), FreqEstLocal is (α, δ)-bounded d_{EM}-DP in the local model and returns an estimator $\tilde{H}$ such that*

$$\max_K \mathbb{E}[d_{\mathsf{EM}}(\tilde{H}, \tilde{K})] \le r\sqrt{\frac{st(\|B^T\|_{1\to 2}^2 - 1)}{mn}} + \sqrt{\frac{s(\|P^T B^T\|_{1\to 2}^2 - 1)}{mn}}, \tag{8}$$

where B is a right inverse of $\mathcal{A}$, $P = I_{\mathcal{B}} \otimes 1_{\mathcal{C}}^+$, and $1_{\mathcal{C}}^+$ is a column vector of 1s indexed by $\mathcal{C}$.

First, we will introduce notation. For a cluster label $b \in \mathcal{B}$, let $\mathcal{X}[b] \subseteq \mathcal{X}$ denote the elements of $\mathcal{X}$ in cluster b. Define $\tilde{F}[b] \in \mathbb{R}^{\mathcal{B}\times\mathcal{C}}$ to be the indices of $\tilde{F}$ in $\mathcal{X}[b]$ (so that indices outside $\mathcal{X}[b]$ are zeroed out). Define $\tilde{K}[b]$ similarly, and observe that $\tilde{F}[b], \tilde{K}[b]$ are not normalized.

For any estimate $\tilde{F}$, consider the following transportation plan from $\tilde{F}$ to $\tilde{K}$: For each $b \in \mathcal{B}$, transfer $\tilde{F}[b]$ to $\tilde{K}[b]$ arbitrarily, and put any excess weight in the bin (b, c') for an arbitrary $c' \in \mathcal{C}$. The cost incurred by this is at most $r\|\tilde{F}[b] - \tilde{K}[b]\|_1 + r|\mu(\tilde{F}[b]) - \mu(\tilde{K}[b])|$, where $\mu(\cdot)$ denotes total mass of its argument. Finally, equalize the weights in the coordinates $\{(b, c') : b \in \mathcal{B}\}$. The cost incurred for this step is at most $(1-r)\sum_{b\in\mathcal{B}} |\mu(\tilde{F}[b]) - \mu(\tilde{K}[b])|$. Thus, the total cost is

$$\sum_{b\in\mathcal{B}} r\|\tilde{F}[b] - \tilde{K}[b]\|_1 + |\mu(\tilde{F}[b]) - \mu(\tilde{K}[b])| \\ = r\|\tilde{F} - \tilde{K}\|_1 + \sum_{b\in\mathcal{B}} |\mu(\tilde{F}[b]) - \mu(\tilde{K}[b])|.$$

Observe that the term $\sum_{b\in\mathcal{B}} |\mu(\tilde{F}[b]) - \mu(\tilde{K}[b])|$ is simply the ℓ_1 distance between $\tilde{F}P$ and $\tilde{K}P$, where $P \in \mathbb{R}^{(\mathcal{B}\times\mathcal{C})\times\mathcal{B}}$ is the matrix that maps a vector to its sum along each coordinate in $\mathcal{B}$. Thus, we may form the the upper bound

$$\begin{aligned}
\mathbb{E}[d_{\mathsf{EM}}(\tilde{F}, \tilde{K})] &\le r\mathbb{E}[\|\tilde{F} - \tilde{K}\|_1] + \mathbb{E}[\|(\tilde{F} - \tilde{K})P\|_1] \\
&\le r\mathbb{E}[\sqrt{st}\|\tilde{F} - \tilde{K}\|_2] + \mathbb{E}[\sqrt{s}\|(\tilde{F} - \tilde{K})P\|_2] \\
&\le r\sqrt{st\mathbb{E}[\|\tilde{F} - \tilde{K}\|_2^2]} + \sqrt{s\mathbb{E}[\|(\tilde{F} - \tilde{K})P\|_2^2]}.
\end{aligned} \tag{26}$$

Now, we will bound (26) given this estimator. In the following, let A_x denote the xth row of the matrix A. Observe that

$$\begin{aligned}
\tilde{F} - \tilde{K} &= \frac{1}{mn}\sum_{i=1}^{mn} z_i B - \tilde{K}AB \\
&= \frac{1}{mn}\sum_{i=1}^{mn} z_i B - \frac{1}{mn}\sum_{i=1}^{mn} e_{k_i} AB \\
&= \frac{1}{mn}\sum_{i=1}^{mn} z_i B - \frac{1}{mn}\sum_{i=1}^{mn} A_{k_i} B \\
&= \frac{1}{mn}\sum_{i=1}^{mn} (z_i - A_{k_i})B
\end{aligned}$$

Define $w_i = z_i - A_{k_i}$, and notice that $\mathbb{E}[w_i] = \mathbb{E}[z_i] - A_{k_i} = 0$. Thus,

$$\begin{aligned}
\mathbb{E}[\|\tilde{F} - \tilde{K}\|_2^2] &= \mathbb{E}[(\tilde{F} - \tilde{K})(\tilde{F} - \tilde{K})^T] \\
&= \left(\frac{1}{mn}\right)^2 \mathbb{E}\left[\left(\sum_{i=1}^{mn} w_i B\right)\left(\sum_{i=1}^{mn} B^T w_i^T\right)\right] \\
&= \left(\frac{1}{mn}\right)^2 \sum_{i,j=1}^{mn} \mathbb{E}[w_i B B^T w_j^T] \\
&= \left(\frac{1}{mn}\right)^2 \sum_{i=1}^{mn} \mathbb{E}[w_i B B^T w_i^T],
\end{aligned}$$

where the last step holds because the w_i are independent. Now, we have

$$\begin{aligned}
\mathbb{E}[w_i B B^T w_i^T] &= \mathbb{E}[z_i B B^T z_i^T] - \mathbb{E}[A_{k_i} B B^T A_{k_i}^T] \\
&= \mathbb{E}[z_i B B^T z_i^T] - e_{k_i} e_{k_i}^T \\
&\leq \|B^T\|_{1,2}^2 - 1.
\end{aligned}$$

Putting it all together, we have

$$\mathbb{E}[\|\tilde{F} - \tilde{K}\|_2^2] \leq \frac{\|B^T\|_{1,2}^2 - 1}{mn}$$

To control the term $\|(\tilde{F} - \tilde{K})P\|_2^2$ in (26), using similar steps, we may write

$$\mathbb{E}[\|(\tilde{F} - \tilde{K})P\|_2^2] \leq \left(\frac{1}{mn}\right)^2 \sum_{i=1}^{mn} \mathbb{E}[w_i B P P^T B^T w_i^T].$$

Similarly, for any i we have

$$\begin{aligned}
\mathbb{E}[w_i B P P^T B^T w_i^T] &= \mathbb{E}[z_i B P P^T B^T z_i^T] - \mathbb{E}[A_{k_i} B P P^T B A_{k_i}^T] \\
&\leq \|P^T B^T\|_{1,2}^2 - 1,
\end{aligned}$$

and this implies

$$\mathbb{E}[\|(\tilde{F} - \tilde{K})P\|_2^2] \leq \frac{\|P^T B^T\|_{1,2}^2 - 1}{mn}.$$

Substituting into (26), we obtain the desired bound.

D.3 Proof of Theorem 6.6

Theorem 6.6. *For the metric space $\mathcal{X} = \mathcal{B} \times \mathcal{C}$, FreqEstLocal with the mechanism $\mathcal{A} = \mathsf{GKRR}_{\alpha_0}$ satisfies (α, δ)-d_{EM} DP in the local model and returns an estimator $\tilde{H}$ such that*

$$\max_K \mathbb{E}[d_{\mathsf{EM}}(\tilde{H}, \tilde{K})] \quad \leq \quad r\sqrt{\frac{st^3}{mn}}\left(\frac{e^{\alpha_0} + s}{e^{\alpha_0} - e^{(1-r)\alpha_0}}\right) \; + \; \sqrt{\frac{s^2 t^2}{mn}}\left(\frac{\sqrt{s + 2(e^{\alpha_0} - 1)}}{e^{\alpha_0} + (t-1)e^{(1-r)\alpha_0} - t}\right), \quad (9)$$

where α_0 is defined in Eq. (6)*.*

For positive constants a, b, c, the matrix A is given by

$$A = aI_{\mathcal{X}} + (bI_{\mathcal{B}} + c\mathbf{1}_{\mathcal{B}}) \otimes \mathbf{1}_{\mathcal{C}},$$

where

$$\begin{aligned} a &= \frac{e^{\alpha_0} - e^{(1-r)\alpha_0}}{e^{\alpha_0} + (t-1)e^{(1-r)\alpha_0} + (s-1)t} \\ b &= \frac{e^{(1-r)\alpha_0} - 1}{e^{\alpha_0} + (t-1)e^{(1-r)\alpha_0} + (s-1)t} \\ c &= \frac{1}{e^{\alpha_0} + (t-1)e^{(1-r)\alpha_0} + (s-1)t}. \end{aligned}$$

The matrix A is actually invertible, and

$$A^{-1} = a'I_{\mathcal{X}} + (b'I_{\mathcal{B}} + c'\mathbf{1}_{\mathcal{B}}) \otimes \mathbf{1}_{\mathcal{C}},$$

where

$$\begin{aligned} a' &= \frac{e^{\alpha_0} + (t-1)e^{(1-r)\alpha_0} + (s-1)t}{e^{\alpha_0} - e^{(1-r)\alpha_0}} \\ b' &= -\frac{(e^{(1-r)\alpha_0} - 1)(e^{\alpha_0} + (t-1)e^{(1-r)\alpha_0} + (s-1)t)}{(e^{\alpha_0} - e^{(1-r)\alpha_0})(e^{\alpha_0} + (t-1)e^{(1-r)\alpha_0} - t)} \\ c' &= -\frac{1}{e^{\alpha_0} + (t-1)e^{(1-r)\alpha_0} - t}. \end{aligned}$$

It is easy to show the identity that $a' + tb' + stc' = 1$. Each row of A^{-1} looks like one copy of $a' + b' + c'$, $t-1$ copies of $b' + c'$, and $(s-1)t$ copies of c'. Thus,

$$\begin{aligned} &\|(A^{-1})^T\|_{1\to 2}^2 - 1 \\ &= (a' + b' + c')^2 + (t-1)(b' + c')^2 + (s-1)t(c')^2 - 1 \\ &= (1 - (t-1)b' - (st-1)c')^2 + (t-1)(b')^2 \\ &\qquad + 2(t-1)b'c' + (t-1)(c')^2 + (s-1)t(c')^2 - 1 \\ &= -2(t-1)b' - 2(st-1)c' + 2(t-1)(st-1)c'b' \\ &\qquad + (t-1)^2(b')^2 + (st-1)^2(c')^2 + (t-1)(b')^2 \\ &\qquad + 2(t-1)b'c' + (t-1)(c')^2 + (s-1)t(c')^2 \\ &\le (tb')^2 + 2st^2b'c' + (stc')^2 - 2tb' - 2stc' \\ &\le (tb' + stc')^2 - 2(tb' + stc') \\ &= (a')^2 - 1. \end{aligned}$$

Substituting, we obtain

$$\begin{aligned} (a')^2 - 1 &\le \left(\frac{te^{(1-r)\alpha_0} + (s-1)t}{e^{\alpha_0} - e^{(1-r)\alpha_0}}\right)^2 + 2\left(\frac{te^{(1-r)\alpha_0} + (s-1)t}{e^{\alpha_0} - e^{(1-r)\alpha_0}}\right) \\ &\le \frac{t^2e^{2\alpha_0} + 2(s-1)t^2e^{\alpha_0} + (s-1)^2t^2 + 2te^{2\alpha_0} + 2(s-1)te^{\alpha_0}}{(e^{\alpha_0} - e^{(1-r)\alpha_0})^2} \\ &\le \left(t\frac{e^{\alpha_0} + s}{e^{\alpha_0} - e^{(1-r)\alpha_0}}\right)^2 \end{aligned}$$

Next, it's easy to see that

$$\begin{aligned} A^{-1}P &= \left(a' I_{\mathcal{X}} + ((b' I_{\mathcal{B}} + c' \mathbf{1}_{\mathcal{B}}) \otimes \mathbf{1}_{\mathcal{C}})\right) (I_{\mathcal{B}} \otimes 1_{\mathcal{C}}) \\ &= a' I_{\mathcal{B}} \otimes 1_{\mathcal{C}} + (b' I_{\mathcal{B}} + c' \mathbf{1}_{\mathcal{B}}) \otimes t 1_{\mathcal{C}} \end{aligned}$$

Each row of the latter consists of one copy of $a' + tb' + tc'$ and $s-1$ copies of tc'. This gives us

$$\begin{aligned} \|(A^{-1}P)^T\|_{1\to 2}^2 - 1 &= (a' + tb' + tc')^2 + (s-1)(tc')^2 - 1 \\ &= (1 - (s-1)tc')^2 + (s-1)(tc')^2 - 1 \\ &= s(s-1)(tc')^2 - 2(s-1)(tc') \\ &\le (stc')^2 - 2(stc'). \end{aligned}$$

Substituting, we obtain

$$\begin{aligned} (stc')^2 - 2(stc') &= \frac{st(st + 2(e^{\alpha_0} + (t-1)e^{(1-r)\alpha_0} - t))}{(e^{\alpha_0} + (t-1)e^{(1-r)\alpha_0} - t)^2} \\ &\le \frac{st^2(s + 2(e^{\alpha_0} - 1))}{(e^{\alpha_0} + (t-1)e^{(1-r)\alpha_0} - t)^2} \end{aligned}$$

Applying Theorem 6.5, we obtain

$$\begin{aligned} &\mathbb{E}[d_{\mathsf{EM}}(\tilde{F}, \tilde{K})] \\ &\le r\sqrt{\frac{st((a')^2 - 1)}{mn}} + \sqrt{\frac{s^2 t(st(c')^2 - 2c')}{mn}} \\ &\le r\sqrt{\frac{st^3}{mn}}\left(\frac{e^{\alpha_0} + s}{e^{\alpha_0} - e^{(1-r)\alpha_0}}\right) + \sqrt{\frac{s^2t^2}{mn}}\left(\frac{\sqrt{s + 2(e^{\alpha_0} - 1)}}{e^{\alpha_0} + (t-1)e^{(1-r)\alpha_0} - t}\right), \end{aligned}$$

finishing the claim. To obtain an asymptotic bound (with budget $\alpha = \varepsilon/r$), we plug in (6), which says that we may set

$$\alpha_0 = \begin{cases} \frac{\alpha}{32\sqrt{m\ln(4m\exp(\alpha)/\delta)}} & \text{if } \alpha \le 32\sqrt{m\ln(4m\exp(\alpha)/\delta)} \\ 2\ln\left(\frac{\varepsilon}{16r\sqrt{m\ln(4m\exp(\alpha)/\delta)}}\right) & 32r\sqrt{m\ln(4m\exp(\alpha)/\delta)} \le \varepsilon \le rm \end{cases}.$$

In the first case, we have

$$\begin{aligned} \frac{e^{\alpha_0} + s}{e^{\alpha_0} - e^{(1-r)\alpha_0}} &\le \frac{s}{r\alpha_0} \\ \frac{\sqrt{s + 2(e^{\alpha_0} - 1)}}{e^{\alpha_0} + (t-1)e^{(1-r)\alpha_0} - t} &\le \frac{2\sqrt{s}}{\alpha_0 t}, \end{aligned}$$

and this implies

$$\begin{aligned} \mathbb{E}[d_{\mathsf{EM}}(\tilde{K}, \tilde{F})] &\le \sqrt{\frac{s^3t^3}{mn}}\frac{1}{\alpha_0} + \sqrt{\frac{s^3t^2}{mn}}\frac{2}{\alpha_0} \\ &\le \frac{64r(st)^{3/2}\sqrt{\ln(4m\exp(\frac{\varepsilon}{r})/\delta)}}{\varepsilon\sqrt{n}}. \end{aligned}$$

In the second, we have

$$\frac{e^{\alpha_0}+s}{e^{\alpha_0}-e^{(1-r)\alpha_0}} = \frac{1+s/e^{\alpha_0}}{1-e^{-r\alpha_0}} \le 2\frac{1+se^{-\alpha_0}}{\min\{1, r\alpha_0\}}$$
$$\frac{\sqrt{s+2(e^{\alpha_0}-1)}}{e^{\alpha_0}+(t-1)e^{(1-r)\alpha_0}-t} \le \frac{\sqrt{2(s+e^{\alpha_0})}}{e^{\alpha_0}}.$$

This implies

$$\begin{aligned}
&\mathbb{E}[d_{\mathsf{EM}}(\tilde{K},\tilde{F})] \\
&\le 2\left(1+\tfrac{1}{r\alpha_0}\right)(1+se^{-\alpha_0})r\sqrt{\frac{st^3}{mn}} + 2\left(e^{-\alpha_0}\sqrt{s}+e^{-\alpha_0/2}\right)\sqrt{\frac{s^2t^2}{mn}} \\
&\le 2(1+se^{-\alpha_0})\sqrt{\frac{st^3}{mn}} + 2\left(e^{-\alpha_0}\sqrt{s}+e^{-\alpha_0/2}\right)\sqrt{\frac{s^2t^2}{mn}} \\
&\le 2(1+\sqrt{s}e^{-\alpha_0/2}+se^{-\alpha_0})\sqrt{\frac{st^3}{mn}} \\
&\le 4(1+se^{-\alpha_0})\sqrt{\frac{st^3}{mn}} \\
&\le 4\sqrt{\frac{st^3}{mn}} + 1024\frac{r^2\sqrt{ms^3t^3}}{\varepsilon^2\sqrt{n}}\ln(4m\exp(\varepsilon/r)/\delta) \\
&\le 4\sqrt{\frac{st^3}{mn}} + 32\frac{r\sqrt{s^3t^3}}{\varepsilon\sqrt{n}}\sqrt{\ln(4m\exp(\varepsilon/r)/\delta)}.
\end{aligned}$$

In both cases, the desired bound has been shown.

D.4 Proof of Lemma 6.7

We use the bound that $d_{\mathsf{EM}}(\tilde{K},\tilde{F}) \le \|\tilde{K}-\tilde{F}\|_1$. In each coordinate, the expected error introduced by the Laplace noise is at most $O(\frac{1}{n\varepsilon})$, and thus $\mathbb{E}[\|\tilde{K}-\tilde{F}\|_1] \le O(\frac{k}{n\varepsilon})$. Normalizing will only reduce this error.

D.5 Proof of Corollary 6.8

Corollary 6.8. *For the metric space $\mathcal{X} = \mathcal{B}\times\mathcal{C}$, $\mathsf{FreqEstLocal}$ with $\mathcal{A} = \mathsf{GKRR}_{\alpha_0}$ with α_0 given in Eq. (7) satisfies (α,δ)-d_{EM} DP in the central model and returns an estimator $\tilde{H}$ with error given in Eq. (9).*

Our mechanism will simply combine the itemsets into one large itemset K with mn elements (and one global user), and then apply the algorithm of Theorem 6.6. By Theorem 4.4, the privacy budget is (α,δ), where

$$\alpha_0 = \begin{cases} \frac{\alpha\sqrt{n}}{32\sqrt{m\ln(4me^{\alpha}/\delta)}} & \text{if } \alpha\sqrt{n} \le 32\sqrt{m\ln(4me^{\alpha}/\delta)} \\ 2\ln\left(\frac{\alpha\sqrt{n}}{16\sqrt{m\ln(4me^{\alpha}/\delta)}}\right) & 32\sqrt{m\ln(4me^{\alpha}/\delta)} < \alpha\sqrt{n} < m\sqrt{n} \end{cases}$$

Following the proof in Section D.3, (and setting $\alpha = \frac{\varepsilon}{r}$), we can show that

$$\mathbb{E}[d_{\mathsf{EM}}(\tilde{K},\tilde{F})] \le 4\sqrt{\frac{st^3}{mn}} + 64\frac{r\sqrt{s^3t^3}}{\varepsilon n}\sqrt{\ln(4m\exp(\varepsilon/r)/\delta)}.$$